



EVALUATING THE PERFORMANCE OF MACHINE LEARNING ALGORITHMS FOR PREDICTING METEOROLOGICAL PARAMETERS

Yannick Mubakilayi¹, Simon Ntumba¹, Pierre Kafunda¹, Gracias Kabulu², Theodore Kabangu², Christian Kabeya³

*Corresponding author
myannick@aims.ac.rw

¹Department of Computer Science, University of Kinshasa, Kinshasa, DR Congo

²Department of Computer science, University of Mbuji mayi, Mbuji mayi, DR Congo

³Department of Mechanics, University of Mbuji mayi, Mbuji mayi, DR Congo

¹B.P.190 Kinshasa 11, University of Kinshasa, Kinshasa, DR Congo

^{2,3}B.P.225 Mbuji mayi, University of Mbuji mayi, Mbuji mayi, DR Congo

Article history:

Received June 20, 2023

Revised July 22, 2023

Accepted August 30, 2023

Keywords:

Machine Learning;

Prediction;

Climate Change

Abstract

Artificial learning techniques are currently used for weather and climate forecasting, etc. In this paper, we will evaluate three algorithms for predicting meteorological parameters based on the humidity parameter. Our dataset was taken from Mbuji mayi airport in the DRC. So, having the real-world data at our disposal, we used these Machine Learning tools to interpret and understand what happened by training the three models separately, and draw the conclusion as to which was the best model. Then, we used the three models to make some predictions about what our environment will be like tomorrow, and to draw conclusions and make decisions about whether or not our climate is already facing climate change. Three models are used: Decision tree, k-nearest neighbor and neural network, the analysis reveals that of the three models tested, the decision tree scored 81.8% after training with an average prediction of 71.5%, in second place we have the K-nearest neighbor with a score of 70% after training with an average prediction of 70, 8% and the closure neural network with 64% training and an average prediction of 66.1%. Thus, the decision tree outperforms the other models in terms of training and prediction of meteorological parameters, and is the best model with a very high performance compared to the other models.

1.0 INTRODUCTION

Today, climate change and global warming are burning issues worldwide as their negative impact alters human lives [1]. Climate change is expected to increase the risk of water-related disasters such as urban floods and severe droughts [2-4], or to have an intense effect on river water quality in many parts of the world [5]. Having a good understanding of the future, particularly with regard to temperature change, is essential to help decision-makers assess and reduce the effects of climate change, and increase the reliability and sustainability of infrastructure [6-8].

In assessing the quality of human life, humidity is known to be essential. Therefore, a reliable and accurate framework for accurately predicting humidity in the future is of great importance [9]. Such predictions with 100% accuracy may be impossible, but prediction errors can be minimized or prediction speed can be intensified [10, 11]. The current demand for this type of information has encouraged the development of new means of prediction and data processing, and led to the presentation of new specialized tools and methods [12-14]. Over the last twenty years, these new tools, known as machine learning (ML) methods, have demonstrated their capabilities and accuracy in various fields of science and engineering, such as forecasting, prediction and pattern recognition problems [15-18]. Numerous studies have used ML methods and algorithms to predict natural phenomena such as humidity, dew point, precipitation and soil temperature.

One of the long-standing and well-known ML methods used in many environmental fields is linear regression, applicable to both univariate and multivariate problems [19-21]. Another non-parametric ML method, k-Nearest Neighbor (kNN), has been chronically used in the ML literature to solve regression problems [22-24]. Support Vector Machine (SVM) regression is another method applied to find relationships between inputs and outputs [24, 25]. Finally, the artificial neural network (ANN) is one of the best-known and most popular methods, capable of capturing non-linear patterns in the features and targets of functional relationships. It has been applied to many ML problems in different domains [26-28]. Researchers have used the above methods to predict natural phenomena in recent years. Jain et al [29] used the coactive neuro-fuzzy inference system (CANFIS) to predict soil temperature in an arid and semi-arid zone. Jain et al. Jain et al. [30] developed an ANN model to predict winter air temperatures. They used data from six hours before the prediction time, including relative humidity, wind speed and solar radiation. Smith, McClendon and Hoogenboom Smith, McClendon and Hoogenboom [31] improved on the predictions of previous studies by extending the data lag to twenty-four hours, and concluded that this improved prediction accuracy. [32] used two ML methods, Support Vector Machine (SVM) regressions and Multilayer Perceptron (M.L.P.), to predict mean monthly temperature at eight observing stations in Australia and two in New Zealand, with implications for the detection of possible climate change reported in these regions. They concluded that SVM produced better predictions. [33] proposed a new approach to air temperature prediction by combining ANFIS and optimization algorithms. The proposed method showed an increase in prediction accuracy. In another study, [34] used kNN and a model matching machine learning algorithm to predict short-term air temperature.

Literature reviews have shown that despite numerous studies focusing on the use of specific ML methods to predict air temperature, few studies compare and evaluate the application of multiple ML methods in the same study [35]. Since temperature has a direct effect on precipitation, which will create a drought event, different ML approaches would help predict a drought event [36]. This work could be used as a reference for other hydrological and hydro-metrological parameters such as precipitation, flow measurement, etc. for this study region. Consequently, this study aims to evaluate the performance of machine learning algorithms based on the humidity parameter of the meteorological station at Bipemba International Airport in Mbujimayi. First, historical data were obtained and analyzed. Then, three different ML methods were trained using the actual data. The predictive ability of the ML methods was then evaluated and compared. Model inputs were selected on the basis of previous studies and available data, humidity, temperature, heavy rain, light rain, rainfall and frequency.

2.0 THEORETICAL

The research entitled evaluating the performance of machine learning algorithms for predicting meteorological parameters defines whether the machine learning model is reliable, where it makes errors and how much it can improve in order to choose the right model for the meteorological problem.

T. Saranya (2020) explored the comparative study of various ML algorithms used in IDS for several applications such as fog computing, Internet of Things (IoT), big data, smart city and 5G network. In addition, he also classified intrusions using ML algorithms such as Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART) and Random Forest.

Sunil gupta & al. (2022) working on how to verify algorithms on a dataset using R and find the most accurate algorithm model for the dataset. It's not easy to know which algorithm to use on the dataset to get the best results. We don't know the best parameters to use for particular algorithms. Here, we have worked on a trial-and-error strategy to select the most appropriate algorithm. The dataset is used to train the models, and a test option is used to evaluate the model. The test metrics are used for comparison.

Muhammad Kamran & al. (2022) present a comparative analysis of ML algorithms for short-term load forecasting (STLF) with regard to forecast accuracy and error. Based on implementation and analysis, they identify that, among other algorithms, STLF provides comparatively better results.

Savan K Patel & al. included in their paper a comparison of the performance of a few machine learning algorithms in referring students' social engagement during the COVID-19 pandemic period. In this study, the comparison of the Naïve Bayes, J48 tree, REPTree, and Random forest algorithms is carried out on a structured data set of over 1,200 instances.

2.1. Machine learning

Machine learning is the scientific study of the algorithms and statistical models used by computer systems to perform a specific task without being explicitly programmed. Learning algorithms are present in many of the applications we use every day. One of the reasons why a search engine like Google works so well every time it is used to search the Internet is that it has a learning algorithm that has learned how to rank web pages. These algorithms are used for a variety of applications, including data mining, image processing, predictive analysis and more. The main advantage of machine learning is that once an algorithm has understood what to do with data, it can do its job automatically.

2.2. Meteorology

Meteorology is the discipline that studies the weather and what it's going to do in the near future. Meteorology includes areas such as live situation analysis, weekly forecasting and seasonal forecasting. As such, it is the study of short-term weather.

3.0 METHODOLOGY

3.1. Study Area

Built in 1958 by Forminière (now MIBA), and handed over to the Ministry of Transport and Communication in 1960 (Ordonnance-Loi n° 0162/12 du 18 nov. 1960), Mbuji mayi national airport (FZWA according to the ICAO code and MJM according to the IATA code) has been one of the 54 airports in the DRC managed by the RVA since 1972. It is one of the country's top five airports. It is located in the RVA district of the commune of Bipemba. Its coordinates are: 6°07'16" S, 23°34'08" E and 677 m altitude. Theoretically, its area covers 300 hectares, but in practice only 169 hectares (1.69 km²) remain for a perimeter of 6.54 km, due to the encroachment of housing.

Its runway, 2000 m long and 45 m wide, oriented S-E-N-W, is known as Runway 17/35 (Figure 2). Due to its altitude (677m), Mbuji mayi airport, situated at the top of an elongated hill, is a center of runoff dispersion, particularly to the northeast and east of the Muya-river watershed [37].

3.2. Machine Learning Methods

a) K nearest neighbor

The nearest neighbor approach was first introduced by [38] and later studied by [39]. This approach is one of the simplest and oldest methods used for model classification. It often gives efficient performance and in some cases its accuracy is higher than that of state-of-the-art classifiers [40] [41]. The KNN classifier categorizes an unlabeled test example using the label of the majority of examples among its k nearest (most similar) neighbors in the training set. Similarity depends on a specific distance metric, therefore, the performance of the classifier depends significantly on the distance metric used [42]. KNN classifier is one of the most popular neighborhood classifiers in pattern recognition [43] and [44], as the technique is very simple and effective in the field of pattern recognition, machine learning, text categorization, data

mining and object recognition. etc. [45] and [46]. However, it has limitations, such as memory requirements and time complexity, as it depends entirely on each example in the training set.

b) Neural network

Modeling of a typical artificial neuron and multilayer neural network is shown in Figure 2. Signal flow from inputs $x_1 \dots x_n$ are assumed to be unidirectional, indicated by arrows pointing to the neuron's output flow (O).. The neuron output signal O is given by

$$O = f(\text{net}) = f\left(\sum_{j=1}^n w_j x_j\right) \quad (1)$$

Where w_j is the weight vector and the function $f(\text{net})$ is the activation function.

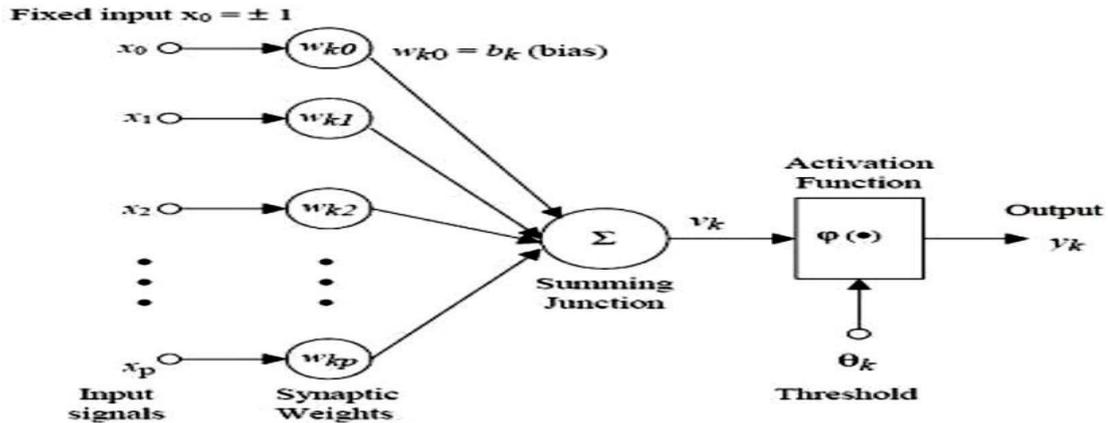


Figure 2. Neural Model

The variable net is defined as the dot product of the weights and the input vector.

$$\text{net} = w^T x = w_1 x_1 + \dots + w_n x_n \quad (2)$$

Where T is the transpose of the matrix.

The output value O is calculated as

$$O = f(\text{net}) = \begin{cases} 1 & \text{if } w^T x \geq \theta \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Where θ is called the threshold. This type of node is called a linear threshold unit. The internal activity of the neuron model is given by

$$v_k = \sum_{j=1}^p w_{kj} x_j \quad (4)$$

In this case, the output of neuron Y_k is the result of the activation function for the values of V_k . [47].

c) Decision tree

The decision tree approach is a binary model (split into two) that indicates how the quantity of a dependent variable can be estimated from the values of the independent variables. There are two types of decision of decision trees: [48] classification trees are the most common and [49] regression trees are used for estimation purposes based on numerical variables [50].

If each leaf of the tree contains linear regression relationships for the prediction of the target variable in that leaf, this is called the tree model. The M5 decision tree algorithm was developed by Quinlan [51]. The M5 algorithm uses single attribute tests that maximize the variance in the target space, creating a regression sequence by iteratively dividing the sample space. A mathematical formula for calculating the standard deviation reduction (SDR) is:

$$\text{SDR} = \text{sd}(T) = \sum_{\frac{|T_i|}{|T|}} \text{sd}(t_i) \quad (5)$$

where T is a set of examples that reach the node, T_i is the subset of examples that have the i th result of the potential set, and sd is the standard deviation. After the tree grows, multiple linear

regression is created for each internal node using the data for that node and all attributes involved in the in the subtest of that node. Each sub-tree is then considered in the pruning to overcome irregularity growth problems. Pruning occurs when the prediction error in the linear relationship at the root of a subtree is less than or equal to the expected error for the subtree. Finally, smoothing is used to compensate for sharp discontinuities between adjacent linear patterns on the leaves of the pruning tree.

d) System design

In this study, we stated by preparing data and applying selected prediction algorithm, and finally viewing the evaluation results taking into account training and prediction as shown in figure below

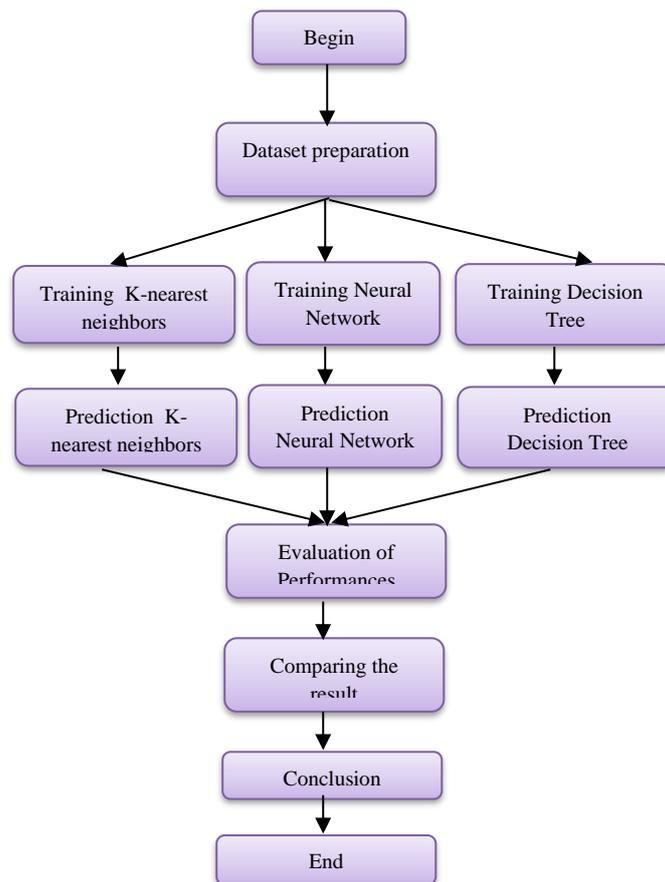


Figure 3. Study's Flowchart

4.0 RESULTANTS

4.1. DataSet Preparation

```
dataset.tail()
```

	Year	Month	Rainfall	weather	Storm	Humidity	Temperature	Wind Speed	Rain Frequency
79	2021	Aout	79	3	2	45.1	34	26	5
80	2021	Septembre	110	4	9	60.0	33	30	13
81	2021	Octobre	121	4	9	71.0	33	25	13
82	2021	Novembre	101	2	14	72.0	31	22	16
83	2021	Decembre	106	4	15	73.0	31	26	19

Figure 4. DataSet Preparation

The data shows us that we have 10 parameters for a total of 84 entries in the space from 2015 to 2021 and concerning the data types, we have integers.

4.2. Models training

K-nearest Neighbors

```
modelKNN=KNeighborsRegressor(n_neighbors=3)
modelKNN.fit(x,y)
print("Score K-nearest neighbors : ", modelKNN.score(x,y))
```

Score K-nearest neighbors : 0.7025624458475073

Figure 5. Score K-nearest neighbors

Neural Network

```
modelRN=MLPRegressor(hidden_layer_sizes=(500,500,2),max_iter=500)
modelRN.fit(x,y)
print("Score Neural Network : ", modelRN.score(x,y))
```

C:\Users\ciman_rgyv8yt\anaconda3\lib\site-packages\sklearn\neural_network_multilayer_perceptron.py:1599: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
y = column_or_1d(y, warn=True)

Score Neural Network : 0.6462994847478969

C:\Users\ciman_rgyv8yt\anaconda3\lib\site-packages\sklearn\neural_network_multilayer_perceptron.py:692: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (500) reached and the optimization hasn't converged yet.
warnings.warn()

Figure 6. Score Neural Network

Decision Tree

```
modelDecisionTree=DecisionTreeRegressor(max_depth=3)
modelDecisionTree.fit(x,y)
print("Score Decion Tree : ", modelDecisionTree.score(x,y))
```

Score Decion Tree : 0.8187131520754635

Figure 7. Score Decision Tree

Table 1. Taining result

Models	Training score
K-nearest neighbors	70,2%
Neural Network	64,6%
Decision Tree	81,8%

As can be seen in Table 1, having chosen humidity as the fixed Target, the initial parameters of the various models were then trained using the "fit" method. After training, the various models were then evaluated using the "score" method: the K-nearest neighbor model scored 70.2%, the Neural Network model scored 64.6% and the Decision Tree model scored 81.8%, with the last model scoring the highest of the three.

Predictions with different models

```
modelRN.predict(x)
array([ 72.70406882,  65.72779353,  69.41148247,  66.9143401 ,  62.56900439,
        60.54008077,  63.71908688,  67.61025226,  64.03327579,  67.83089028,
        68.24691165,  70.38849743,  71.407766 ,  66.57177793,  68.26970464,
        66.63287854,  61.89062605,  55.36629382,  55.60772737,  62.89402794,
        65.13495662,  64.55502437,  66.57676602,  63.60513645,  68.17390997,
        64.97065171,  65.12844137,  63.03439192,  60.03280083,  56.35519136,
        63.2456423 ,  66.52179804,  65.29473222,  65.52264439,  67.43819355,
        68.99010284,  69.46794874,  63.64027048,  68.06519688,  65.77937156,
        64.86658164,  52.62414169,  52.24217761,  63.05695108,  62.84720024,
        65.12499979,  65.47459526,  69.85672589,  70.71999114,  64.75894987,
        76.97636519,  64.02512709,  64.52770657,  54.56588657,  59.74321631,
        62.60544948,  64.22537543,  69.75320782,  66.68713247,  70.5016144 ,
        70.60420616,  65.76114806,  67.42629426,  65.36648457,  60.41052591,
        55.00246469,  57.76661203,  65.52446169,  64.21993731,  70.58955868,
        66.88307181,  68.77075628,  68.40202519,  63.4618088 ,  66.26061207,
        75.21960676,  60.4522741 ,  55.3613406 ,  52.40431745,  53.35682351,
        60.19751767,  67.2498175 ,  66.57192321,  68.05988278])
```

Figure 8. Neural network prediction

After training our "Neural Network" model, we proceeded with the prediction over a multi-year period of seven years. We can see a constant decrease in relative humidity just after January, reaching the critical threshold of 63% in July. This may be understandable given the rainy season, but it gradually rises again to 68% in August, before dropping again to 64% and 65% in September and October, despite being in the rainy season. This is something we experienced in the city, as September and October 2022 were almost rain-free, and it rises again to 76% in November, which is a particularly wet and rainy month, but we notice a drop in December to reach the critical threshold of 62%.

```
modelArbreDecision.predict(x)
array([ 82. , 76. , 75.33333333, 72.8 , 63. ,
        61. , 65. , 72.8 , 68. , 72.8 ,
        74.5 , 77. , 79. , 75. , 74.5 ,
        72.8 , 61. , 51. , 50. , 63. ,
        70. , 66. , 72.8 , 66.6 , 72.25 ,
        74. , 72. , 66. , 58. , 53. ,
        64. , 70.33333333, 70. , 68. , 73.2 ,
        75.33333333, 76. , 72. , 73.2 , 70.33333333,
        67. , 45. , 43. , 64. , 65. ,
        67. , 70. , 76. , 78. , 74. ,
        89. , 68. , 66. , 49. , 58. ,
        63. , 68. , 76. , 73.2 , 77. ,
        78. , 73.2 , 72.25 , 70. , 59. ,
        50. , 54. , 68. , 68. , 77.4 ,
        72.25 , 74. , 75.33333333, 71. , 70.33333333,
        89. , 59. , 51. , 43. , 45.1 ,
        60. , 72.25 , 72. , 73.2 , ])
```

Figure 9. Prediction with the decision tree

Using the decision tree, we can see that in the first month of 2022, humidity reaches 82%, while in February it drops to 61%. In March, it rises to 74%, before dropping briefly to 72% in the fourth month of the year. It rises again in June to 74% and drops again in July to 64%, which is understandable given the dry season, then rises again in August to 75% with the onset of rain, before dropping to 67% between September and October, given the scarcity of rain observed in the city during this period and the heat peaks. It rises again in November to reach a record 89%, due to the abundance of rain, and drops again to its lowest level in December to 63%.

```

modelKPPV.predict(x)
array([[80.05714286],
       [70.85714286],
       [75.57142857],
       [72.71428571],
       [64.],
       [59.42857143],
       [67.8],
       [72.57142857],
       [67.57142857],
       [73.28571429],
       [73.85714286],
       [77.05714286],
       [78.34285714],
       [72.71428571],
       [74.],
       [72.71428571],
       [63.57142857],
       [49.87142857],
       [51.14285714],
       [64.],
       [70.14285714],
       [69.],
       [72.71428571],
       [67.8],
       [73.85714286],
       [70.57142857],
       [70.],
       [65.14285714],
       [59.42857143],
       [51.14285714],
       [66.94285714],
       [72.71428571],

```

Figure 10. K-nearest neighbors prediction

In Figure 10, predicted with the K-nearest neighbors model, we see humidity at its lowest during the dry season, i.e. the months of May, June and July, rising in August to 72% with the onset of the rains, falling in September to 67%, which may explain the scarcity of rains and heat peaks during this month, rising again to 73% between October and November to explain the high temperatures observed during this time of the year, and rising again in December.

Table 2. Prediction result

Month	K-nearest neighbors	Neural network	Decision tree
January	80	72	82
February	70	60	61
March	75	68	74
April	72	66	72
May	64	65	70
June	59	64	74
July	67	63	64
August	72	68	75
September	67	64	67
October	73	65	67
November	73	76	89
December	77	62	63
Mean	70,8	66,1	71,5

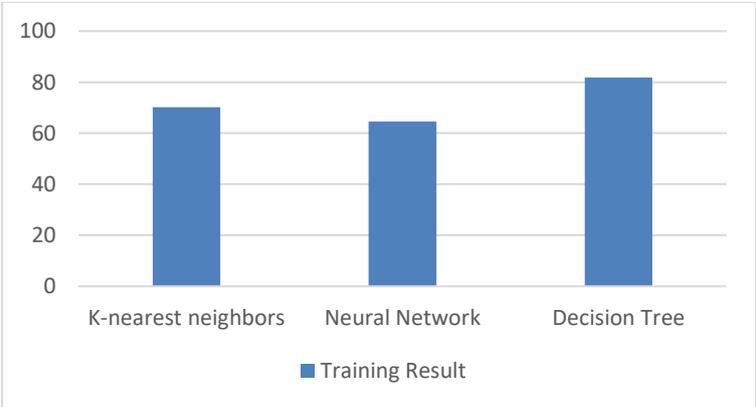


Figure 11. Training Result

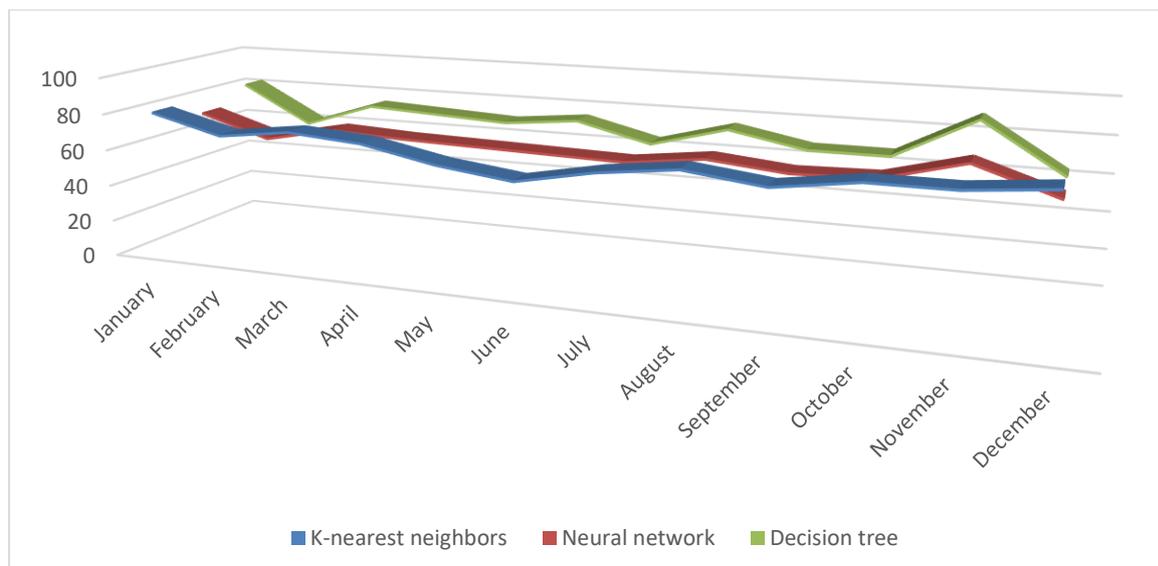


Figure 12. Prediction result

5.0 CONCLUSION

This study focused on evaluating the performance of machine learning algorithms in forecasting meteorological parameters in Mbuji Mayi City, meteorological data including humidity, temperature, heavy rain, light rain, precipitation, frequency, etc. were extracted from the weather station at Bipemba International Airport. The data covered almost seven years from 2015 to 2021. The main purpose of this data is that this study consisted of evaluating the performance of different ML techniques compared to training and prediction using humidity as input. Therefore, three ML methods were used for training and prediction, including K-nearest neighbors, neural networks, and Decision tree. Analysis showed that the decision tree in training gave him a score of 81.8% and the average prediction was 71.5% for him. In second place, the K nearest neighbor score he had 71.5%. Training was 70% and average prediction was 70.8%, Cloture neural network was training 64% and average prediction was 66.1%. Therefore, decision trees outperform other models in terms of training and predicting meteorological parameters, and are the best models with very high performance compared to others. This can be seen in figures 11 and 12.

REFERENCES

- [1] S. Thober, R. Kumar, N. Wanders, A. Marx, M. Pan, O. Rakovec, L. Samaniego, J. Sheffield, E. F. Wood, M. Zink, Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degrees global warming, *Environ. Res. Lett.* 13 (2018) 014003.
- [2] K. Mohammed, A. K. M. S. Islam, G. M. T. Islam, L. Alfieri, M. J. U. Khan, S. K. Bala, M. K. Das, Future Floods in Bangladesh under 1.5°C, 2°C, and 4°C Global Warming Scenarios, *J. Hydrol. Eng.* 23 (2018) 04018050.
- [3] B. Azari, M. Tabesh, Urban storm water drainage system optimization using a sustainability index and LID/BMPs, *Sustain. Cities Soc.* 76 (2022) 103500.
- [4] K. Azizi, C. I. Meier, Improving the Characterization of Urban Flash Floods through Application of Local Knowledge, in *AGU Fall Meeting Abstracts*, (2020), 2020, H162-0011.
- [5] S. Ebrahimi, M. Khorram, Variability effect of hydrological regime on river quality pattern and its uncertainties: case study of Zarjoob River in Iran, *J. Hydroinformatics* 23 (2021) 1146–1164.
- [6] M. Tabesh, B. Azari, H. Rezaei, Hydraulic performance assessment of stormwater networks, in *Proceedings of the International Conference on Civil, Architecture and Disaster Management (ICCADM-16)*, Hong Kong, (2016), 17–18.

- [7] B. Azari, M. Tabesh, Storm Water Drainage Networks Hydraulic Performance Assessment, International Conference on Sustainable Development and Urban Construction, Iran. (2015)
- [8] K. Azizi, C. I. Meier, Urban Pluvial Flood Risk Assessment: Challenges and Opportunities for Improvement Using a Community-Based Approach, in World Environmental and Water Resources Congress 2021, (2021), 350–361. doi: 10.1061/9780784483466.033.
- [9] D. Absalon, B. Ślesak, Air temperature increase and quality of life in an anthropogenically transformed environment: A case study, Polish J. Environ. Stud. 21 (2012) 235–239.
- [10] Q. Song, B. S. Chissom, Forecasting enrollments with fuzzy time series — Part I, Fuzzy Sets Syst. 54 (1993) 1–9.
- [11] S.-M. Chen, J.-R. Hwang, Temperature prediction using fuzzy time series, IEEE Trans. Syst. Man, Cybern. Part B 30 (2000) 263–275.
- [12] N. Zhong, S. S. Yau, J. Ma, S. Shimojo, M. Just, B. Hu, G. Wang, K. Oiwa, Y. Anzai, Brain Informatics-Based Big Data and the Wisdom Web of Things, IEEE Intell. Syst. 30 (2015) 2–7.
- [13] Z. Su, Q. Xu, Q. Qi, Big data in mobile social networks: a QoE-oriented framework, IEEE Netw. 30 (2016) 52–57.
- [14] J. Wen, J. Yang, B. Jiang, H. Song, H. Wang, Big Data Driven Marine Environment Information Forecasting: A Time Series Prediction Network, IEEE Trans. Fuzzy Syst. 29 (2021) 4–18.
- [15] M. Alizamir, O. Kisi, A. N. Ahmed, C. Mert, C. M. Fai, S. Kim, N. W. Kim, A. El-Shafie, Advanced machine learning model for better prediction accuracy of soil temperature at different depths, PLoS One 15 (2020) 1–25.
- [16] B. Azari, M. Tabesh, Optimal design of stormwater collection networks considering hydraulic performance and BMPs, Int. J. Environ. Res. 12 (2018) 585–596.
- [17] A. Rahman, A. Khan, A. A. Raza, Parkinson's disease detection based on signal processing algorithms and machine learning, CRPASE Trans. Electr. Electron. Comput. Eng. 6 (2020) 141–145.
- [18] S. R. Moosavia, D. A. Woodb, S. A. Samadanic, Modeling Performance of Foam-CO₂ Reservoir Flooding with Hybrid Machine-learning Models Combining a Radial Basis Function and Evolutionary Algorithms, methods 4 (2020) 5.
- [19] K. A. Marill, Advanced Statistics: Linear Regression, Part II: Multiple Linear Regression, Acad. Emerg. Med. 11 (2004) 94–102.
- [20] M. M. Ahmadi, H. Shahriari, Y. Samimi, A novel robust control chart for monitoring multiple linear profiles in phase II, Commun. Stat. - Simul. Comput. (2020) 1–12.
- [21] S. Mehri, M. M. Ahmadi, H. Shahriari, A. Aghaie, Robust process capability indices for multiple linear profiles, Qual. Reliab. Eng. Int. 37 (2021) 3568–3579.
- [22] J. Liang, Q. Yang, T. Sun, J. D. Martin, H. Sun, L. Li, MIKE 11 model-based water quality model as a tool for the evaluation of water quality management plans, J. Water Supply Res. Technol. 64 (2015) 708–718.
- [23] I. Saini, D. Singh, A. Khosla, QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases, J. Adv. Res. 4 (2013) 331–344.
- [24] X. Wang, L. Ma, X. Wang, Apply semi-supervised support vector regression for remote sensing water quality retrieving, in 2010 IEEE International Geoscience and Remote Sensing Symposium, (2010), 2757–2760. doi: 10.1109/IGARSS.2010.5653832.
- [25] S. Liu, H. Tai, Q. Ding, D. Li, L. Xu, Y. Wei, A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction, Math. Comput. Model. 58 (2013) 458–465.
- [26] A. Najah, A. El-Shafie, O. A. Karim, A. H. El-Shafie, Application of artificial neural networks for water quality prediction, Neural Comput. Appl. 22 (2013) 187–201.
- [27] K. P. Singh, A. Basant, A. Malik, G. Jain, Artificial neural network modeling of the river water quality—A case study, Ecol. Modell. 220 (2009) 888–895.
- [28] B. Fotovvati, M. Balasubramanian, E. Asadi, Modeling and Optimization Approaches of Laser-Based Powder-Bed Fusion Process for Ti-6Al-4V Alloy, Coatings 10 (2020)
- [29] P. Hosseinzadeh Talaei, Daily soil temperature modeling using neuro-fuzzy approach, Theor. Appl. Climatol. 118 (2014) 481–489.

- [30] A. Jain, R. W. McClendon, G. Hoogenboom, R. Ramyaa, Prediction of frost for fruit protection using artificial neural networks, *Am. Soc. Agric. Eng. St. Joseph, MI, ASAE Pap.* (2003) 3–3075.
- [31] B. A. Smith, R. W. McClendon, G. Hoogenboom, Improving air temperature prediction with artificial neural networks, *Int. J. Comput. Intell.* 3 (2006) 179–186.
- [32] S. Salcedo-Sanz, R. C. Deo, L. Carro-Calvo, B. Saavedra-Moreno, Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms, *Theor. Appl. Climatol.* 125 (2016) 13–25.
- [33] A. Azad, H. Kashi, S. Farzin, V. P. Singh, O. Kisi, H. Karami, H. Sanikhani, Novel approaches for air temperature prediction: A comparison of four hybrid evolutionary fuzzy models, *Meteorol. Appl.* 27 (2020) e1817.
- [34] Y. Wang, Y. Bai, L. Yang, H. Li, Short time air temperature prediction using pattern approximate matching, *Energy Build.* 244 (2021) 111036.
- [35] M. Moeini, A. Shojaeizadeh, M. Geza, Supervised Machine Learning for Estimation of Total Suspended Solids in Urban Watersheds, *Water* 13 (2021) 147.
- [36] P. Kumar, S. F. Shah, M. A. Uqaili, L. Kumar, R. F. Zafar, Forecasting of Drought: A Case Study of Water-Stressed Region of Pakistan, *Atmosphere (Basel)*. 12 (2021) 1248.
- [37] L. Mukunayi, J. BanokaNsona, Kakule K, Kanda N, Impact of Mbujimayi Airport on Groundfill Cavities in Bipemba Community, *The International Journal of Engineering and Science*, 8(2019) 53-61
- [38] E. Fix and J. Hodges, "Discriminatory Analysis: Nonparametric Discrimination: Consistency Properties," 4, 1951.
- [39] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 21-27, 1967.
- [40] Y. Hamamoto, S. Uchimura, and S. Tomita, "A Bootstrap Technique for Nearest Neighbor Classifier Design," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 19, no. 1, pp. 73-79, 1997.
- [41] E. Alpaydin, "Voting Over Multiple Condensed Nearest Neighbors," *Artificial Intelligence Review*, vol. 11, pp. 115-132, 1997.
- [42] K. Q. Weinberger and L. K. Saul, "Distance Metric Learning for Large Margin Nearest Neighbor Classification," *Journal of Machine Learning Research*, vol. 10, pp. 207-244, 2009.
- [43] A. Kataria and M. D. Singh, "A Review of Data Classification Using K-Nearest Neighbour Algorithm," *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 6, pp. 354-360, 2013.
- [44] N. Bhatia and A. Vandana, "Survey of Nearest Neighbor Techniques," (*IJCSIS*) *International Journal of Computer Science and Information Security*, vol. 8, no. 2, pp. 302-305, 2010.
- [45] A. B. A. Hassant, "Visual Speech Recognition," in *Speech Technologies*, I. Ipsic, Ed. Rijeka: InTech - Open Access Publisher, 2011, vol. 2, ch. 14.
- [46] A. B. A. Hassanat, "Visual Passwords Using Automatic Lip Reading," *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, vol. 13, no. 1, pp. 218-231, 2014.
- [47] P. Kumar P., *Artificial Neural Network Based Numerical Solution of Ordinary Differential Equations*, Thesis, National institute of technology, 2012
- [48] Unlukara, A.; Kurunc, A.; Cemek, B. Green Long Pepper Growth under Different Saline and Water Regime Conditions and Usability of Water Consumption in Plant Salt Tolerance. *J. Agric. Sci.* 2015, 21, 167–176.
- [49] Bruin, H.; Trigo, I. A New Method to Estimate Reference Crop Evapotranspiration from Geostationary Satellite Imagery: Practical Considerations. *Water* 2019, 11, 382. [CrossRef]
- [50] Witten, I.H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Morgan Kaufmann Series in Data Management Systems: Washington, DC, USA, 2005.
- [51] Quinlan, J. Simplifying decision trees. *Int. J. Man-Mach. Stud.* 1987, 27, 221–23