Rainfall Prediction of Konkan and Goa Region from 1901 to 2017

DR. SANTOSH SINGH[1], MITHLESH VISHWAKARMA[2], SHRAVAN SHIVANAND KAMAT[3], HRUSHI BHOLA[4]

[1] *H.O.D, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*

[2] *Assistant professor, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*

[3, 4] *PG Student, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India*

Abstract- This research dives deep into analyzing and predicting rainfall patterns in Konkan and Goa from 1901 to 2017. Given the area's ecological importance and the region's reliance on agriculture, understanding historical trends becomes crucial. Using data-driven approaches, particularly machine learning algorithms such as Random Forest, Back Propagation, SVR, Linear Regression, Decision Tree, this study aims to create precise predictive models from meteorological data. We have found different accuracies for different models based on the above-mentioned algorithms for which we have plotted suited graphs for data visualization for better understanding patterns of each model. These models seek to forecast annual rainfall, offering invaluable insights for resource management, agricultural planning, and disaster readiness in Konkan and Goa.

Indexed Terms- Climate Modelling, Goa Region, Konkan Region, Machine Learning Algorithms, Meteorological Data Analysis, Rainfall Prediction

I. INTRODUCTION

The Konkan and Goa region, nestled along the southwestern coast of India, presents a distinctive blend of topography and coastal influence that intricately shapes its climate dynamics. This region, marked by its diverse and complex rainfall patterns, plays a crucial role in the lives of its inhabitants who heavily depend on agriculture as a primary livelihood. The confluence of geographical features and the proximity to the Arabian Sea results in a climatic mosaic that demands a nuanced understanding of historical rainfall trends. The imperative for accurate and timely predictions of annual rainfall becomes paramount, given the region's reliance on agriculture and the necessity to inform sustainable practices and water resource management. Understanding historical trends in rainfall has grown more and more important as a result of the changing climate conditions that have emerged over time. This study begins a thorough examination of rainfall data in Konkan and Goa from 1901 to 2017, a century and a half. Finding patterns and trends that can offer important insights into the climate behavior of the area is the main objective. This research aims to go beyond conventional methodologies and create predictive models for annual rainfall by utilizing the power of advanced machine learning algorithms, such as Random Forest, Back Propagation, Support Vector Regression (SVR), Linear Regression, Decision Tree. The application of machine learning techniques in climate science marks a paradigm shift, offering the potential for more precise and reliable forecasts. These forecasts, generated through a fusion of historical data and cutting-edge algorithms, are envisioned to serve as a linchpin for decision-makers, farmers, and disaster authorities. By forecasting future rainfall, these predictive models aim to empower communities, enabling them to proactively adapt to evolving weather conditions and bolster their resilience. The machine learning algorithms employed in this research each bring distinct functionalities and advantages to the table. The Random Forest method, for instance, harnesses the collective power of multiple decision trees, ensuring accuracy and resilience in the face of complex datasets and missing values. Back Propagation, a cornerstone in training multi-layer artificial neural networks, iteratively adjusts weights to minimize errors, showcasing efficacy in deep learning networks. Support Vector Regression

(SVR), a robust algorithm, seeks optimal hyperplanes to segregate different classes in high-dimensional spaces, proving versatile in both classification and regression tasks.

Linear Regression, a stalwart in statistical modelling, establishes linear relationships between variables, offering simplified interpretations that unveil insights into these relationships. This research, anchored in advanced machine learning methodologies, endeavors to not only analyze historical rainfall data but also to project climate changes in Konkan and Goa attributable to the overarching specter of climate change. The pursuit of more precise forecasts than conventional models aims to furnish practical strategies that can genuinely benefit the communities and people in these regions. In doing so, this research seeks to enhance their resilience to the ever-evolving challenges posed by shifting weather patterns.

In conclusion, this research aims not only to unravel historical rainfall patterns but also to pioneer a new frontier in climate science by harnessing the power of machine learning. By providing accurate and timely rainfall predictions, the study aspires to equip stakeholders with the knowledge needed for informed decision-making, contributing to the resilience of communities in the face of evolving weather conditions.

II. RELATED WORK

Rainfall prediction holds immense importance in the Konkan and Goa regions due to its impact on agriculture, water resource management, ecosystem health, disaster preparedness, tourism, infrastructure planning, and cultural practices. Accurate predictions assist farmers in planning crops and irrigation, aid in managing water resources, preserve biodiversity, enable disaster preparedness, support tourism planning, ensure infrastructure resilience, and help communities maintain cultural traditions linked to seasonal changes. Overall, rainfall prediction profoundly influences various aspects of life, economy, and the environment, contributing significantly to the region's well-being and sustainability.

K. W. Wong et al. [10] proposed models consisting of self-organizing map, Backpropagation neural networks and Fuzzy system. These models consist of one model using soft computing techniques and another model using ANN techniques. Model using soft computing techniques gave a RMSE of 72.95 with a relative error of 0.31 while model using ANN techniques gave a RMSE of 78.65 with a relative error of 0.46.

Suhaila Zainudin et al. [9] used various techniques such as Support Vector Machine, Naïve Bayes, Decision Tree, Neural Network, and Random Forest to predict rainfall. Here Random Forest has 1043 correctly classified instances at 10% training data.
We are going to predict rainfall using various techniques such as Random Forest, Back Propagation, Support Vector Regressor (SVR), Linear Regression and Decision Tree. and compare them on basis of different error measures.

III. METHODOLOGY

A. Data Collection

The collected historical rainfall data from 1901 to 2017 covers more than a century, and it offers a comprehensive and extensive picture of the dynamics of precipitation in the Konkan and Goa region. In addition to capturing the annual rainfall records, this large dataset enables a thorough exploration of inter-annual variability, showing possible patterns and anomalies in the climate. A thorough foundation like this creates opportunities for in-depth climatological research and improves our knowledge of the region's vulnerability to extreme weather events and climate change. In order to develop policies for the sustainable management of water resources and climate resilience in the Konkan and Goa region, researchers, climatologists, and policymakers will find great use in this information.

B. Data Preprocessing

Conducted data cleaning to address missing values, outliers, and inconsistencies. Normalized and standardized the data to ensure uniformity and facilitate algorithm convergence. Applied feature engineering to extract relevant information and enhance the dataset for machine learning model input. We extracted Konkan and Goa Region's Data from Country's Rainfall dataset after cleaning unnecessary data columns like non-Monsoon Months and Collective Months' Average Data Columns.

C. Exploratory Data Analysis (EDA)

By utilizing descriptive statistics and visualizations, a thorough analysis of the historical rainfall patterns in the Konkan and Goa region was conducted. This approach revealed potential trends, seasonality, and anomalies in the data. Descriptive statistics provided quantitative insights, while visualizations offered an intuitive representation of the precipitation dynamics from 1901 to 2017. This comprehensive analysis aids in understanding climatic variations, identifying periodic trends, and pinpointing instances of extreme weather events over the studied period.

D. Feature Selection

Conducted a thorough examination of influential features impacting rainfall in Konkan and Goa. Leveraging statistical methods and domain knowledge, selected key features include annual average rainfall, essential monsoon months, and their collective averages. This refined dataset serves as a solid foundation for predictive models, capturing both annual trends and seasonal dynamics in the region's rainfall patterns.

E. Machine Learning Models

a) Random Forest

Random Forest regression predicts 'JUN' from 'YEAR' in a 80-20% training-testing split with 100 trees. R-squared quantifies variance explained, providing a comprehensive performance measure.

b) Back Propagation (Neural Networks)

Backpropagation optimizes neural network training by iteratively adjusting weights based on gradients, enhancing predictive accuracy by minimizing mean squared error through 'adam' optimization.

c) Support Vector Machine (SVM)

Support Vector Regression models the 'YEAR'-'JUN' relationship in Konkan and Goa dataset. R-squared and mean squared error evaluate performance, including a graphical representation of predictions and actual values.

d) Linear Regression

Linear regression models 'YEAR'-'JUN' in 80-20% split, forecasting June rainfall for 2017. R-squared measures variance explanation, with a scatter plot depicting model efficacy.

e) Decision Tree

Decision Tree Regressor predicts 'JUN' from 'YEAR' in 80-20% split, using R-squared and mean squared error for evaluation. Results, including accuracy percentage, provide a thorough assessment.

F. Model Evaluation

To evaluate the performance of each prediction model, relevant measures such as Mean Squared Error, R-squared, and accuracy were used. Using cross-validation approaches, the models' resilience and generalizability were confirmed. The stringent assessment procedure ensures that the models are dependable in identifying and forecasting precipitation trends in the Konkan and Goa area.

IV. RESULTS

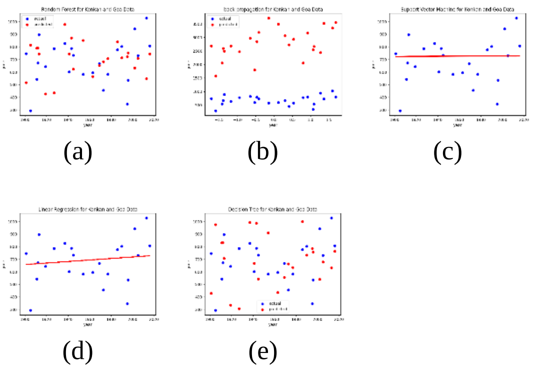
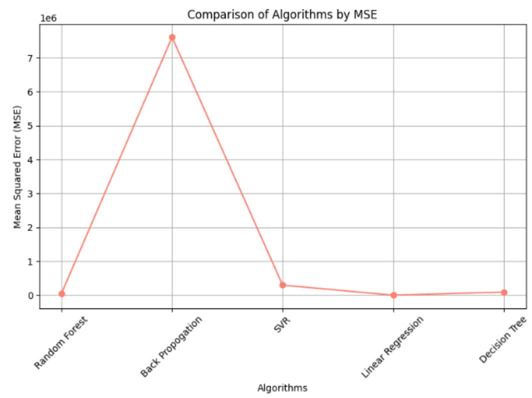
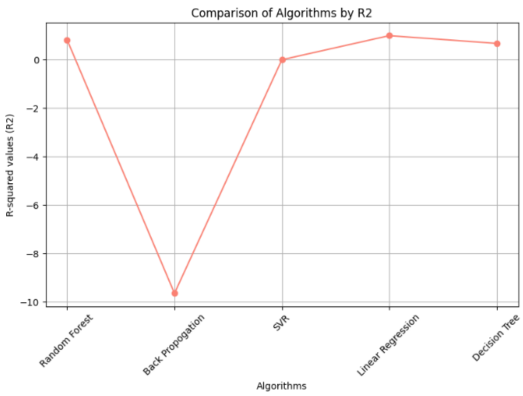


Figure 1. Graph represents actual versus predicted rainfall in month of June. (a)Random Forest (b)Back propagation (c)SVR (d)Linear regression (e)Decision tree



(a)



(b)

Figure 2. graph represents comparison of algorithms with respect to R squared value. (a)Mean Squared Error (b)R squared value.

Algorithm	MSE	R^2
Random Forest	54533.500293916244	0.8176949889999806
Back Propagation	7607112.498888947	-24.430510059209375
SVR	297845.5798349936	0.00430485086277832
Linear Regression	4.904922538498031e-24	1.0
Decision Tree	90084.72833333332	0.6988475468981729

Table 1. MSE and R^2 value of various algorithms.

V. CONCLUSION AND FUTURE WORK

Accurate rainfall prediction in the Konkan and Goa region holds immense significance across various sectors, impacting agriculture, water resource management, disaster preparedness, tourism, infrastructure planning, and cultural practices. The selection of prediction methods, considering factors like data availability, computational requirements, and regional complexities, plays a pivotal role in influencing the accuracy and applicability of forecasts. Based on the results as shown in Figure 1, Figure 2 and Table 1, the Random Forest algorithm has the lowest mean squared error and the highest R-squared value. This means that it is the most accurate algorithm for predicting rainfall in the Konkan and Goa region. Improving prediction accuracy involves integrating diverse techniques, such as machine learning algorithms, physical models, and remote sensing technologies, to capture complex rainfall patterns specific to the region. Robust validation methods and data assimilation techniques are crucial for refining models continuously, enhancing reliability, and addressing local variations. By leveraging advancements in technology and integrating multiple approaches, the goal is to achieve more accurate and reliable rainfall predictions, ultimately contributing to informed decision-making, sustainable development, and resilience against climatic variations in the Konkan and Goa region.

REFERENCES

[1] Ari Yair Barrera-Animas Et. al. 5 more authors, Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting, Machine Learning with Applications, Volume 7, 2022 [https://www.sciencedirect.com/science/article/pii/S266682702100102X]

[2] W. Wang and S. Yang, "Research on Air Quality Forecasting Based on Big Data and Neural Network," 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 2020 [https://ieeexplore.ieee.org/document/9239752]

[3] S. K. Sunori, P. B. Negi, S. Maurya, A. Mittal, N. Bisht and P. Juneja, "Developing Soft Computing based Models for Prediction of Pollutant PM10 of Air," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2021 [https://ieeexplore.ieee.org/document/9453037]

[4] Gouda, K C, Sanjeeb Kumar Sahoo, Payoshni Samantray, and Himesh Shivappa. 2017. "Comparative Study of Monsoon Rainfall Variability over India and the Odisha State" Climate 5, no. 4: 79. https://doi.org/10.3390/cli5040079

[5] Q. Tao, F. Liu, Y. Li and D. Sidorov, "Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU," in IEEE Access, vol. 7 [https://ieeexplore.ieee.org/document/8732985]

[6] Shivang Agarwal Et. al. 10 more authors, Air quality forecasting using artificial neural networks with real time dynamic error correction in highly polluted regions, Science of The Total Environment, Volume 735, 2020 [https://www.sciencedirect.com/science/article/abs/pii/S0048969720329715]

[7] C. Zhao, M. van Heeswijk and J. Karhunen, "Air quality forecasting using neural networks," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 2016 [https://ieeexplore.ieee.org/document/7850128]

[8] Y. -T. Tsai, Y. -R. Zeng and Y. -S. Chang, "Air Pollution Forecasting Using RNN with LSTM," 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress, Athens, Greece, 2018 [https://ieeexplore.ieee.org/document/8512020]

[9] Zainudin, S., Jasim, D.S. and Bakar, A.A., 2016. Comparative analysis of data mining techniques for Malaysian rainfall prediction. *Int. J. Adv. Sci. Eng. Inf. Technol*, *6*(6), pp.1148-1153. [https://www.researchgate.net/profile/Suhaila-Zainudin-2/publication/311917800_Comparative_Analysis_of_Data_Mining_Techniques_for_Malaysian_Rainfall_Prediction/links/586c9c6308ae8fce4919ed86/Comparative-Analysis-of-Data-Mining-Techniques-for-Malaysian-Rainfall-Prediction.pdf]

[10] Wong, K.W., Wong, P.M., Gedeon, T.D. and Fung, C.C., 2003. Rainfall prediction model using soft computing technique. *Soft Computing*, *7*, pp.434-438. [https://link.springer.com/article/10.1007/s00500-002-0232-4]