Rainfall Prediction Using NASA Earth Observation Data and Ensemble Learning

Ayşegül KAYA [1]*, Ali ÇETİNKAYA[2]

[1]* *Department of Computer Engineering, Selcuk University Faculty of Technology, Selçuklu, Konya*
[2] *Department of Computer Engineering, Selcuk University Faculty of Technology, Selçuklu, Konya*

(213311040@ogr.selcuk.edu.tr)

Abstract – Outdoor event planning faces significant challenges due to weather uncertainties, with traditional forecasting methods often proving inadequate for location-specific, long-term predictions. This study presents a machine learning-powered rainfall prediction system leveraging 44 years (1981-2025) of NASA Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) Earth observation data. The system addresses the critical need for accurate, data-driven weather forecasting to support event planners in making informed decisions. An ensemble learning approach combining Random Forest and XGBoost algorithms was implemented to predict rain probability for specific dates and locations. The model processes 44+ meteorological features including temperature at 2 meters (T2M), relative humidity at 2 meters (RH2M), cloud coverage (CLOUD_AMT), surface pressure (PS), and wind parameters, with advanced feature engineering techniques such as moisture index and temperature range calculations. Experimental results demonstrate strong predictive performance with 86.0% accuracy, 83.4% precision, 77.9% recall, 80.5% F1-score, and 93.2% Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) score. The system was validated through retrospective testing and successfully predicted weather conditions for outdoor events in Konya, Turkey with high accuracy. Cross-regional validation across multiple cities confirmed the model's robustness and generalizability. The developed system provides three core functionalities: specific date prediction with confidence levels, optimal date recommendation within a given month, and comprehensive event planning reports. This work demonstrates that historical Earth observation data, when processed through ensemble machine learning techniques, can deliver superior performance compared to traditional statistical methods in weather prediction applications.

Keywords – Ensemble Learning; XGBoost / Random Forest; NASA MERRA-2 Earth Observation; Long-Term Rainfall Forecasting

REFERENCES

[1] NASA Langley Research Center (LaRC) Prediction of Worldwide Energy Resources (POWER) Project, "POWER Data Access Viewer," 2024. [Online]. Available: https://power.larc.nasa.gov/api/pages/

[2] NASA Global Modeling and Assimilation Office, "MERRA-2: Modern-Era Retrospective analysis for Research and Applications, Version 2," Goddard Earth Sciences Data and Information Services Center (GES DISC), 2015. Available: https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/

[3] W. McKinney, "Data structures for statistical computing in Python," in Proc. 9th Python in Science Conference, 2010, pp. 51–56. DOI: 10.25080/Majora-92bf1922-00a

[4] "pandas: Powerful Python data analysis toolkit," pandas development team, 2024. [Online]. Available: https://pandas.pydata.org/docs/

[5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794. DOI: 10.1145/2939672.2939785

[6] "XGBoost Documentation," XGBoost developers, 2024. [Online]. Available: https://xgboost.readthedocs.io/

[7] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[8] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. DOI: 10.1023/A:1010933404324

[9] D. Wolpert, "Stacked generalization," Neural Networks, vol. 5, no. 2, pp. 241–259, 1992. DOI: 10.1016/S0893-6080(05)80023-1

[10] S. Rasp, P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey, "WeatherBench: A benchmark data set for data-driven weather forecasting," Journal of Advances in Modeling Earth Systems, vol. 12, no. 11, 2020. DOI: 10.1029/2020MS002203

[11] A. Grover, A. Kapoor, and E. Horvitz, "A deep hybrid model for weather forecasting," in Proc. 21st ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2015, pp. 379–386. DOI: 10.1145/2783258.2783275

[12] V. M. Krasnopolsky, "The application of neural networks in the Earth system sciences," in Neural Networks Emulations for Complex Multidimensional Mappings. Springer, 2013, pp. 1–18.

[13] M. Gevrey, I. Dimopoulos, and S. Lek, "Review and comparison of methods to study the contribution of variables in artificial neural network models," Ecological Modelling, vol. 160, no. 3, pp. 249–264, 2003. DOI: 10.1016/S0304-3800(02)00257-0

[14] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, 2015. DOI: 10.1016/j.neunet.2014.09.003

[15] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Morgan Kaufmann, 2016.