

HYBRID KNN BASED MACHINE LEARNING MODELS FOR SHORT-TERM WATER QUALITY PREDICTION

1.Dr.P.Tamilselvan

Assistant Professor

Department of Computer Science

Hindusthan College of Arts & Science

Coimbatore

2.Dr.M.Rajeshkumar

Assistant Professor

Department of Computer Science

KPR College of Arts Science and Research

Coimbatore

Abstract

Water sources are the foundation of humans's lifestyles and economic improvement, and are closely associated with health and the environment. Accurate prediction of water exceptional is the important thing to enhancing water control and pollution manipulate. In this paper, three novel hybrid machine learning models are proposed to obtain more accurate short-term water quality prediction results. The following water fine parameters were utilized to evaluate the general water exceptional in phrases of potability on this have a look at. The pH and hardness testing data Solids, chloramines, sulphate, conductivity, organic carbon, trihalomethanes, turbidity, and potability had been the parameters. To depict the water satisfactory, those parameters are used as a characteristic vector. The performance of device learning strategies such as Random forest, Naive Bayes and K- Nearest Neighbour(KNN) for predicting water pleasant components of Bhavani River placed inside the India. KNN demonstrated its energy in nonlinear modeling and achieved nicely in predicting a couple of water first-rate parameters. Experiments had been accomplished using a actual dataset containing information from various locations, in addition to a artificial dataset generated at random the usage of parameters. According to the findings, device mastering procedures are capable of correctly predicting the potability.

Keywords: Water Quality, Machine Learning, Naive Bayes, Random forest and K-Nearest Neighbour(KNN).

1.INTRODUCTION

Water is the largest useful resource of life, important for helping the existence of most existing creatures and people. Living organisms want water with sufficient exceptional to hold their lives. There are positive limits of pollutions that water species can tolerate. Exceeding those limits affects the life of those creatures and threatens their lives. Most ambient water our bodies

which include rivers, lakes, and streams have unique exceptional requirements that indicate their fine. Moreover, water specifications for other programs/usages own their requirements. For example, irrigation water must be neither too saline nor incorporate toxic materials that can be transferred to flora or soil and hence destroying the ecosystems. Water exceptional for industrial uses also calls for different houses based totally on the particular industrial strategies. Some of the less high-priced sources of clean water, together with ground and floor water are herbal water property. However, such assets can be polluted by using human/industrial activities and other herbal procedures.

Water quality analysis is a complicated subject matter due to the various factors that impact it. This idea is inextricably linked to the diverse functions for which water is used. Different desires necessitate one of kind standards. There is a lot of take a look at being finished on water first-class prediction. Water first-class is normally determined by means of a set of bodily and chemical parameters that are closely related to the water's meant utilization. The suitable and unacceptable values for each variable need to then be mounted. Water that meets the pre-decided parameters for a particular application is considered appropriate for that application. If the water does now not satisfy these requirements, it must be dealt with before it is able to be used. Water first-rate can be assessed the use of a spread of physical and chemical residences. As an end result, analyzing the behaviour of each man or woman variable independently isn't feasible in practice to appropriately describe water first-class on a spatial or temporal basis.

Water exceptional can be classified into four kinds—potable water, palatable water, infected water, and inflamed water. Potable water is safe to drink, high-quality to flavor, and usable for home functions. Palatable water is esthetically pleasing; it considers the presence of chemical substances that do not reason a danger to human fitness. Contaminated water is that water containing unwanted physical, chemical, biological, or radiological materials, and it is not worthy for drinking or home use. Infected water is infected with pathogenic organism. The parameters of water pleasant are proven in Figure 1.1.

Physical parameters	Chemical parameters	Biological parameters
Turbidity, Temperature, Color, Taste and odor, Solids, Electrical conductivity	pH, Acidity, Alkalinity, Chloride, Chlorine residual, Sulfate, Nitrogen, Fluoride, Iron and manganese, Copper and zinc, Hardness, Dissolved oxygen, Biochemical oxygen demand, Chemical oxygen demand, Toxic inorganic substances, Radioactive substances	Bacteria, Algae, Viruses, Protozoa

Figure 1.1 Types of water quality parameters

2. LITERATURE REVIEW

Water prediction using artificial intelligence strategies which includes artificial neural network (ANN), group approach of information coping with (GMDH) and assist vector machine (SVM) for predicting water first-class additives of Tireh River positioned within the southwest of Iran. To broaden the ANN and SVM, special sorts of transfer and kernel functions had been examined, respectively. Reviewing the consequences of ANN and SVM indicated that both fashions have appropriate performance for predicting water excellent additives. The lowest accuracy of fashions was associated with GMDH [1].

The exceptional of water has deteriorated significantly due to pollution and plenty of different problems. As a result of this, there is a want for a version that may make accurate projections about water excellent. The system mastering overall performance of tactics due to XGB, RF, SVC, ADA, and Decision Trees in predicting the components of a water satisfactory dataset. After obtaining accuracy, we ought to improve our model using hyper parameter tuning to attain the favored accuracy [2].

The complete weighting method that combines the entropy weighting technique and the Pearson correlation coefficient method confirmed effectiveness in choosing a characteristic set for water excellent prediction, enhancing the predictive performance of the fashions. Through comparative research, the LSTM version emerged as the pinnacle-performing version for water best prediction, as it should be forecasting variations in distinctive water excellent variables in a stable manner. However, in addition studies and programs are necessary to explore optimized

feature-selection methods, enhance device gaining knowledge of fashions, and enhance the accuracy and reliability of water great prediction [3].

The importance of water for the human frame and also Quality of the water is very an awful lot essential due to the fact if we drink water without knowing that its miles secure for ingesting we may want to get ill. There are masses of water-borne diseases like Cholera, Typhoid, Giardia, E.Coli, Hepatitis A, and so forth. These sorts of sicknesses manifest if we drink non-drinkable water. So understanding the excellent of the water is the maximum vital factor. But the primary trouble lies right here. An opportunity technique the use of synthetic intelligence to predict water quality. This technique makes use of a big and without problems to be had water exceptional index which is set by way of the World Health Organization [4].

The hyper parameter tuning became finished in SVM for improving the accuracy the usage of Grid Search technique. Artificial and machine learning models can be used for the prediction of the equal to make human life healthier and simpler. In this investigation is a specific Machine Learning algorithm on Water Quality Prediction dataset has been carried out. The validation on take a look at dataset provides 68% accuracy [5].

Safe consuming-water gets right of entry to be vital to health, a basic human right and an aspect of powerful policy for fitness safety. It is vital as a fitness and development trouble at a national, regional and local stage. Automatic Machine Learning facilities offer gadget getting to know with push of a button, or, on a minimal stage, make sure to retain algorithm execution, statistics pipelines, and code, usually, are kept from sight and are predicted to be the stepping stone for normalizing AI. The overall performance of Random wooded area class, K-nearest neighbor in supervised getting to know and Tpot in AutoML were evaluated to be expecting the water excellent. Python is expecting more accuracy degree than the AI framework [6].

Supervised device gaining knowledge of requires schooling machines by means of offering giant datasets and parameters. This venture uses various type algorithms such as Random Forest, Decision Tree and Naive Bayes. These algorithms are used to locate the patterns and understand them and then the usage of specific styles we find accuracy and better overall

performance. Roodeplaat Dam water first-class class was modeled using five system mastering algorithms run the use of Anaconda Jupyter Notebook/ Python software program. The water exceptional of the dam was discovered to have appropriate, medium, and horrific instructions. The prediction accuracy of all the system studying algorithms besides linear regression becomes superb. However, the accuracy of K-nearest associates model changed into the nice of all at all test sizes [7][8].

Predicting water pleasant has turn out to be very essential in controlling water pollutants. In these artwork, advanced artificial intelligence (AI) algorithms are advanced to expect water excellent index (WQI) and water excellent kind (WQC). For the WQI prediction, synthetic neural community models, specifically nonlinear autoregressive neural community (NARNET) and long brief-term reminiscence (LSTM) deep studying algorithm had been evolved. In addition, three machine learning algorithms, namely, aid vector device (SVM), K-nearest neighbor (K-NN), and Naive Bayes, was used for the WQC forecasting. After analyzing the robustness and efficiency of the proposed version for predicting the WQI [9].

Portability determines the nice of water that is one of the most critical sources for lifestyles. Traditionally, testing water first-class required a costly and time-ingesting lab analysis. This take a look at appeared into an opportunity gadget learning method for predicting water excellent using just a few simple water quality standards. To estimate, a set of consultant supervised gadget getting to know algorithms was used. It would come across water of bad satisfactory before it became released for consumption and notify the correct authorities. It will hopefully reduce the number of folks who drink low-first-rate water, reducing the risk of illnesses like typhoid and diarrhea. In this situation, the use of prescriptive analysis primarily based on projected values would bring about destiny abilities to assist choice and policy makers [10].

The hyper parameters nonetheless call for to be carried out by using augmenting the AI version with other advanced meta-heuristic optimization algorithms. Overall, this observes integrates several analytical and modeling strategies that might become beneficial to institutions which can be dedicated to river basin management inside Malaysia. The findings of this

exploration in a sizeable scope IoT-primarily based net staring at framework using simply the sensors of the vital boundaries. The attempted calculations could foresee the water exceptional right away dependent on the ongoing statistics took care of from the IoT framework. It might come across low-fine water till it changed into dispensed for use and warn involved specialists. It will preferably result in diminishing of individuals burning-thru low pleasant water and thusly de-enhance nerve racking diseases like typhoid and the runs. In such way, the usage of a prescriptive exam from the ordinary qualities could activate destiny workplaces to help choices and policymakers [11][12].

The effective and realistic machine for comparing the character of groundwater and its advancement in the usage of physicochemical barriers are as it should be upheld with the aid of GIS methods, multivariate displaying, and machine studying. Analytical Hierarchy Process is an important geological approach for identifying ground-level elevation stages, and they according to-formed higher boundary conditions and calibration stages. Overall, numerical outcomes of predicting the water first-rate the use of WQI sensor generation, IoT, device gaining knowledge of, deep learning, and GIS technologies are encouraging. These technologies may be utilized to create particular water best prediction models, consistent with several studies conducted within the beyond [13].

The ML classifiers used within the examine were aid vector gadget, random forest, logistic regression, CATBoost, XGBoost, selection tree, and multi-layer perceptron. Performances metrics, which include precision, sensitivity, F1 rating, and accuracy, have been envisioned the usage of the confusion matrix of each classifier. Moreover, precision-don't forget, ROC, and average precision curves were studied. The analysis also can be accelerated into multiple guidelines, consisting of hardware implementation, Internet of Things (IoT), biomedical studies, and other fields [14].

Two fashions are used to be expecting water temperature, dissolved oxygen, pH cost, particular conductance, turbidity, and FDOM, and the prediction effects are as compared with other benchmark models. Besides, this paper takes SDE as an evaluation index to research the stableness of the model. CEEMDAN-RF plays nice in the prediction of temperature, dissolved oxygen and precise conductance [15].

Our take a look at demonstrates the application of using a model stacking technique for predictive modelling of seaside water quality. Since model stacking averages out noise from its

base models, it is theoretically more promising than character models in generating predictions with more accuracy and robustness. Minimizing overfitting is a unique problem for stacking a library of models, for this reason strategies such as move validation, regularization and bagging ought to be used to assess the generalization overall performance of a stacking version [16].

The deep studying prediction technique based totally on multisource information fusion proposed on this observe can expect water quality efficiently and speedy thru indicators which are clean to acquire and measure. It can remedy the issues of lengthy processing instances and high costs associated with pattern series and chemical evaluation. Compared with the linear method and conventional gadget mastering algorithms, deep gaining knowledge of algorithms can reap higher predictive overall performance with decrease RMSE [17].

3. METHODOLOGY

Classification predictive modelling is the venture of approximating a mapping feature from enters variables to discrete output variables. Before using ML fashions to research the information, earlier moves have been taken to put together the facts as enter to the model. This blanketed dividing the data into schooling and trying out sets to train the three fashions and examine the overall performance. In addition, the dataset changed into wiped clean by deleting wrong values and replacing empty cells with the median of the dataset's input variables. Different ML fashions were then used to be expecting Water Quality Classification. The recommended method for the modern have a look at is proven in Figure3.1.

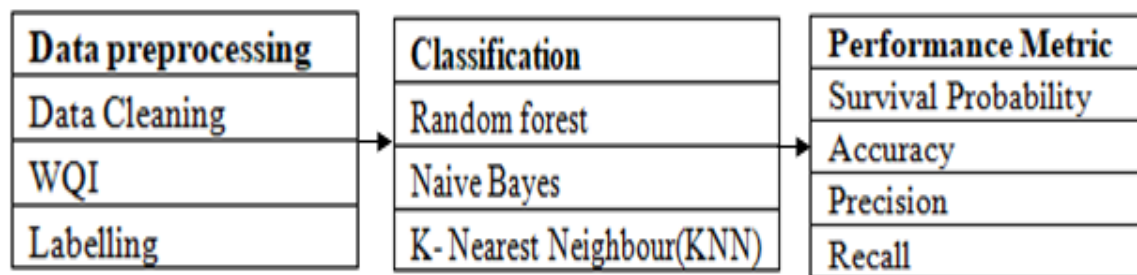


Figure 3.1 Methodology of the proposed system

3.1 Dataset

This ingesting water quality records become gathered from bhavani between 2023 and 2024. An overall of 1756 samples had been accumulated and analyzed for Physical parameters, Chemical parameters and Biological parameters. More statistics about the dataset can be

obtained from the link cited in the facts availability section. In addition, the records had been subjected to a preprocessing step to improve data excellent. Data preprocessing encompasses all steps required to keep away from inappropriate entries by cleansing and labeling the facts.

3.2 Classifiers

Naïve Bayes

The Naïve Bayes (NB) classifier belongs to the probabilistic family of classifiers based totally on Bayes' Theory. The most important characteristic of this classifier is the belief that all variables are conditionally independent which the motive for calling it 'Naïve' is. Because of that, the parameters may be discovered one by one, extra definitely and quicker. The NB classifier calculates a posterior opportunity of the elegancex) following the Bayes' Rule as in Equation

$$P(C|X) = \frac{P(X|C)*P(C)}{P(X)}$$

Where P(C) is the prior class' probability, P(X|C) is the probability of predictor given class, and P(X) is the prior probability of predictor.

Random forest

A random forest is a Meta estimator that fits some of choice tree classifiers on numerous sub-samples of the dataset and makes use of averaging to enhance the predictive accuracy and control over-becoming. Trees within the woodland use the first-class split strategy, i.e. Equal to passing splitter="nice" to the underlying Decision Tree Regressor. The sub-sample length is managed with the max_samples parameter if bootstrap=True (default), in any other case the complete dataset is used to build every tree. Classification are conducted by way of the random wooded area technique that operates by way of building an ensemble of choice timber in training by way of swapping and changing the covariates to enhance the prediction performance. The weighted common of tree outputs is used to reap the intention, as RF has many ability functions generated from distinctive nodes. This version calls for many educated timber and positive amount of the variable in each tree. The Random forest classifier is a reliable algorithm that surpasses numerous different type algorithms in phrases of accuracy. Additionally, the RF prediction procedure steps to start with take the take a look at characteristics and predict the final results the use of the regulations of each randomly generated selection tree, then store the expected end result; after that, evaluate the votes for each ability target. Take the Random

woodland algorithm final forecast as the highest chosen predicted objective. The parameters of Randomforest are $n_estimators=100$, $Criterion='gini'$, $max_depth=None$, $min_samples_split=2$, $min_samples_leaf=1$, $min_weight_fraction_leaf=0$, $max_features='auto'$, $max_leaf_nodes=None$, $min_impurity_decrease=0$, $Bootstrap=True$, $oob_score=False$, $n_jobs=None$, $random_state=None$, $Verbose=0$, $warm_start=False$, $class_weight=None$, $ccp_alpha=0$, $max_samples=None$.

K- Nearest Neighbour

The machine learning models help researchers regulate water quality and address irrigation water quality issues. A groundwater quality prediction using machine learning techniques, and the initial step involved in the application of regression approach to predict groundwater images and the process adopted are shown in Figure 3.2.

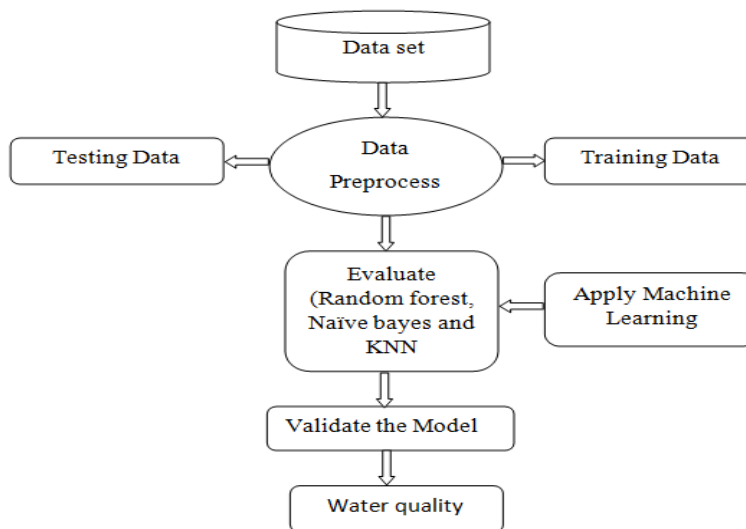


Figure 3.2.1 Implementation of machine learning models

A feed-ahead network is the KNN that includes many layers of neurons, in which one neuron's output, is propagated to the other neuron's input that is in the subsequent layer. In KNN, the enter layer's nodes only propagate the enter values of the first hidden layer's nodes. In the hidden layers, each node's enter–output dating may be provided. The proposed gadget is intended to determine potability. It is split into stages, one for schooling and the alternative for testing. The following procedures are done in both sections. Data on training pH and hardness checking out information Solids, chloramines, sulphate, conductivity, natural carbon, trihalomethanes, turbidity, and potability are all terms that can be used to testing. The facts set

changed into selected as follows: The collection of critical parameters that have an effect on water high-quality, identity of the range of information samples, and definition of the elegance labels for each statistics pattern gift inside the information are all factors that move into deciding on the water best data set, which is a prerequisite to model production. Ten indicator parameters make up the facts sets used in this have a look at. PH value and hardness are examples of those elements. Solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, turbidity, and potability are all terms that may be used to describe the properties of a substance. The proposed technique, but, isn't confined with the aid of the quantity of parameters or the choice of parameters.

The dataset is separated into k-disjointed units of identical size, every with kind of the identical class distribution, using this method. This division's subsets are utilized as the check set in turn, with the closing subsets serving because the schooling set. These are Random forest and Naive Bayes methods. In terms of the underlying relational structure among the indicator parameters and the class label, every of those strategies takes a one-of-a-kind approach. As an end result, each approach's overall performance for the equal records set is in all likelihood to differ. Validating the overall performance of various classifiers on unknown records set: Data mining offers several metrics for validating the performance of different classifiers on unknown facts set. A repeated cross-validation procedure inside the Matlab caret package deal was used to create the studying and trying out surroundings. The following method becomes used to use the type algorithm:

1. The statistics set become split into elements: training and trying out
2. The training set turned into subjected to repeated cross-validation, with the variety of iterations constant to Classifiers were educated on this way.
3. The model's most beneficial parameter configuration become decided on, ensuing in the most accuracy.
4. The version changed into scrutinized.

To estimate river water nice class, three statistics mining methods were used: Random forest, Naive Bayes and K- Nearest Neighbour(KNN). These methods are both parametric and nonparametric classifiers, and their goal is to develop a function that maps enter variables to output variables from an education dataset. Because the characteristic's shape is unknown, exceptional algorithms make different assumptions approximately the function's shape and the

way schooling statistics is learned to supply the output. The parametric learning classifier makes more confident assumptions approximately the information. If the assumptions for any records set are true, those classifiers will make rectification judgments. However, if the assumptions are incorrect, the equal classifier performs poorly. In order to examine class tasks, those classifiers do not depend upon the amount of the pattern facts set; as a substitute, their operating concepts are their assumptions. This classifier is susceptible to prediction mistakes together with bias, similarly to its parametric man or woman. To make type choices, it in reality calculates the distance among traits. Because each set of rules' mode of operation differs, a comparison of all of them is vital to decide which one is higher at approximating the underlying feature for the identical education and testing water pleasant datasets.

4. EXPERIMENTAL RESULT AND DISCUSSION

The most usually metric used to determine the performance of classifier is accuracy. Since the accuracy is inappropriate when records is imbalanced, we've got used every other metrics to examine the overall performance. The fashionable approach for evaluating classifier on imbalanced class is Receiver Operating Characteristic. It indicates that KNN has regular accuracy even though the information has been randomized 34 instances. Random Forest Tree can classify the result higher than the alternative classifiers. Recall degree shows how frequently a effective class example in the dataset changed into anticipated as a high-quality elegance instance by means of the classifier. Precision measure suggests how frequently an instance that turned into anticipated as fantastic that is genuinely positive. Here we evaluate current and proposed algorithms as: 1. NavieBayes, 2. Random Forest, 3. KNN.

Survival Probability:In this section, we compare memory usage for each set of rules with the same datasets as the runtime checks. Our algorithm, it ensures Survival Probability as precise as that of the ultra-modern algorithm. Moreover, our set of rules affords the most wonderful results in lots of instances.

No of Samples	Naive Bayes	Random Forest	KNN
100	0.72	0.77	0.82
200	0.73	0.78	0.83
300	0.75	0.79	0.84
400	0.77	0.80	0.86
500	0.78	0.81	0.87

Table 4.1 Survival Probability Results

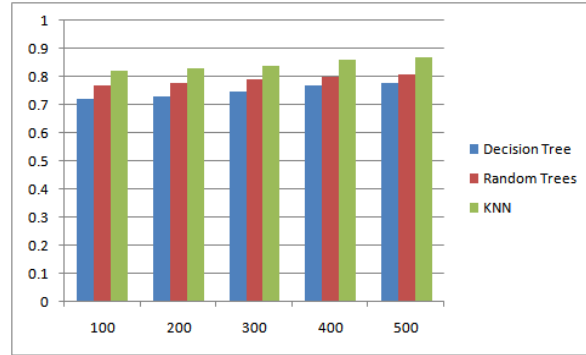


Figure 4.1 Survival Probability

Accuracy (%):We can examine that our proposed outperforms the others in nearly all of the cases. Our proposed linear shape to its bushes instead of the previous tree form as a way to limit get entry to instances to look nodes. As an end result, its benefits have an effective impact on lowering runtime in complete experiments.

No of Samples	Naive Bayes	Random Forest	KNN
100	70.7	72.6	85.9
200	70.8	73.1	86.3
300	71.2	73.9	87.7
400	71.5	74.2	88.6
500	71.9	74.9	89.4

Table 4.2 Accuracy Table

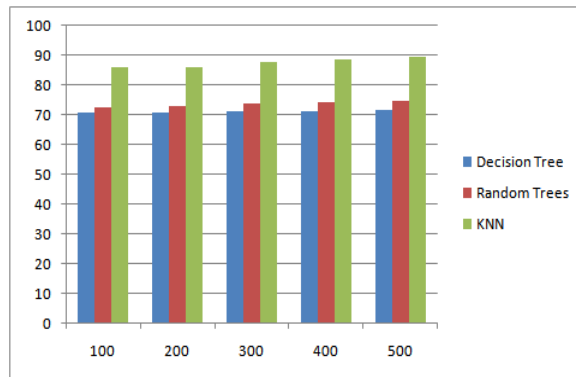


Figure 4.2 Accuracy

Precision(%):Proposed set of rules indicates the best precision at the same time as the others have especially negative performance, which shows that our scheme can keep those increasing attributes more efficiently than the alternative structures of the competitor algorithms. Through the above experimental consequences, we recognize that the proposed set of rules outperforms the others with admire to increasing transactions and gadgets in phrases of scalability in addition to runtime and memory usage for the actual datasets.

No of Samples	Naive Bayes	Random Forest	KNN
100	71.6	72.4	86.4
200	72.7	73.3	87.1
300	73.4	73.7	88.5
400	74.3	74.2	89.2
500	74.9	75.6	90.1

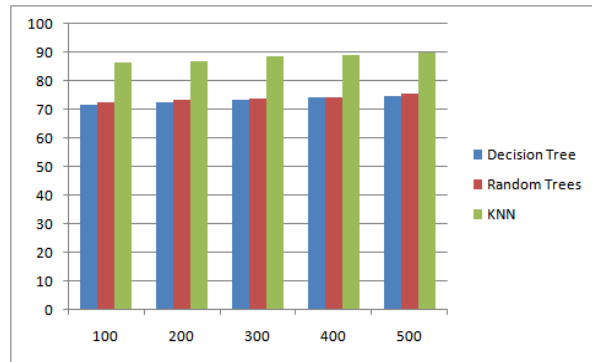


Table 4.3 Precision Table Figure 4.3 Precision

Recall(%): Through the above experimental outcomes, we recognize that the proposed algorithm outperforms the others with appreciate to growing transactions and gadgets in terms of scalability as well as runtime and memory usage for the actual datasets.

No of Samples	Naive Bayes	Random Forest	KNN
100	77.1	81.5	91.1
200	78.3	82.4	92.5
300	79.4	83.4	93.8
400	80.6	84.7	95.2
500	80.8	85.2	96.4

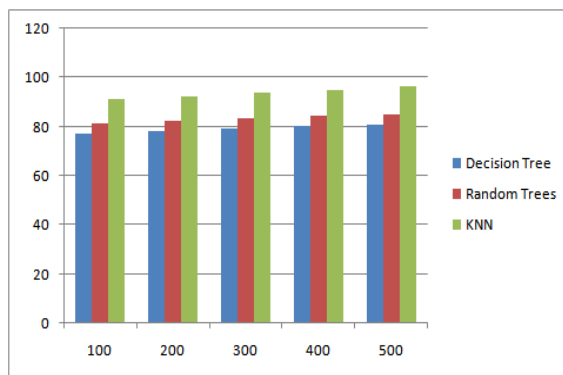


Table 4.4 Recall Table Figure 4.4 Recall

In our experimental consequences showed that KNN first-rate performance in terms of accuracy, precision, recall, memory utilization, and scalability. Proposed work giant improvement in classification accuracy, precision, recall and in survival chance. This massive impact indicates that use of a KNN in selection guide systems has more than one benefit, each in phrases of gadget accuracy and in phrases of machine transparency.

5. CONCLUSION

In this paper, the overall performance of Machine studying consisting of, Random Forest, Naive Bayes and KNN changed into evaluated to be expecting the water exceptional of drinking water. To this stop, the most dataset-associated famous additives, inclusive of oxygen, pH, conductivity, ORP, and turbidity etc., were collected. Results indicated that the carried out

fashions have suitable performance for predicting water pleasant additives, however, the excellent performance was related to the KNN. Furthermore, a contrast of the performance of carried out fashions indicated that the outcomes of KNN models have been extra reliable in comparison with the Random Forest and Naïve Bayes. In this case, using a prescriptive evaluation based totally on projected values would bring about future abilities to help decision and coverage makers.

References

1. Amir Hamzeh Haghiabi, Ali Heider Nasrolahi and Abbas Parsaie, "Water quality prediction using machine learning methods", *Water Quality Research Journal*, pp:1-11, January 2018, DOI:10.2166/wqrj.2018.025.
2. Jinal Patel, Charmi Amipara, Tariq Ahamed Ahanger, Komal Ladhva, Rajeev Kumar Gupta, Hashem O. Alsaab, Yusuf S. Althobaiti and Rajnish Ratna, "Machine Learning Based Water Potability Prediction Model by Using Synthetic Minority Oversampling Technique and Explainable AI", *Hindawi Computational Intelligence and Neuroscience*, pp:1-15, 2022, <https://doi.org/10.1155/2022/9283293>.
3. Xianhe Wang, Ying Li, Qian Qiao, Adriano Tavares, Yanchun Liang, "Water Quality Prediction Based on Machine Learning and Comprehensive Weighting Methods", *Entropy*, Volume 25, Issue 8, 2023, <https://doi.org/10.3390/e25081186>.
4. Nishant Rawat, Mangani Daudi Kazembe, Pradeep Kumar Mishra, "Water Quality Prediction using Machine Learning", *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, Volume 10 Issue VI June 2022-, ISSN: 2321-9653.
5. Srinidhi Iyer, Simran Kaushik, Poonam Nandal, "Water Quality Prediction Using Machine Learning", *MR International Journal of Engineering and Technology*, Vol. 10, No. 1, May 2023, <https://doi.org/10.58864/mrijet.2023.10.1.8>.
6. Dr. Kalaivazhi Vijayaragavan, N. Praveen, M. V. Sudharsan and P. S. Vijayan, "Machine Learning Model for Water Quality Prediction using Python and AI framework", *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, Volume 2, Issue 3, May 2022, ISSN (Online) 2581-9429.
7. M Srikanth and K Anish Reddy, "Water Quality and Potability using Machine Learning", *International Journal of Creative Research Thoughts*, Volume 11, Issue 5 May 2023 | ISSN: 2320-2882.

8. KassahunBirhanuTadesse and MegersaOlumana Dinka, "Water Quality Class Modeling Using Machine Learning Algorithms at Roodeplaas Dam, South Africa",2023, doi: 10.20944/preprints202306.1016.v1.
9. Aldhyani TH, Al-Yaari M, Alkahtani H, MaashiM., "Water quality prediction using artificial intelligence algorithms", Applied Bionics and Biomechanics, Dec-2020, <https://doi.org/10.1155/2020/6659314>.
10. Sai SreejaKurra, SambangiGeethika Naidu, SravaniChowdala, SreeChithraYellanki, Dr. B. Esther Sunanda," Water Quality Prediction using Machine Learning", International Research Journal of Modernization in Engineering Technology and Science, Volume:04, Issue:05, May-2022, ISSN: 2582-5208.
11. Ali NajahAhmeda, FaridahBintiOthmanb, HaithamAbdalmohsinAfanb, RusulKhaleelIbrahimb, Chow Ming Faic, MdShabbirHossaind, Mohammad Ehterame,AhmedElshafieb,"Machine learning methods for better water quality prediction", Journal of Hydrology, 2019,<https://doi.org/10.1016/j.jhydrol.2019.124084>.
12. Yogalakshmi S. and Mahalakshmi A," Efficient Water Quality Prediction for Indian Rivers Using Machine Learning", Asian Journal of Applied Science and Technology (AJAST), Volume 5, Issue 1, Pages 100-109, January-March 2021, ISSN: 2456-883X.
13. Poornima Jayaraman, Kothalam Krishnan Nagarajan , Pachaivannan Partheeban , Vallidevi Krishnamurthy,"Critical review on water quality analysis using IoT and machine learning models", Elsevier-International Journal of Information Management Data Insights, January 2024,<https://doi.org/10.1016/j.jjime.2023.100210>.
14. Nida Nasir, Afreen Kansal, Omar Alshaltone, Feras Barneih, Mustafa Sameer, Abdallah Shanableh, Ahmed Al-Shamma'a,"Water quality classification using machine learning algorithms", Elsevier-Journal of Water Process Engineering,2022, <https://doi.org/10.1016/j.jwpe.2022.102920>.
15. Hongfang Lu and Xin Ma,"Hybrid decision tree-based machine learning models for short-term water quality prediction", Elsevier-Chemosphere, 2020, <https://doi.org/10.1016/j.chemosphere.2020.126169>.
16. Wang, Leizhi, et al. "Improving the robustness of beach water quality modeling using an ensemble machine learning approach." Science of the Total Environment 765 (2021): 142760.

17.Jiang, Yiqi, et al. "A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks." *Journal of Cleaner Production* 318 (2021): 128533.