World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Classification of Kepler exoplanet searching from galaxy using machine learning model

Sasmita Kumari Nayak *

Computer Science and Engineering, Centurion University of Technology and Management, Odisha, India.

World Journal of Advanced Research and Reviews, 2024, 21(01), 227-230

Publication history: Received on 22 November 2023; revised on 01 January 2024; accepted on 04 January 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.21.1.0012

Abstract

The Kepler exoplanet mission was designed specifically to search the Milky Way galaxy for numerous small planets the size of Earth that are either in or close to the habitable zone. An exoplanet is a planet that orbits stars outside of our solar system. The dataset, which was gathered from the Kaggle website, has 50 columns and 9564 samples. The target variable in the dataset, KOI-disposition, comprises candidate, false positive and confirmed data. We discovered that 5000 samples out of 9564 samples were false positives; each confirmed sample had 2282 candidates. While there are many stars and planets in the Milky Way galaxy, we have only looked at a small number of them. In order to search for exoplanets beyond our stellar atmosphere, we have employed machine learning algorithms on stars and planets, such as decision trees, random forests, KNN classification, and Naive Bayes classification.

Keywords: Kepler Exoplanets data; KNN; Random Forest; Decision Tree.

1. Introduction

2009 saw the launch of the NASA-built satellite known as the Kepler Space Observatory. The telescope is dedicated to searching for exoplanets in star systems beside our own, with the ultimate goal of possibly finding other habitable planets besides our own. The mission's findings yielded information on a broad spectrum of planets and planetary systems orbiting one or more stars with varying sizes, temperatures, and ages [1]. The insights from earlier studies on leaf disease classification with machine learning algorithms are presented at this part. The section begins by explaining the underlying theory of machine learning algorithms [2-12]. The majority of machine learning Algorithms used for prediction as well as classification is then explored. Deep learning and optimization algorithms researches are also conducted by the image classification. The readers can get the idea of machine learning and deep learning in details from the papers [13-19].

The telescope has been operational since 2014 on a "K2" extended mission; however, the original mission ended in 2013 due to some technical issues. As part of its extended mission, the telescope is still operational and continues to gather new data. Our project's primary goal is to search for exoplanets using kepler data to obtain the target variable, koi_disposition, which contains false positive, confirmed, and candidate. koi_slogg, koi_impact, koi_depth, koi_period, koi_duration, koi_impact, and koi_score. Estimates of parent distributions are highly uncertain until the effects of these biases are accurately identified and the entire range of orbital periods is taken into account.

This article explains the initial attempts to use machine learning algorithms for classifying the Kepler Exoplanet. Section 4 concludes by classifying the feasibility of machine learning algorithms in an effort to offer a better solution.

* Corresponding author: Sasmita Kumari Nayak

2. Machine Learning Model

Machine learning is an artificial intelligence approach and a subfield of computer science. This method has the benefit of allowing a model to address issues that are not amenable to explicit algorithms, and it can be applied in multiple domains. A thorough analysis of various deterministic and machine learning techniques for predicting food, crop yield, weather, and hepatitis is provided in [20-28]. Even in situations where representation is not feasible, machine learning models identify relationships between inputs and outputs; this characteristic allow the use of machine learning models in many cases, for example in data mining and forecasting problems, spam filtering, classification problems, and pattern recognition. Because one must work with large datasets and machine learning models can handle pre-processing and data preparation, the classification and data mining aspects of this field are especially intriguing. Following this stage, forecasting issues can be solved using the machine learning models.

2.1. Decision Tree

A decision tree is a type of decision support tool that shows potential decisions, outcomes, or reactions using a model or chart that resembles a tree [29].

2.2. Random Forest

The random forest is a model made up of many decision trees. This algorithm combines the output of multiple decision trees to generate the final output [24, 30].

2.3. K-Nearest Neighbors (KNN)

One of the most basic machine learning algorithms for regression and classification problems is K-Nearest Neighbors (KNN). Using similarity metrics, KNN algorithms classify new data points using existing data. An efficient machine learning technique for both regression and classification problems is K-nearest neighbors, or CNN. The idea behind KNN is to use the distance between an unknown sample and the K closest samples in the training set to classify it. The process of classification involves designating the most prevalent class among the K closest neighbors. Because KNN is a lazy learning algorithm, all that needs to be done for training is storing the training data. KNN is quick and memory-efficient because the real classification or regression of fresh samples is done at the prediction stage. KNN can handle both linear and nonlinear data, and it is simple to comprehend and apply. Nevertheless, the selection of K, the size of the features, and features that are not relevant can affect how well it performs. Unlabeled observations are classified using a KNN classifier by putting them in the same class as the labeled examples that are the closest to them. Both the training and test dataset's characteristics are gathered. For instance, the crunchiness and sweetness of fruit, vegetables, and grains can be used to identify them [31].

3. Results and Discussion

For this project, gathering datasets is essential. We have gathered 9564 samples and 50 feature columns from the Kaggle website. Our goal in using this dataset is to find the exoplanet, so we have tried a number of different algorithms. We have been using Google Colab to run the program for experimental purposes. Regression modeling has been used by many to analyze this dataset, but since exoplanets are in categorical format, we considered classification [32]. The accuracy of all machine learning model for classifying the Kepler Exoplanet is shown in Table 1.

Table 1 Classification Results Using Machine Learning Models

Machine Learning Classification Models	Accuracy of the Model (%)
KNN	67%
Decision Tree	82%
Random Forest	86%

From the above table, we can easily conclude that Random Forest model has the best accuracy in comparison with other implemented machine learning models. The accuracy score of Random Forest as well as accuracy score also found out to be maximum.

4. Conclusion

This paper is a companion work to an experiment conducted on a dataset containing kepler object of interest data. Features seen from the Kepler satellite are included in the dataset. With the aid of this dataset, we can visualize the data and gain a thorough understanding of it through basic data analysis exploration. The machine learning models were chosen based on how well the model would perform in a binary classification when compared to other experiments under similar conditions. Cross-validation was used to run the models flawlessly and determine the ideal meta-parameters for each model. The Kepler mission determined the number of stars that supported planets and, more importantly, estimated the frequency of planets similar to Earth.

References

[1] Borucki, William, et al. "KEPLER: search for Earth-size planets in the habitable zone." Proceedings of the International Astronomical Union 4.S253 (2008): 289-299

[2] Nayak, S. K., Padhy, S. K., & Panigrahi, S. P. (2012). A novel algorithm for dynamic task scheduling. Future Generation Computer Systems, 28(5), 709-717.

[3] Nayak, S. K., Panda, C. S., & Padhy, S. K. (2018). Efficient multiprocessor scheduling using water cycle algorithm. Soft Computing Applications, 131-147.

[4] Nayak, S. K., Panda, C. S., & Padhy, S. K. (2018). Efficient multiprocessor scheduling using water cycle algorithm. In Soft Computing Applications (pp. 131-147). Springer, Singapore.

[5] Nayak S.K., Padhy S.K., Panda C.S. (2018) Efficient Multiprocessor Scheduling Using Water Cycle Algorithm. In: Pant M., Ray K., Sharma T., Rawat S., Bandyopadhyay A. (eds) Soft Computing: Theories and Applications. Advances in Intelligent Systems and Computing, vol 583. Springer, Singapore.

[6] Nayak, S. K., Panda, C. S., & Padhy, S. K. (2019, February). Dynamic Task Scheduling Problem Based on Grey Wolf Optimization Algorithm. In 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP) (pp. 1-5). IEEE.

[7] S. K. Nayak, C. S. Panda and S. K. Padhy, "Dynamic Task Scheduling Problem Based on Grey Wolf Optimization Algorithm," 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok, India, 2019, pp. 1-5, doi: 10.1109/ICACCP.2019.8882992.

[8] Nayak, S. K., & Panda, C. S. (2021). Dynamic task scheduling using nature inspired algorithms. J. Math. Comput. Sci., 11(1), 893-913.

[9] Nayak, S. K. Multiprocessor Scheduling Using Nature Inspired Optimization. Thesis. Sambalpur University, Odisha, 2021. Web. http://hdl.handle.net/10603/464925

[10] Nayak, S. K., Panda, C. S., & Padhy, S. K. (2018). Multiprocessor Scheduling using Krill Herd Algorithm (KHA). International Journal of Computer Sciences and Engineering, 6(6), 7-17.

[11] S.K. Nayak, C.S. Panda, S.K. Padhy, "Multiprocessor Scheduling using Krill Herd Algorithm (KHA)," International Journal of Computer Sciences and Engineering, Vol.6, Issue.6, pp.7-17, 2018. CrossRef-DOI: https://doi.org/10.26438/ijcse/v6i6.717

[12] Nayak, S. K., & Panda, C. S. (2019). Multiple Processor Scheduling with Optimum Execution Time and Processor Utilization Based on the SOSA. International Journal of Recent Technology and Engineering, ISSN: 2277-3878. Volume-8, Issue-2, July, 5463-5471.

[13] Swain, S., Nayak, S. K., & Barik, S. S. (2020). A review on plant leaf diseases detection and classification based on machine learning models. Mukt shabd, 9(6), 5195-5205.

[14] S Sucharita, S Nayak, S Panda et al., "Human Face Recognition using LBPH", International Journal of Recent Technology and Engineering, vol. 8, no. 6, pp. 3208-3212, 2020.

[15] Stitiprajna Panda*, Swati Sucharita Barik, Sasmita Kumari Nayak, Aeisuriya Tripathy, and Gourav Mohapatra. 2020. Human Face Recognition using LBPH. International Journal of Recent Technology and Engineering (IJRTE) 8, 6 (2020), 3208-3212. DOI:https://doi.org/10.35940/ijrte.f8117.038620

[16] Jena, T. R., Barik, S. S., & Nayak, S. K. (2020). Electricity consumption & prediction using machine learning models. Acta Tech. Corviniensis-Bull. Eng, 9, 2804-2818.

[17] Nayak, S. K., Barik, S. S., & Beura, M. (2020). Analysis of Infectious Hepatitis Disease with High Accuracy Using Machine Learning Techniques. TEST Engineering & Management, 83, 83.

[18] Nayak, S. K. (2020). Analysis and high accuracy prediction of coconut crop yield production based on principle component analysis with machine learning models. International Journal of Modern Agriculture, 9(4), 359-369.

[19] Sasmita Kumari Nayak, Swati Sucharita Barik, Mamata Beura," Weather Forecasts Based on Rainfall Prediction Using Machine Learning Methodologies," Adalya Journal 9 (6), Page No: 72 – 80, ISSN NO: 1301-2746.

[20] Mishra, S. P., Siddique, M., Beura, M., & Nayak, S. K. (2021). Analysis of Indian Food Based on Machine learning Classification Models. Journal of Scientific Research and Reports, 27(7), 1-7.

[21] Nayak, S. K. (2023). Classification of cyclones using machine learning techniques. World Journal of Advanced Research and Reviews, 20(2), 433-440.

[22] Sasmita Kumari Nayak. (2020). CONSTRICTING THE BIVARIATE ANALYSISOFHIGHEST CROP YIELD PRODUCTION BASED ON DIFFERENT ZONES OF INDIA. International Journal of Modern Agriculture, 9(4), 216 - 226. Retrieved from https://modern-journals.com/index.php/ijma/article/view/205

[23] Sasmita Kumari Nayak, Mohammed Siddique. (2020). EFFECT OF STOCK INDEX PARAMETERS ON FORECASTING THE HIGH STOCK VALUE OF VISA STEEL USING DEEP LEARNING NEURAL NETWORK MODEL. International Journal of Modern Agriculture, 9(4), 227 - 236. Retrieved from https://modern-journals.com/index.php/ijma/article/view/207

[24] Sowmya Jagadeesan, B. B. (2022). A Perishable Food Monitoring Model Based on IoT and Deep Learning to Improve Food Hygiene and Safety Management. IJFANS International Journal of Food and Nutritional Sciences, 11(8), 1164-1178.

[25] Jagadeesan, S., Barman, B., Agarwal, R. K., Srivastava, Y., Singh, B., Nayak, S. K., & Venu, N. A Perishable Food Monitoring Model Based On Iot And Deep Learning To Improve Food Hygiene And Safety Management. interventions, 8, 9.

[26] Rajesh, J., Ashraf, M. S., Kaur, L., Rout, S., Nayak, S. K., Kaur, G., & Saikanth, D. R. K. APPLICATION OF FUZZY LOGIC IN SMART AGRICULTURE TO RECOGNISE TOMATO FRUIT RIPENESS. IJFANS International Journal of Food and Nutritional Sciences, 11(1), 2360-2367.

[27] N. Dash, S. K. Nayak and J. Majumdhar, "Detection of Cut Transition in Videos Using Optical Flow and Clustering," 2021 Asian Conference on Innovation in Technology (ASIANCON), PUNE, India, 2021, pp. 1-7, doi: 10.1109/ASIANCON51346.2021.9544553.

[28] Majumdhar, J., & Nayak, S. K. (2021, August). A Novel Method on Summarization of Video Using Local Ternary Pattern and Local Phase Quantization. In 2021 2nd International Conference on Range Technology (ICORT) (pp. 1-6). IEEE.

[29] Chakravarthy, A., Panda, B. S., & Nayak, S. K. (2023). Review and Comparison for Alzheimer's Disease Detection with Machine Learning Techniques. International Neurourology Journal, 27(4), 403-409.

[30] Nayak, S. K. (2023). Nature inspired algorithms in dynamic task scheduling: A review. World Journal of Advanced Research and Reviews, 20(2), 829–833.

[31] Nayak, S. K. (2023). Exploring and forecasting of solar radiation with machine learning methods. World Journal of Advanced Research and Reviews, 20(2), 824–828.

[32] Borucki, William J. "Kepler: A brief discussion of the mission and exoplanet results." Proceedings of the American Philosophical Society 161.1 (2017): 38.