Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

AI-Driven Groundwater Level Enhancement System using Advanced Prediction Algorithms

Ranganathan S[1], Ranjith Kumar K[2], Vignesh M[3]

Department of Artificial Intelligence and Data Science, Karpagam College of Engineering,

Coimbatore, India

E-mail: [1]ranganathansaravanan18@gmail.com, [2]ranjithkumar201200@gmail.com, [3]vigneshm90160@gmail.com,

Abstract

This research focuses on predicting water sources in various areas by analyzing historical data on groundwater levels, rainfall, and borewells. The study explores the relationships between groundwater levels and environmental factors, emphasizing the influence of rainfall on aquifer recharge. Borewell data, including depth and water quality, is incorporated to identify potential water sources. The research involves data cleaning, exploratory analysis, and machine learning to predict groundwater levels based on diverse features such as rainfall patterns and geographical characteristics. Spatial analysis using GIS tools visualizes the distribution of groundwater levels and rainfall. The model's performance is evaluated, considering metrics and local hydrogeological conditions, with an emphasis on integrating borewell data. Continuous monitoring and updates ensure the model's ongoing relevance. This integrated approach aims to provide insights for sustainable water resource management, assisting decision-makers in planning water sources in diverse areas.

Keywords: Groundwater level, Linear regression, Water, Prediction. Artificial Intelligence

Extended version of article received from 4th INTERNATIONAL CONFERENCE on Artificial Intelligence,5G Communication and Network Technologies
DOI: https://doi.org/10.36548/jscp.2024.1.005
Received: 02.03.2024, received in revised form: 06.04.2024, accepted: 15.04.2024, published: 29.04.2024

1. Introduction

Water scarcity is a pressing global concern, exacerbated by increasing population demands, climate variability, and anthropogenic activities. Addressing this challenge requires a nuanced understanding of the intricate interactions between environmental factors influencing groundwater levels, precipitation patterns, and the availability of potential water sources such as borewells. In response to this need, the research embarks on a comprehensive exploration of historical data to predict water sources in diverse areas [4-10]. The focal point of the study is the analysis of groundwater levels, rainfall, and borewell data, seeking to unravel the complex relationships governing aquifer recharge dynamics. By employing advanced data science techniques, including data cleaning, exploratory data analysis, and machine learning, we aim to develop a predictive model capable of forecasting groundwater levels based on an array of diverse features such as rainfall patterns and geographical characteristics [11-13]. In addition to these predictive insights, the research integrates borewell data, encompassing factors like depth and water quality, to identify and assess potential water sources. The spatial distribution of groundwater levels and rainfall is visualized through Geographic Information System (GIS) tools, offering a geographical context to the analysis [14-16]. This spatial component enriches the understanding of water resource dynamics and aids decision-makers in crafting region-specific strategies for sustainable water management. An integral aspect of our approach is the ongoing relevance and adaptability of our model. We emphasize continuous monitoring and updates to ensure the model's accuracy remains attuned to evolving hydrogeological conditions [17,18]. This integrated and dynamic methodology seeks to provide decision-makers with actionable insights for planning water sources in diverse areas, thereby contributing to sustainable water resource management practices. As we delve into the intricate web of environmental and hydrogeological data, the research aspires to be a valuable resource in the ongoing global effort to secure and manage water resources effectively. Through a synthesis of historical data, machine learning, and spatial analysis, we aim to contribute meaningful solutions to the challenges posed by water scarcity, fostering a more resilient and sustainable water future[19,20].

2. Related Works

Farouk Boukredera [1] et al. presented the depth-of-cut when utilizing machine learning models for bit testing and ROP optimization, as well as for selecting the most suitable drilling settings. When compared to the current state-of-the-art models and practices, the suggested method offers a practical means of reducing time, cost, and effort. This investigation, conducted by F.S. Boukredera and his team in 2021, indicates that the suggested procedure has the potential to significantly lower drilling expenses and save time, presenting a more cost-effective and secure alternative to traditional field-testing techniques. The simulations reveal an average increase of 43% in torsional stability and ROP when the improved parameters are implemented. The study's findings highlight the potential of the suggested workflow in mitigating the adverse impacts of drill-string vibrations on drilling efficiency, emphasizing its ability to reduce costs and save time while providing a more affordable and secure alternative to conventional field-testing methods.

Yazdani [2] et al. proposed the integration of realm drilling parameter optimization for motorized bottom hole assembly. They employ a multi-objective optimization method to tackle this task. Their approach involves crafting a comprehensive model that considers various factors such as the rate of penetration, the intricacies of the positive displacement motor, mechanical specific energy, and hydraulic system. To develop this model, a blend of experimental data and numerical simulations was used. To fine-tune and extract the optimal parameters that enhance drilling efficiency while simultaneously reducing drilling time, the authors employed a genetic algorithm.

Aditi Nautiyala [3] et al. Exposed the application of an ensemble model consisting of 25 decision trees. They employed the Broyden-Fletcher-Goldfarb-Shanno algorithm (LBFGS) solver with a decision tree as the base learner. After conducting their analysis, the authors determined that among the three machine learning models they investigated, the Random Forest (RF) model demonstrated the highest level of accuracy. To build their model, the researchers utilized drilling parameters such as WOB, RPM, Flow Rate, and Depth. They adopted a 10-fold cross-validation approach for data training and model evaluation, ensuring its accuracy and generalizability. The study highlights the potential of machine learning in enhancing the efficiency and cost-effectiveness of oil and gas well drilling operations, which is a crucial aspect of drilling optimization.

HsinYu Chen [7] et al. demonstrate the environmental ecosystems and the progress of civilization greatly hinge on our water resources. Groundwater, in particular, plays a crucial role in supplementing our often-limited surface water reserves, catering to the needs of households, industries, and agriculture. However, unlike some other natural resources, groundwater accumulates slowly over many years and cannot be swiftly replenished. Consequently, the sustainable management of groundwater supplies is a critical consideration for long-term development.

Arturo Magan [5] et al. presented the utilization of Well Control Space Out technology, a cutting-edge Internet of Things platform that integrates cameras, additional sensors, machine learning, data analytics, and edge computing. What sets this work apart is its emphasis on deep learning, specifically the application of deep learning models for automated tool joint recognition. The significant outcome of this research effort is the development of a fully operational Internet of Things system. This system goes the extra mile by automatically identifying tool joints and providing advisory systems for addressing well control issues, thereby ensuring the safety and efficiency of drilling operations.

3. Dataset Description

3.1 Ground Water Dataset

In the water industry's well drilling AI integration project, comprehensive data collection is pivotal. It encompasses groundwater levels, geological and hydrogeological data, climate conditions, and more. Ensuring data quality and compliance is imperative for developing a robust AI model. This aids in optimizing well drilling processes and informed groundwater resource management decisions, prioritizing privacy and security. Figure.1 depicts the ground water dataset collected.

Well ID	S.No	STATE	DISTRICT	BLOCK	LOCATION	LATITUDE	LONGITUDE	pH	EC	CO3
ID	ID	Category	Text	Text	Text	Number	Number	Number	Number (Integer)	Number (Integer)
5101 unique values	5372 unique values	Madhya Prad... 20.3% Chhattisgarh 12.8% Other 66.9%	Raigarh 1.7% JAIPUR 1.3% Other 97%	blank 1.3% Dharmajaigarh 0.5% Other 98.1%	5136 unique values	4347 unique values	4378 unique values	3.06 – 9.25	40.0 – 26200.0	0.0 – 348
W113126092431802	1	A&N Islands	South Andaman	PORT BLAIR	Kodiaghat	11.54500	92.74583	7.74	349.0	0.0
W113616092411001	2	A&N Islands	South Andaman	PORT BLAIR	Sippighat	11.60250	92.73000	7.53	660.0	0.0
W113640092443301	3	A&N Islands	South Andaman	PORT BLAIR	Brichganj	11.62083	92.73028	7.5	270.0	0.0
W114421092422801	4	A&N Islands	South Andaman	FERRARGUNJ	Wimberleyganj	11.73250	92.70861	7.6	311.0	0.0
W113955092391801	5	A&N Islands	South Andaman	FERRARGUNJ	Ograbraj	11.66361	92.65806	7.82	2501.0	0.0
W113137092432201	6	A&N Islands	South Andaman	PORT BLAIR	Burmanala	11.55778	92.72917	7.71	264.0	0.0
W113227092433001	7	A&N Islands	South Andaman	PORT BLAIR	Calicut	11.60750	92.72556	7.79	535.0	0.0
W114348092404801	8	A&N Islands	South Andaman	FERRARGUNJ	Saitankhari	11.71167	92.67111	7.78	440.0	0.0
W123020092544201	9	A&N Islands	N&M Andaman	MAYABUNDER	Rangat	12.50611	92.90861	7.92	592.0	0.0
W123014092513301	10	A&N Islands	N&M Andaman	MAYABUNDER	Bakultala	12.50417	92.85833	7.78	456.0	0.0
W123129092581301	11	A&N Islands	N&M Andaman	MAYABUNDER	Nimbutala	12.49444	92.95417	7.78	531.0	0.0
W123030092573801	12	A&N Islands	N&M Andaman	MAYABUNDER	Padmanavapuram	12.60667	92.94917	7.29	506.0	0.0
W124940092512801	13	A&N Islands	N&M Andaman	MAYABUNDER	Rest Camp	12.83278	92.85917	7.75	431.0	0.0
W113427092395101	14	A&N Islands	South Andaman	PORT BLAIR	Gupta Para	11.55972	92.65861	7.22	147.0	0.0
W121001092455601	15	A&N Islands	N&M Andaman	RONGET	Baratang (Nilambur)	12.16833	92.76694	7.83	548.0	0.0
W121858092472001	16	A&N Islands	N&M Andaman	RONGET	Uttara	12.16139	92.79000	7.59	253.0	0.0

Figure 1. Ground Water Dataset

3.2 Rainfall Dataset

Rainfall data is pivotal for understanding precipitation's impact on agriculture, water management, and climate. It includes geographic coordinates, date/time, and rainfall amount. Rain gauges measure rainfall intensity and duration at various intervals. Data reliability is ensured with quality indicators and metadata. Analysis reveals climatic trends and extreme events, aiding decision-making. Figure.2 depicts the rainfall dataset collected.

year	month	day	tempavg	DPavg	humidity avg	SLPavg	visibilityavg	windavg	Rainfall
Number (Integer)	Number (Integer)	Number (Integer)	Number (Integer)	Number (Integer)	Number (Integer)	Number (Integer)	Category	Number (Integer)	Number
2011 – 2017	1 – 12	1 – 31	13 – 39	7 – 28	33 – 99	991 – 1018	5 39.6% 4 26.4% Other 34%	0 – 37	0.0 – 151.8
2011	2	13	21	13	60	1009	3	5	0.0
2011	6	5	29	26	83	1004	6	8	0.0
2011	6	7	29	24	79	1003	6	10	0.0
2011	6	8	27	24	85	1000	5	8	0.0
2011	6	9	26	25	91	998	4	5	0.0
2011	6	10	28	26	86	999	5	8	0.0
2011	6	11	29	25	81	1000	5	11	0.0
2011	6	12	29	26	78	1000	6	11	0.0
2011	6	13	30	26	80	999	6	6	0.0
2011	6	14	28	26	82	998	5	6	0.0
2011	6	15	30	24	74	995	5	16	0.0
2011	6	16	27	24	88	994	4	24	0.0
2011	6	17	26	23	86	995	4	23	0.0
2011	6	18	26	25	93	997	3	16	0.0
2011	6	19	28	26	81	998	5	13	0.0
2011	6	20	29	25	82	1000	6	11	0.0

Figure 2. Rainfall Dataset

3.3 Borewell Dataset

Borewell data provides vital information about groundwater access. It includes location, depth, and geological details. Water quality, yield, and usage indicate efficiency and purpose. Maintenance records and permits ensure compliance. Time series data reveals trends, aiding resource management and environmental sustainability efforts.Figure.3 depicts the bore well dataset collected.

Name of State	No. of wells Analysed	Depth to Water Level (	Depth to Water Level (	Depth to Water Level (	Depth to Water Level (	Depth to Water Level (	Depth to Water Level (	Depth to Water Level (	Depth to Water Level
Text	Number (Integer)	Number	Number	Number (Integer)	Number (Integer)	Number (Integer)	Number (Integer)	Number (Integer)	Category
Andhra Prade... 3.1% Arunachal Pra... 3.1% Other 93.8%	2 16569	0 4.21	4.6 119.02	0 4855	1 6391	0 3376	0 1213	0 480	0 59.4% 1 12.5% Other 28.1%
Andhra Pradesh	717	0	75.65	377	221	80	26	10	3
Arunachal Pradesh	3	0.8	6.47	1	1	1	0	0	0
Assam	160	0.05	15.84	94	54	10	2	0	0
Bihar	548	0.25	15	132	348	58	10	0	0
Chandigarh	12	2.74	53.64	0	3	1	4	3	1
Chhattisgarh	721	0.07	29.57	46	476	172	24	3	0
Dadra & Nagar Haveli #	17	1.03	7.6	6	10	1	0	0	0
Daman & Diu	9	1.3	4.6	1	7	1	0	0	0
Delhi	81	1.61	65.16	1	18	21	21	14	6
Goa	65	1.32	13.58	8	31	22	4	0	0
Gujarat	742	0	44.2	136	288	204	79	34	1
Haryana	315	0.1	108	24	66	52	77	66	30
Himachal Pradesh	89	0.41	29.02	18	30	21	17	3	0
Jammu & Kashmir	163	0.09	34.5	34	86	29	8	6	0
Jharkhand	279	0	14.05	61	171	44	3	0	0
Karnataka	1318	0.01	32.5	440	470	322	82	4	0
Kerala	1334	0.18	31.35	243	416	561	109	5	0
Madhya Pradesh	1296	0	54.1	155	525	470	127	18	1

Figure 3. Borewell Dataset

4. Methodology

4.1 Common Approach for All Data

Combining groundwater, rainfall, and borewell data for machine learning involves several steps. First, collect and integrate data from various sources, ensuring they can be easily merged. Clean the data by fixing errors and handling missing values. Create new features to improve analysis. Align timestamps for accurate time series analysis and aggregate data as needed.

Define the target variable, like predicting groundwater levels, and merge datasets. Split the data for testing. Select features for model training and choose appropriate algorithms. Evaluate model performance using metrics like Mean Squared Error or accuracy.

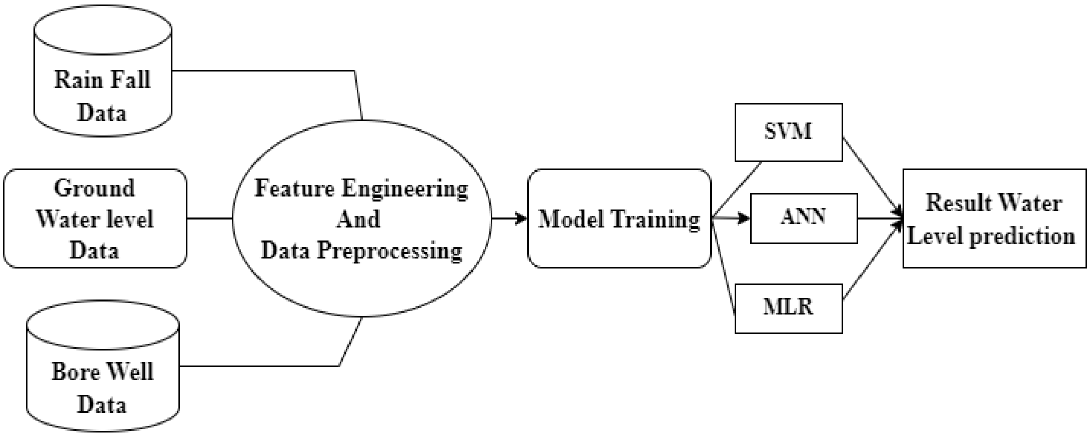

Figure 4. Block Diagram

4.2 Proposed Method

Data was initially collected from the Water Information and Management System (WIMS) an integrated database that provides all water related data. The collected data was merged, and then preprocessed checking for the missing values and detecting the outliers. The proposed method employed KNN algorithms for checking the missing values and utilized Z-score to identify and handle the outliers. The dimensionality reduction was performed applying the principal component analysis. The prepared dataset was split in the ratio of 2: 1 for the purpose of training and testing respectively. The Figure.4 shows the block diagram of the proposed. For classification purposes, several algorithms, such as ANN, MLR and SVM were deployed.

Using Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), and Support Vector Machines (SVM) for predicting water sources with borewell, groundwater, and rainfall data is crucial. MLR helps understand simple relationships, while ANN handles complex patterns. However, SVM stands out for its ability to effectively handle high-dimensional data and nonlinear relationships, making it particularly well-suited for accurate water source prediction.

When using Python and scikit-learn to predict water sources, we find that the Gaussian (RBF) kernel is the best choice. As it is good at understanding complicated relationships between data points, which is important for understanding borewell, groundwater, and rainfall data. We use a method called grid search to find the best settings for our model, and the

Gaussian kernel works well here. We can also easily use it with Geographic Information System (GIS) tools to analyze spatial data. By keeping an eye on changes in the environment, we can make sure our model stays useful over time, helping us manage water resources better.

5. Result and Discussion

The models were coded using the Python programming language and the Scikit-learn library, and Matplotlib. The code was executed using Google Colab.

Figures 5 to 7 illustrate the analysis of groundwater features, water hardness, and pH values, respectively

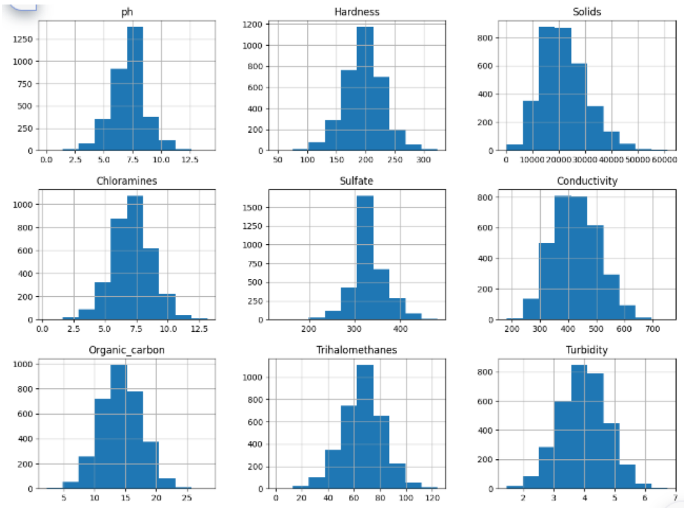

Figure 5. Analysis the features

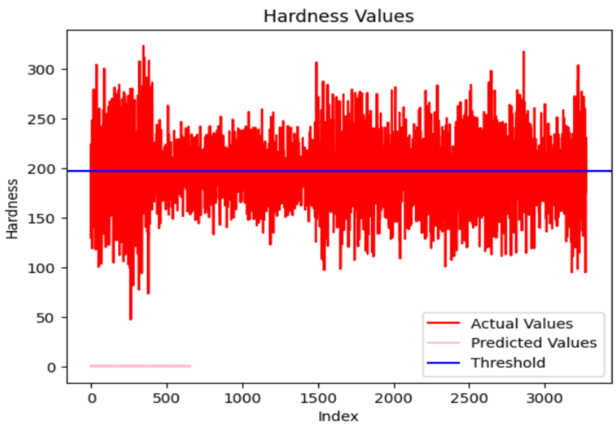

Figure 6. Hardness of Water

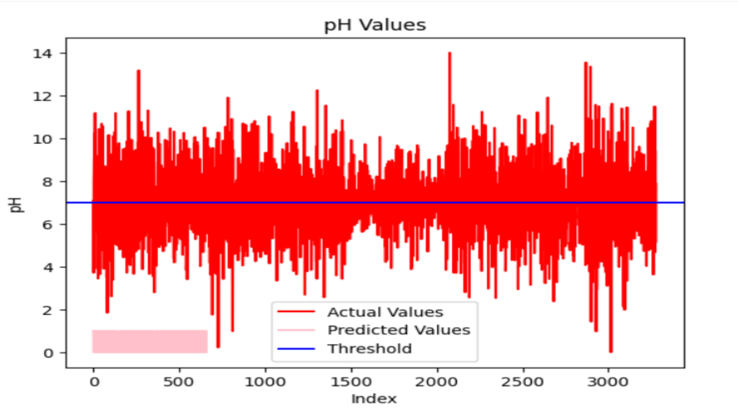

Figure 7. pH Value of Water

Figures 8 and 9 illustrate the total number of bore wells and the depth of bore wells, respectively.

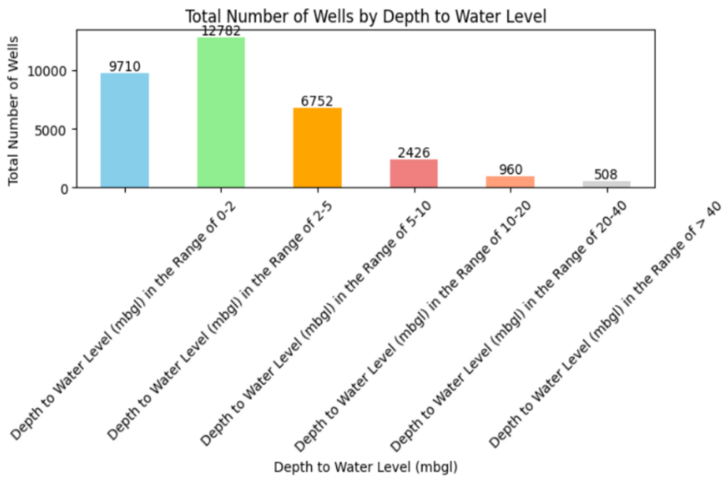

Figure 8. Total Number of Bore Wells

- The above bar chart shows the total number of wells categorized by depth to water level which is to be present in the dataset.

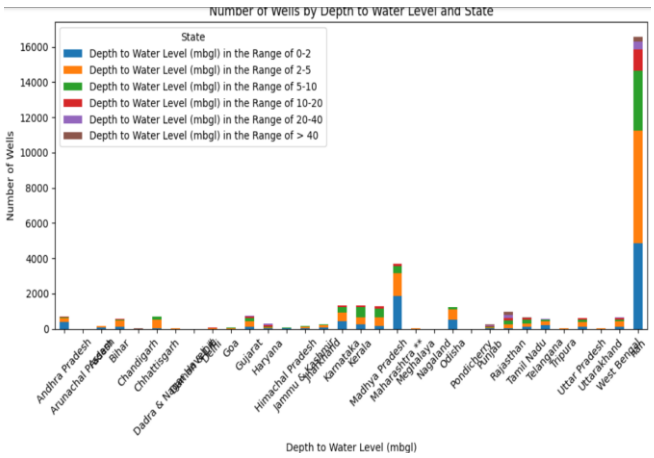

Figure 9. Depth of Bore well Water

- The above chart explains the depth of water level which is to be present in the different states in India.

Figure 10 illustrates the predicted water quantity available in select districts of Tamil Nadu

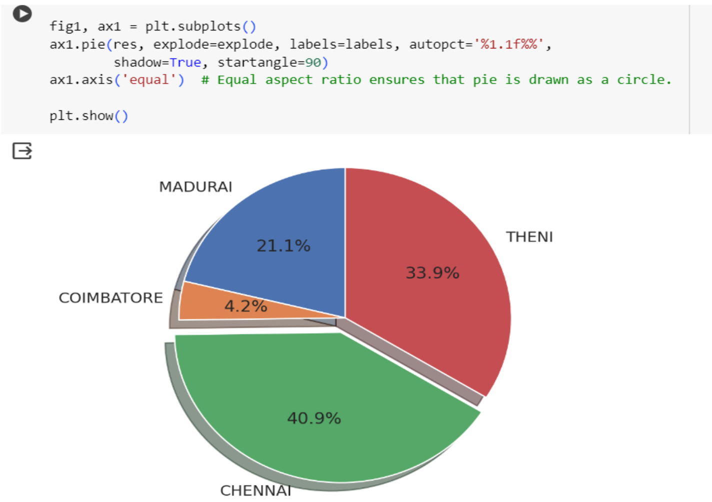

Figure 10. Water Quantity

The Figure. 11 illustrates the rainfall data features observed

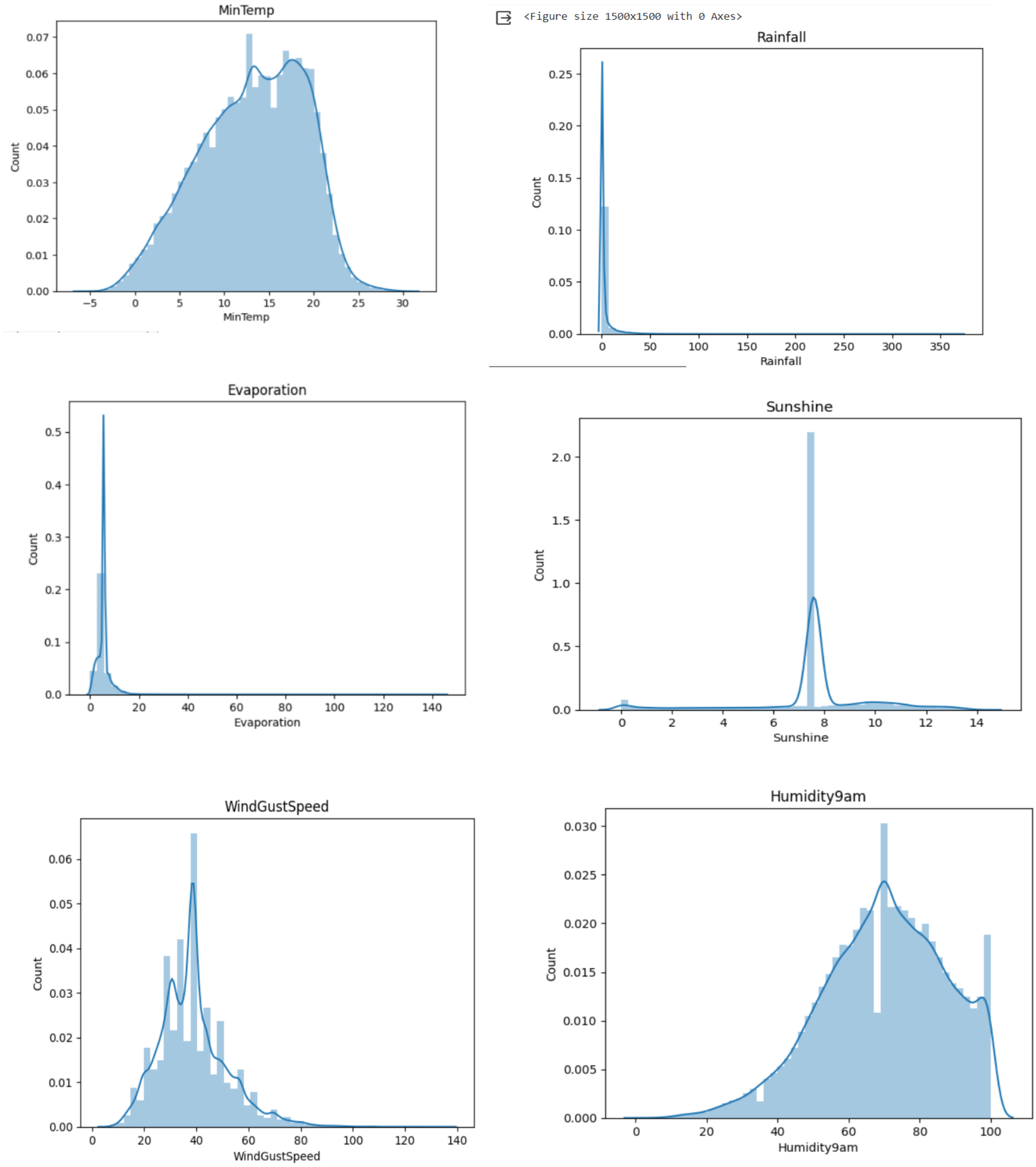

Figure 11. Rainfall Data Features

Evaporation measures water vapor returning to the atmosphere. Sunshine tracks sunlight duration. Rainfall quantifies water precipitation. MinTemp records minimum temperature. Humidity indicates air moisture content. Wind speed measures horizontal air movement. These features collectively inform weather conditions, influencing evaporation rates, precipitation patterns, and temperature variations, crucial for environmental and water resource management.

6. Conclusion

This water source prediction project represents a multifaceted and integrated approach to understanding and forecasting groundwater levels in diverse areas. By analyzing historical data encompassing groundwater levels, rainfall patterns, and borewell information, the study explores the intricate relationships between environmental factors and aquifer recharge dynamics. The incorporation of borewell data, including depth and water quality, enhances the identification of potential water sources, providing a holistic view of the groundwater landscape. The application of the Support Vector Machine (SVM) algorithm, specifically the Support Vector Regression (SVR) variant, adds a robust predictive element to the project. Through meticulous data cleaning, exploratory analysis, and machine learning, the model becomes adept at forecasting groundwater levels based on diverse features, including rainfall patterns and geographical characteristics. The spatial analysis using GIS tools further facilitates a visual understanding of the distribution of groundwater levels and rainfall across regions. The emphasis on continuous monitoring and updates ensures the ongoing relevance of the predictive model, enabling adaptation to evolving hydrogeological conditions. By considering local metrics and integrating borewell data, the project's evaluation framework becomes tailored to the specific characteristics of each region. Ultimately, this integrated approach seeks to provide valuable insights for sustainable water resource management, empowering decision-makers with informed tools for planning and managing water sources in a dynamic and diverse landscape. The project stands as a comprehensive endeavor to address the complexities of water resource prediction and management, contributing to the broader goal of ensuring sustainable and resilient water systems.

References

[1] Farouk Said Boukredera a , Mohamed Riad Youcefi b , Ahmed Hadjadj c ,Chinedu Pascal Ezenkwu d , Vahid Vaziri a , Sumeet S. Aphale : Enhancing the drilling efficiency through the application of machine learning and optimization algorithm.Volume 126, Part C, November 2023, 107035

[2] H. Yazdani, M. Fattahi, B.S. Alavi, R. Khosravi, J. Qian, M. Sarmadivaleh,and M. Rostami ,An approach for optimization of controllable drilling parameters for motorized bottom hole assembly in a specific formation,Volume20, December 2023, 101548

[3] Nautiyal, Aditi, and Amit Kumar Mishra. "Machine learning application in enhancing drilling performance." Procedia Computer Science 218 (2023): 877-886.

[4] Amadi, Kingsley, Ibiye Iyalla, Radhakrishna Prabhu, Mortadha Alsaba, and Marwa Waly. "Development of predictive optimization model for autonomous rotary drilling system using machine learning approach." Journal of petroleum exploration and production technology 13, no. 10 (2023): 2049-2062.

[5] Magana-Mora, Arturo, Michael Affleck, Mohamad Ibrahim, Greg Makowski, Hitesh Kapoor, William Contreras Otalvora, Musab A. Jamea, Isa S. Umairin, Guodong Zhan, and Chinthaka P. Gooneratne. "Well control space out: A deep-learning approach for the optimization of drilling safety operations." IEEE Access 9 (2021): 76479-76492.

[6] Kommineni, Madhuri, K. Veniha Reddy, K. Jagathi, B. Dushyanth Reddy, A. Roshini, and V. Bhavani. "Groundwater level prediction using modified linear regression." In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 1164-1168. IEEE, 2020.

[7] Chen, Hsin-Yu, Zoran Vojinovic, Weicheng Lo, and Jhe-Wei Lee. "Groundwater Level Prediction with Deep Learning Methods." Water 15, no. 17 (2023): 3118.

[8] Dr Silpaja Chandrasekar, PhD,AI-Powered Groundwater Prediction forSustainable Irrigation in Water-Scarce Regions:Oct 3 2023. https://www.azoai.com/news/20231003/AI-Powered-Groundwater-Prediction-for-Sustainable-Irrigation-in-Water-ScarceRegions.aspx#:~:text=In%20summary%2C%20this%20research%20highlights,for%20Bilate%2C%20aiding%20irrigation%20decisions.

[9] Iqbal, Naeem, Anam-Nawaz Khan, Atif Rizwan, Rashid Ahmad, Bong Wan Kim, Kwangsoo Kim, and Do-Hyeun Kim. "Groundwater level prediction model using

correlation and difference mechanisms based on boreholes data for sustainable hydraulic resource management." IEEE Access 9 (2021): 96092-96113.

[10] Lashari, Shan E. Zehra, Ali Takbiri-Borujeni, Ebrahim Fathi, Ting Sun, Reza Rahmani, and Mehdi Khazaeli. "Drilling performance monitoring and optimization: a data-driven approach." Journal of Petroleum Exploration and Production Technology 9 (2019): 2747-2756.

[11] Kombo, Omar Haji, Santhi Kumaran, Yahya H. Sheikh, Alastair Bovim, and Kayalvizhi Jayavel. "Long-term groundwater level prediction model based on hybrid KNN-RF technique." Hydrology 7, no. 3 (2020): 59.1-24

[12] M. Sakizadeh, M. M. A. Mohamed, and H. Klammler, ‘‘Trend analysis and spatial prediction of groundwater levels using time series forecasting and a novel spatio-temporal method,’’Water Resour. Manage., vol. 33, no. 4, pp. 1425–1437, Mar. 2019.

[13] Banerjee, K., V. Bali, N. Nawaz, S. Bali, S. Mathur, R. K. Mishra, and S. Rani. "A Machine-Learning Approach for Prediction of Water Contamination Using Latitude, Longitude, and Elevation. Water 2022, 14, 728." (2022). 1-20

[14] A. Nadiri, K. Naderi, R. Khatibi, and M. Gharekhani, ‘‘Modelling groundwater level variations by learning from multiple models using fuzzy logic,’’ Hydrological Sci. J., vol. 64, no. 2, pp. 210–226, Jan. 2019.

[15] S. Lee, K.-K. Lee, and H. Yoon, ‘‘Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors,’’ Hydrogeology J., vol. 27, no. 2, pp. 567–579, Mar. 2019.

[16] S. Maroufpoor, A. Fakheri-Fard, and J. Shiri, ‘‘Study of the spatial distribution of groundwater quality using soft computing and geostatistical models,’’ ISH J. Hydraulic Eng., vol. 25, no. 2, pp. 232–238, May 2019.

[17] M. Ehteram, H. Karami, S. F. Mousavi, S. Farzin, A. B. Celeste, and A.- E. Shafie, ‘‘Reservoir operation by a new evolutionary algorithm: Kidney algorithm,’’ Water Resour. Manage., vol. 32, no. 14, pp. 4681–4706, Nov. 2018.

[18] Y. Tang, C. Zang, Y. Wei, and M. Jiang, ''Data-driven modeling of groundwater level with least-square support vector machine and Spatial–temporal analysis,''Geotechnical Geological Eng., vol. 37, no. 3, pp. 1661–1670, Jun. 2019.

[19] J. Cai, J. Luo, S. Wang, and S. Yang, ''Feature selection in machine learning: A new perspective,'' Neurocomputing, vol. 300, pp. 70–79, Jul. 2018.

[20] S. Hörmann, P. Kokoszka, and G. Nisol, ''Testing for periodicity in functional time series,'' Ann. Statist., vol. 46, no. 6A, pp. 2960–2984, Dec.2018.