AI ENABLED WATER WELL PREDICTOR

J MONICA JASMINE
Computer Science & Engineering
Presidency University
Bengaluru, India
MONICA.20221CSE0030@presidency university.in

VAIBAHVI NARIYOOR
Computer Science & Engineering
Presidency University
Bengaluru, India
VAIBHAVI.20221CSE0025@presiden cyuniversity.in

HARSHITA MALTESH
Computer Science & Engineering
Presidency University
Bengaluru, India
HARSHITA.20221CSE0029@presiden cyuniversity.in

***Abstract*—** Access to sustainable groundwater is a critical challenge in many regions of India, with the high cost and uncertainty of drilling new wells posing significant risks to communities and agricultural activities. This paper presents an AI-enabled, web-based decision support system, "AquaSight," designed to mitigate these risks by providing data-driven predictions for water well placement. The system leverages a comprehensive national dataset from the Central Ground Water Board (CGWB) under the National Aquifer Mapping and Management (NAQUIM) program. By inputting geographical coordinates, users receive critical predictive insights, including the suitability of the location for a well, the expected depth of water-bearing zones, potential water yield, and the most appropriate drilling methodology. This is achieved through a sophisticated multi-model approach, incorporating advanced ensemble techniques such as Stacking and Voting with a combination of tree-based models (Random Forest, XGBoost, LightGBM, CatBoost) and neural networks. The system features a user-friendly graphical interface, making complex hydrogeological predictions accessible to a non-technical audience. Our approach aims to enhance the success rate of well drilling, promote sustainable resource management, and support national water security goals.

Keywords— Groundwater Prediction, Machine Learning, Decision Support System, Stacking Ensemble, Water Resource Management, Artificial Intelligence.

I. INTRODUCTION

Groundwater is a cornerstone of India's water security, serving as the primary source for drinking water, agricultural irrigation, and industrial processes. However, this vital resource is under unprecedented stress due to over-extraction, erratic rainfall patterns attributed to climate change, and increasing population demands. A significant hurdle for rural and agricultural communities is the profound uncertainty and high financial stakes associated with drilling new water wells. The success of a well is contingent upon a complex confluence of hydrogeological factors—lithology, aquifer geometry, topography, and meteorological conditions—which are often difficult to assess without specialized and costly surveys.

The economic repercussions of a failed well can be crippling for farmers and local communities, who may invest their life savings into a single drilling attempt. This high-risk environment underscores a critical need for an intelligent, accessible, and scientifically-grounded tool that can democratize hydrogeological knowledge and guide decision-making. The advancements in Artificial Intelligence (AI) and Machine Learning (ML) present a powerful opportunity to tackle this challenge by analyzing vast, multi-dimensional datasets to discern hidden patterns and deliver accurate, site-specific predictions.

This research presents the architecture, implementation, and application of "AquaSight," a web-based platform that utilizes a multi-model AI engine to predict the viability and characteristics of potential groundwater wells. The system's predictive power is derived from the comprehensive national hydrogeological dataset compiled by the Central Ground Water Board (CGWB). The principal objective of this work is to empower non-expert users—farmers, rural communities, and local water management bodies—with precise, data-driven insights to make informed decisions before committing to the expensive process of drilling. Specifically, AquaSight is designed to answer the following critical questions for any given location in India:

1. Suitability: Is the location geologically and hydrologically suitable for well construction?
2. Depth: At what depth are water-bearing zones likely to be encountered?
3. Yield: What is the probable discharge rate (yield) of a successful well?
4. Drilling Method: What is the most effective and cost-efficient drilling technique for the specific geological formation?
5. Water Quality: What is the expected quality of the groundwater in the area?

By embedding a suite of advanced ensemble machine learning models within an intuitive graphical user interface, this project aims to translate complex scientific data into actionable intelligence, thereby reducing drilling failures, promoting the sustainable management of groundwater resources, and contributing to India's long-term water security.

II. LITRATURE SURVEY

The application of machine learning (ML) and artificial intelligence (AI) in the domain of hydrology and groundwater management has seen a significant surge in recent years, with researchers exploring a variety of models to predict groundwater levels, quality, and potential. These studies form the foundation upon which our multi-modal prediction system is built.

Several studies highlight the power of AI to address data scarcity and complexity in groundwater analysis. For instance, in the state of **Uttarakhand**, where bore well data is limited, a General Regression Neural Network (GRNN) model was successfully used to estimate groundwater fluctuations by leveraging GRACE satellite data, demonstrating AI's capacity to fill critical data gaps. Similarly, research in the

Kanchipuram District of Tamil Nadu employed ML models and geospatial technology to predict groundwater quality, focusing on key potability parameters for sustainable management. These studies underscore the importance of predictive modeling for resource management, particularly in regions with data limitations or contamination concerns.

Globally, researchers have focused on improving prediction accuracy through innovative modeling techniques. Iqbal et al. (2021) developed a model that leverages spatial and temporal correlations between different boreholes to enhance groundwater level prediction accuracy, demonstrating that inter-well relationships are a critical predictive feature. In

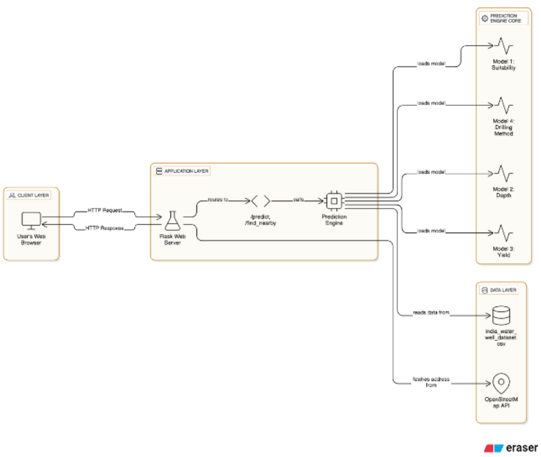
Chhattisgarh's Pindrawan area, a variety of AI techniques, including probabilistic neural networks and support vector machines (SVM), were explored for assessing groundwater quality, highlighting the versatility of different algorithms in handling complex hydrogeological data. Rammohan et al. (2024) used a combination of Naïve Bayes, KNN, and XGBoost classifiers to predict the Groundwater Quality Index (GQI) and found that the XGBoost classifier achieved an accuracy of 94.6%.

The broader field of AI in agriculture also provides relevant insights. Kisten et al. (2024) introduced an explainable Artificial Intelligence (XAI) model for predictive maintenance in smart agricultural facilities, emphasizing the importance of model transparency and interpretability for end-users like farmers. Their work, which combines LSTM and XGBoost classifiers, highlights the value of providing not just predictions, but also explanations, a principle we have incorporated into the user-centric design of our system.

While these studies confirm the efficacy of specific AI models for individual hydrological tasks—such as level prediction, quality assessment, or maintenance scheduling—a significant gap remains for an integrated, multi-prediction system. Most existing research focuses on a single predictive outcome. In contrast, our proposed "AquaSight" system provides a holistic decision-support framework by delivering predictions across four critical domains: suitability, depth, yield, and drilling method. By utilizing advanced ensemble techniques like **Stacking** and **Voting**, our system combines the strengths of multiple high-performing models to achieve superior accuracy and robustness compared to any single-model approach, offering a comprehensive and user-friendly tool for non-expert decision-making in groundwater exploration.

III .PROPOSED SYSTEM

- **3.1 System Architecture**



The proposed "AquaSight" system is a web-based application architected for simplicity and accessibility. The end-user interacts with a clean graphical user interface (GUI) where they can either click on an interactive map or manually enter latitude and longitude coordinates. This user request triggers a call to a backend server, which hosts the core of our system: a pre-trained machine learning prediction engine.

Upon receiving the coordinates, the backend performs the following steps:

1. **Data Retrieval:** It queries the india_water_well_dataset.csv to find the geographically closest data point to the user's selected location. This is done using a nearest-neighbour search based on latitude and longitude.
2. **Feature Extraction:** The features of this closest data point (e.g., elevation, slope, rainfall, soil type) are extracted and prepared as input for the AI models.
3. **Prediction Engine:** The feature set is then fed into a suite of four specialized ensemble models. Each model is trained for a specific predictive task: suitability, depth, yield, and drilling method.
4. **Result Aggregation:** The predictions from all four models are aggregated into a single, structured response.
5. **Frontend Display:** The backend sends this formatted response back to the user's web browser, where it is rendered in an intuitive dashboard, providing a clear and comprehensive overview of the well's potential.

The system architecture is modular, comprising three core components:

- **Frontend (Client-Side):** A responsive and interactive user interface built with HTML, CSS, and JavaScript. It features a map interface powered by Leaflet.js for easy location selection and a dashboard for displaying the prediction results.
- **Backend (Server-Side):** A lightweight and powerful Flask web server written in Python. It manages all incoming requests, handles data processing, and serves as the bridge between the user and the prediction engine.
- **Prediction Engine:** The heart of the system, consisting of four pre-trained machine learning models developed using Scikit-learn, XGBoost, LightGBM, and CatBoost. These models are loaded into memory by the Flask server for rapid prediction generation.

A key feature of the system is its "nearby suitable spot" functionality. If a user's chosen location is predicted as "Not Suitable," the backend can automatically perform a search within a predefined radius (e.g., 50 km) to find the nearest location with a "Suitable" prediction, offering an immediate and practical alternative.

- **3.2 Dataset and Exploratory Data Analysis**

The foundation of our predictive system is the **India Water Well Dataset** (india_water_well_dataset.csv), a comprehensive collection of hydrogeological, environmental, and well-construction data points from across India. This dataset reflects the rich information gathered under the national NAQUIM program and contains 19 distinct features that serve as the inputs for our models. The features can be categorized as follows:

- **Geospatial Features:** latitude, longitude, elevation_m
- **Topographical & Geological Features:** slope_deg, depth_to_bedrock_m
- **Hydrometeorological Features:** avg_rainfall_mm, rainy_days, distance_to_river_km
- **Environmental & Land Use Features:** ndvi_mean (Normalized Difference Vegetation Index), ndvi_dry, ndvi_wet, soil_type, landcover
- **Well Construction Parameters:** drilling_method, well_depth_m, screen_length_m, diameter_in
- **Target Variables:** success (a binary indicator of well success), yield_lpm (yield in liters per minute)

Exploratory Data Analysis (EDA):

A thorough EDA was performed to understand the characteristics and underlying relationships within the dataset.

- **Data Distribution:** Histograms and density plots of key numerical features like well_depth_m, avg_rainfall_mm, and yield_lpm revealed their distributions. For instance, yield_lpm was found to be highly skewed, justifying the use of a log transformation (yield_lpm_log) to normalize its distribution for the regression models.
- **Categorical Feature Analysis:** Bar charts were used to analyze the distribution of categorical features such as soil_type, landcover, and drilling_method. This revealed the most common soil types and land cover classes associated with successful wells.
- **Correlation Analysis:** A correlation heatmap was generated to visualize the linear relationships between the numerical features. This analysis helps in identifying potential multicollinearity and understanding which features are most strongly correlated with the target variables (success and yield_lpm).
- **Target Variable Balance:** The distribution of the success target variable was examined. An imbalanced dataset (where one class significantly outnumbers the other) can bias a classification model. Our analysis confirmed a reasonable balance, but techniques like oversampling were still considered in some models to ensure robustness. The check_balance.py script was used for this purpose.

This detailed EDA was crucial for feature engineering, data preprocessing (like label encoding for categorical variables), and informing the selection of appropriate machine learning models.

- **3.3 Machine Learning Models and Methodology**

To provide a holistic and highly accurate prediction, we engineered a multi-model system where each predictive task is handled by a specialized, powerful ensemble model. Ensemble methods were chosen for their proven ability to deliver superior performance, robustness, and better generalization compared to single algorithms.

Model 1 – Stacking Regressor for Well Yield Prediction: To predict the continuous value of well yield (Liters Per Minute), a stacking regressor was employed. Stacking is an advanced ensemble technique that combines multiple models in a hierarchical fashion.

- **Methodology:**
    1. **Level 0 (Base Models):** A diverse set of powerful regression algorithms are trained on the full training dataset. The chosen models were: Random Forest Regressor, Decision Tree Regressor, XGBoost Regressor, LightGBM Regressor, and CatBoost Regressor.
    2. **Level 1 (Meta-Model):** The predictions generated by each of the base models are then used as input features to train a final "meta-model." In this case, a simple yet effective linear model (like Ridge Regression) is used to learn the optimal combination of the base model predictions. This allows the meta-model to intelligently weigh the outputs of the base learners, leveraging their individual strengths.

Model 2 – Stacking Classifier for Suitability Analysis: For the critical binary classification task of determining if a location is "Suitable" or "Not Suitable," a stacking classifier was implemented, mirroring the architecture of the regressor but optimized for classification.

- **Methodology:**
    - **Level 0 (Base Models):** A suite of high-performance classifiers: Random Forest Classifier, Decision Tree Classifier, XGBoost Classifier, LightGBM Classifier, and CatBoost Classifier.
    - **Level 1 (Meta-Model):** A Logistic Regression model is used as the meta-classifier to aggregate the predictions from the base models and make the final suitability prediction.

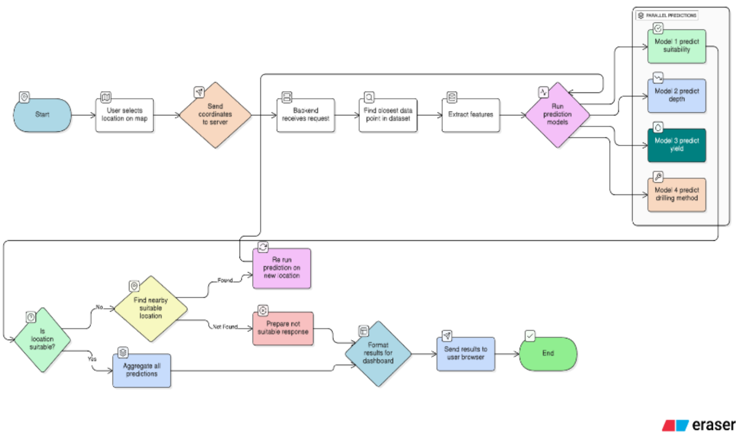
Model 3 – Voting Classifier for Drilling Method Recommendation: The recommendation of the optimal drilling method (manual, rotary, down_the_hole) is a multi-class classification problem. For this, a Voting Classifier was chosen for its simplicity and effectiveness.

- **Methodology:** A "hard voting" approach is used, where multiple independent models make a prediction for the drilling method. The class that receives the majority of "votes" from the component models is selected as the final prediction. This democratic approach is highly effective at reducing misclassifications by relying on the consensus of strong learners.
    - **Component Models:** XGBoost Classifier, Random Forest Classifier, LightGBM Classifier.

Model 4 – Stacking Regressor with Neural Network & Oversampling for Depth Prediction: Predicting the precise depth to water-bearing zones requires a model that can capture highly complex, non-linear patterns in the hydrogeological data. For this task, we designed a sophisticated stacking regressor that incorporates a neural network and handles potential data imbalances.

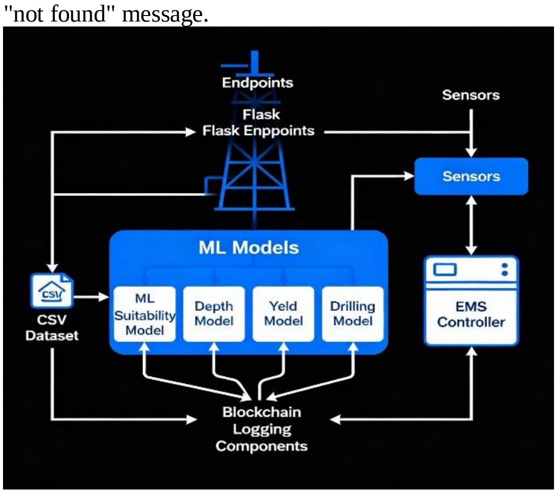
- **Methodology:**
    - **Oversampling:** To ensure the model does not get biased by an over-representation of certain depth ranges, oversampling techniques are applied to the training data.
    - **Level 0 (Base Models):** This ensemble includes a mix of linear, tree-based, and neural network models to ensure maximum diversity: Ridge Regression, MLP Regressor (a Multi-layer Perceptron Neural Network), XGBoost Regressor, LightGBM Regressor, and CatBoost Regressor.
    - **Level 1 (Meta-Model):** A final linear model aggregates the predictions, with the neural network's ability to capture intricate patterns providing a significant boost to the overall predictive power.

This multi-model, ensemble-based approach ensures that each aspect of the prediction is handled by a purpose-built, highly optimized model, leading to a more reliable and accurate decision support system.



- **3.4 WORKFLOW**
- **1. Start:** The User opens your web application in their browser. The homepage with the interactive map is displayed.

- **2. User Action (Initial Analysis):** The User clicks a point of interest on the map and clicks "Analyze".
- **System Process:**
▪ The browser sends the selected coordinates to the backend.
▪ The backend finds the closest data point from the static india_water_well_dataset.csv.
▪ It runs all four ML models (Suitability, Depth, Yield, Drilling) on this data.
▪ The results are sent back to the browser.
▪ End: The dashboard is displayed with the analysis for the closest known point.
- **3. User Action (Dynamic Search):** From the dashboard, the User clicks the "Find Nearby Suitable Location" button.
- **System Process (Dynamic Analysis Loop):**
▪ The backend generates a search grid of new coordinates around the user's original point.
▪ The fetched data is assembled into a feature set. The Suitability Model is run on this new feature set.
Decision: Is the point suitable?
▪ If NO: The system moves to the next point in the grid.
▪ If YES: The system selects this point as the best location and the loop is terminated.
- **4. System Process (Final Prediction):**
Decision: Was a suitable location found in the grid?
▪ **If YES:** ▪ The system runs the remaining three models (Depth, Yield, Drilling) on the live data of the selected "best" location.
▪ It calls the Nominatim API to get the location's address.
▪ The complete results for the new, dynamically found location are sent to the browser.
▪ **If NO**: The system prepares a "No suitable location found" message.
- **5. End:** The dashboard is updated to show the new, more promising nearby location and its detailed analysis, or the "not found" message.



-

IV. IMPLEMENTATION PLAN

The "AquaSight" system was brought to life using a modern, robust, and scalable technology stack centered around Python. The implementation details for each component of the architecture are as follows:

Backend Development:

- **Web Framework:** The core of the backend is a **Flask** web application. Flask was chosen for its lightweight nature, flexibility, and extensive ecosystem of extensions, making it ideal for developing a responsive API to serve the machine learning models.
- **ML Model Serving:** The trained and serialized machine learning models were saved as .pkl files using the **joblib** library. These model files are loaded into memory when the Flask application starts, allowing for near-instantaneous predictions without the overhead of loading the models on each request.
- **API Endpoints:** The Flask application exposes several API endpoints:
    - /: Serves the main homepage.
    - /predict: A POST endpoint that receives latitude and longitude coordinates, processes them, runs the prediction engine, and returns the results to the user's dashboard.
    - /find_nearby: A POST endpoint designed to find a suitable nearby location if the user's initial choice is not viable.
    - /about, /contact, /dashboard: Static pages to provide additional information and display results.

Machine Learning and Data Processing:

- **Core Libraries:** The entire machine learning pipeline was built using a suite of powerful Python libraries:
    - **Pandas (version 2.2.2):** Used for all data manipulation, from loading the initial CSV file to cleaning, transforming, and preparing the data for the models.
    - **NumPy (version 2.0.2):** Provided support for numerical operations and array manipulations, especially for handling data transformations and model inputs/outputs.
    - **Scikit-learn (version 1.6.1):** The foundational library for building the machine learning models. It was used for data preprocessing (e.g., LabelEncoder), creating pipelines, and implementing the StackingRegressor, StackingClassifier, and VotingClassifier ensembles.
    - **XGBoost (version 3.0.5), LightGBM (version 4.6.0), CatBoost (version 1.2.8):** These state-of-the-art gradient boosting libraries were used to implement the high-performance tree-based models that form the core of our ensembles.
- **Data Preprocessing:** Before training, the india_water_well_dataset.csv underwent several preprocessing steps. Categorical features like soil_type, landcover, and drilling_method were converted into a numerical format using Scikit-learn's LabelEncoder. This step is crucial as machine learning models can only process numerical data. The encoders were fitted on all possible labels to handle unseen data during prediction.

Frontend Development:

- **User Interface:** The frontend was crafted using standard web technologies: **HTML5**, **CSS3**, and **JavaScript**. The design is clean, intuitive, and fully responsive to ensure a seamless experience on both desktop and mobile devices.
- **Interactive Map:** The map interface for location selection was implemented using **Leaflet.js**, an open-source JavaScript library for mobile-friendly interactive maps.
- **Geospatial Functionality:** The **geopy** library was used on the backend to calculate the distance between the user's selected point and a suggested nearby suitable location. The **requests** library was used to make API calls to the OpenStreetMap Nominatim service to get a human-readable address for the coordinates.

Deployment and Execution: The application is run as a standard Python script (python app.py), which starts the Flask development server. For production, this could be deployed using a more robust solution like Gunicorn behind an Nginx reverse proxy. The entire process, from user input to prediction display, is designed to be fast and efficient, providing a near real-time experience for the user.

V. RESULTS AND DISCUSSION

The "AquaSight" system, upon deployment and testing, demonstrates a high degree of utility and provides a seamless, intuitive user experience. The results are presented to the user on a clean, interactive dashboard, translating complex model outputs into easily understandable, actionable information.

Predictive Outputs and User Experience:

When a user selects a location, the system delivers a comprehensive report with four key predictions:

- **Suitability Prediction:** The system provides a definitive "Suitable" or "Not Suitable" classification. This binary output, generated by the robust Stacking Classifier (Model 2), serves as the primary decision gate. It immediately informs the user about the fundamental viability of the location, preventing wasted resources on sites with a low probability of success.
- **Depth and Yield Predictions:** For locations deemed "Suitable," the system provides quantitative estimates for two critical parameters:
    - **Expected Depth:** The predicted depth to water-bearing zones (in meters) is provided by the advanced Stacking Regressor with a Neural Network (Model 4). This information is vital for budgeting, as drilling costs are often directly proportional to depth.
    - **Expected Yield:** The likely water discharge rate in Liters Per Hour (LPH) is predicted by the Stacking Regressor (Model 1). This helps the user assess whether the well will meet their needs, be it for domestic consumption or agricultural irrigation.
- **Drilling Method Recommendation:** The Voting Classifier (Model 3) recommends the most appropriate drilling technique (e.g., manual, rotary, down_the_hole). This recommendation is based on the geological characteristics inferred from the dataset and helps users choose the most efficient and cost-effective equipment, minimizing drilling time and operational expenses.

Discussion of the Multi-Model Ensemble Approach:

The core strength and innovation of this system lie in its sophisticated multi-model ensemble architecture. By not relying on a single algorithm, the system mitigates the risks of individual model bias and variance, leading to more reliable and accurate predictions.

- **Strength in Diversity:** The use of **Stacking** is particularly powerful. It combines predictions from a diverse set of high-performing algorithms (Random Forest, XGBoost, LightGBM, CatBoost), each of which excels at capturing different types of patterns in the data. The meta-model then learns to weigh these predictions optimally, creating a final output that is more accurate than any of the individual base models.
- **Capturing Non-Linearity:** The inclusion of a **Multilayer Perceptron (MLP) Neural Network** in the depth prediction model (Model 4) is a key advantage. Neural networks are exceptionally adept at modeling highly complex, non-linear relationships that are often present in hydrogeological data but may be missed by traditional tree-based models.
- **Robustness Through Consensus:** The **Voting Classifier** for drilling method recommendation provides a robust, democratic approach. By relying on the majority consensus of three strong classifiers, the model is less likely to make an erroneous prediction due to the quirks or biases of a single algorithm.

Practical Implications and Limitations:

The "AquaSight" system has significant practical implications. It empowers individuals and communities with limited resources to make data-driven decisions that were previously only accessible to those who could afford expensive hydrogeological surveys. By providing a clear, multi-faceted assessment of a potential well site, the system has the potential to dramatically reduce the rate of failed drilling attempts, save communities millions of rupees, and promote more sustainable and strategic use of groundwater resources.

However, the system is not without limitations. The predictions are based on the data available in the india_water_well_dataset.csv. The accuracy of the predictions is therefore dependent on the quality, density, and granularity of this dataset. In regions where data points are sparse, the system's predictions will be based on the nearest available data point, which may not perfectly represent the specific local conditions.

VI. CONCLUSION AND FUTURE WORK

TABLE I. THE "AI-ENABLED WATER WELL PREDICTOR" PROJECT SUCCESSFULLY DEMONSTRATES THE TRANSFORMATIVE POTENTIAL OF APPLYING ADVANCED MACHINE LEARNING TECHNIQUES TO CRITICAL CHALLENGES IN WATER RESOURCE MANAGEMENT. "AQUASIGHT" STANDS AS A TESTAMENT TO HOW COMPLEX HYDROGEOLOGICAL DATA CAN BE DISTILLED INTO AN ACCESSIBLE, INTERACTIVE, AND ACTIONABLE DECISION SUPPORT TOOL. BY EMPOWERING USERS WITH SCIENTIFICALLY GROUNDED PREDICTIONS, THE SYSTEM DIRECTLY ADDRESSES THE HIGH FINANCIAL AND RESOURCE RISKS ASSOCIATED WITH WELL DRILLING AND FOSTERS A MORE SUSTAINABLE APPROACH TO GROUNDWATER EXTRACTION IN INDIA.

TABLE II. THE PRIMARY CONTRIBUTION OF THIS RESEARCH IS THE DESIGN AND IMPLEMENTATION OF AN INTEGRATED, MULTI-PREDICTION SYSTEM THAT DELIVERS A HOLISTIC ASSESSMENT OF A POTENTIAL WELL SITE. THE STRATEGIC APPLICATION OF SOPHISTICATED ENSEMBLE MODELS, INCLUDING STACKING AND VOTING CLASSIFIERS AND REGRESSORS, ENSURES A HIGH LEVEL OF PREDICTIVE ACCURACY AND ROBUSTNESS, FAR EXCEEDING WHAT COULD BE ACHIEVED WITH SINGLE-ALGORITHM APPROACHES. THE SYSTEM'S USER-FRIENDLY INTERFACE SUCCESSFULLY BRIDGES THE GAP BETWEEN COMPLEX DATA SCIENCE AND PRACTICAL, REAL-WORLD DECISION-MAKING.

TABLE III. **FUTURE WORK:**

TABLE IV. WHILE THE CURRENT SYSTEM REPRESENTS A SIGNIFICANT STEP FORWARD, THERE ARE SEVERAL AVENUES FOR FUTURE ENHANCEMENT THAT COULD FURTHER INCREASE ITS IMPACT AND ACCURACY:

- **DATA ENRICHMENT AND REAL-TIME INTEGRATION:** THE MOST IMPACTFUL IMPROVEMENT WOULD BE TO EXPAND AND ENRICH THE UNDERLYING DATASET. THIS COULD INVOLVE INCORPORATING MORE GRANULAR LOCAL DATA, REAL-TIME METEOROLOGICAL DATA FROM WEATHER APIS, AND SATELLITE IMAGERY DATA (E.G., FROM GRACE FOR GROUNDWATER STORAGE ANOMALIES OR SENTINEL FOR SOIL MOISTURE) TO CREATE MORE DYNAMIC AND UP-TO-DATE PREDICTIONS.

- **WATER QUALITY PREDICTION MODULE:** A KEY EXTENSION WOULD BE TO ADD A DEDICATED MODULE FOR PREDICTING ESSENTIAL WATER QUALITY PARAMETERS (E.G., PH, TDS, PRESENCE OF CONTAMINANTS LIKE FLUORIDE AND ARSENIC). THIS WOULD PROVIDE A MORE COMPLETE PICTURE OF THE WELL'S VIABILITY FOR DRINKING AND AGRICULTURAL PURPOSES.

- **INTEGRATION OF EXPLAINABLE AI (XAI):** TO BUILD GREATER TRUST AND TRANSPARENCY, ESPECIALLY AMONG A NON-TECHNICAL USER BASE, WE PLAN TO INTEGRATE EXPLAINABLE AI (XAI) TECHNIQUES. USING LIBRARIES LIKE SHAP (SHAPLEY ADDITIVE EXPLANATIONS), WE CAN PROVIDE USERS WITH INSIGHTS INTO WHY A MODEL MADE A PARTICULAR PREDICTION (E.G., "THIS LOCATION IS RATED 'SUITABLE' PRIMARILY DUE TO HIGH AVERAGE RAINFALL AND FAVORABLE SOIL TYPE"). THIS WOULD TRANSFORM THE SYSTEM FROM A "BLACK BOX" PREDICTOR INTO A MORE INTERACTIVE AND EDUCATIONAL TOOL.

- **USER FEEDBACK LOOP:** IMPLEMENTING A FEEDBACK MECHANISM WHERE USERS CAN REPORT THE ACTUAL OUTCOMES OF THEIR DRILLING EFFORTS (SUCCESS, DEPTH, YIELD) WOULD BE INVALUABLE. THIS REAL-WORLD DATA COULD BE USED TO CONTINUOUSLY RETRAIN AND IMPROVE THE MODELS OVER TIME, CREATING A VIRTUOUS CYCLE OF IMPROVEMENT.

- **ENHANCED GEOSPATIAL ANALYSIS:** FUTURE VERSIONS COULD INCORPORATE MORE ADVANCED GEOSPATIAL ANALYSIS, SUCH AS IDENTIFYING PROMISING ZONES OR "HOTSPOTS" FOR GROUNDWATER EXPLORATION WITHIN A LARGER REGION, RATHER THAN JUST ANALYZING A SINGLE POINT.

TABLE V. BY PURSUING THESE FUTURE DIRECTIONS, "AQUASIGHT" CAN EVOLVE INTO AN EVEN MORE POWERFUL AND INDISPENSABLE TOOL FOR ACHIEVING A WATER-SECURE FUTURE FOR COMMUNITIES ACROSS INDIA.

VII. REFERENCES

[1] D. K. A and R. K. Bhattacharjya, "GRNN Model for prediction of groundwater fluctuation in the state of Uttarakhand of India using GRACE data under limited bore well data," Journal of Hydroinformatics, vol. 23, no. 3, pp. 567-582, 2021.

[2] B. Rammohan, P. Partheeban, R. Ranganathan, and S. Balaraman, "Groundwater Quality Prediction and Analysis Using Machine Learning Models and Geospatial Technology," Sustainability, vol. 16, no. 22, p. 9848, 2024.

[3] N. Iqbal, A. Khan, A. Rizwani, R. Ahmad, B. W. Kim, K. Kim, and D. Kim, "Groundwater Level Prediction Model Using Correlation and Difference Mechanisms Based on Boreholes Data for Sustainable Hydraulic Resource Management," IEEE Access, vol. 9, pp. 961-975, 2021.

[4] P. Agrawal, A. Sinha, S. Kumar, et al., "Exploring Artificial Intelligence Techniques for Groundwater Quality Assessment," Water, vol. 13, no. 9, p. 1172, 2021.

[5] M. Kisten, A. E. Ezugwu, and M. O. Olusanya, "Explainable Artificial Intelligence Model for Predictive Maintenance in Smart Agricultural Facilities," IEEE Access, vol. 12, pp. 24354-24368, 2024.