From the studied algorithms, LightGBM was the one that performed the best. Besides the metrics obtained, its speed and efficiency in dealing with large and imbalanced datasets, its generalization capacity and it is data interpretability, due to the parameters an training procedure that constitutes LightGBM, made him significantly better then the other classic ML algorithms. So, from the results obtained in this work, it can be perceived that LightGBM constitutes a good choice for dealing with fault classification problem.

5 Conclusion

From this study it can be concluded that the LightGBM algorithms represent a promising avenue for recognizing correlations between the electric and environmental parameters of a solar PV system and, from there, classifying faults accordingly. After a systematic evaluation of studied algorithms, it can also be concluded that ML algorithms prove to be an excellent option for detecting and classifying faults in PV systems. However, it can be stated that there is no one-size-fits-all solution. Each algorithm possesses its own set of advantages and disadvantages.

As a final note, it is important to highlight that the behaviour of the algorithms and their results in the various metrics are in accordance with the data used, and that their behaviour, as well as the results obtained may vary. Indeed, these algorithms are highly susceptible to the data assigned to them.

References

[1] Sara Gallardo-Saavedra, Luis Hernandez-Callejo, Oscar Duque-Perez, Quantitative failure rates and modes analysis in photovoltaic plants. Energy, Volume 183, (2019), pp. 825-836, 0360-5442, https://doi.org/10.1016/j.energy.2019.06.185.

[2] Andre Eugenio Lazzaretti, Clayton Hilgemberg da Costa, Marcelo Paludetto Rodrigues, A monitoring system for online fault detection and classification in photovoltaic plants. Sensors (Switzerland), 20:1–30, 9 (2020), https://doi:10.3390/s20174688.

[3] Fouzi Harrou and Ying Sun and Bilal Taghezouit and Ahmed Saidi and Mohamed Elkarim Hamlati, Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches. Renewable Energy, Elsevier Ltd, (2018), https://doi:10.1016/j.renene.2017.09.048, 18790682.

[4] Paulo Monteiro, Github with Dissertation Scripts, github.com/paulo5930/Dissertation Scripts.

[5] Paulo Monteiro, Pattern Recognition Machine Learning Algorithms for Fault Classification of PV System, Master's Thesis, Faculty of Engineering of the University of Porto, 02-2023.

[6] Wes McKinney, Python for Data Analysis, 3rd edn. O REILLY, (2022), https://wesmckinney.com/book.

[7] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, et al. Scipy 1.0: fundamental algorithms for scientific computing in python, Nature Methods, 17:261–272, (2020), https://doi.org/10.1038/s41592-019-0686-2.

[8] Walter H Delashmit and Michael T Manry, Recent developments in multilayer perceptron neural networks. In Proceedings of the Seventh Annual Memphis Area Engineering and Science Conference, MAESC (2005).

[9] Bahzad Charbuty and Adnan Abdulazeez, Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2. 20-28, (2021), https://doi:10.38094/jastt20165

[10] Sickit Learn developers Homepage, DecisionTreeClassifier, https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html, last accessed 2023/10/19.

[11] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. KNN Model- Based Approach in Classification. Lecture Notes in Computer Science 2888, Pages 986 - 996, (2003).

[12] Jair Cervantes, Farid Garcia-Lamont, Lisbeth Rodríguez-Mazahua, and Asdrubal Lopez. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, Elsevier B.V., (2020), https://doi:10.1016/j.neucom.2019.10.118.

[13] Zhi-Hua Zhou: Machine Learning, Springer, Singapore (2021), https://link.springer.com/book/10.1007/978-981-15-1967-3.

[14] Essam Al Daoud: Comparison between xgboost, lightgbm and catboost using a home credit dataset. World Academy of Science, Engineering and Technology, Open Science Index 145, International Journal of Computer, and Information Engineering, 13(1), 6 - 10, (2019).

[15] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu, Lightgbm. A highly efficient gradient boosting decision tree, https: //github.com/Microsoft/LightGBM, last accessed 2023/10/19.

[16] LightGBM Homepage, lightgbm. LGBMClassifier, MicrosoftCorporation, https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMClassifier.html, last accessed 2023/10/19.

[17] Ankit Gupta, Evaluation Metrics For Multi-class Classification, Homepage Gaggle, https://www.kaggle.com/code/nkitgupta/evaluation-metrics-for-multi-class- classification, last accessed 2023/10/19.

[18] Teemu Kanstren, A Look at Precision, Recall, and F1-Score, Towards Data Science Homepage, https://towardsdatascience.com/a-look-at-precision-recall-and- f1-score-36b5fd0dd3ec, last accessed 2023/10/19.

[19] Sarang Narkhede, Understanding Confusion Matrix, Towards Data Science Homepage, https://towardsdatascience.com/understanding-confusion-matrix- a9ad42dcfd62, last accessed 2023/10/19.