



Mathematical Tool for Predicting the Weather Condition of Coastal Regions of Nigeria: a Case Study of Bayelsa State, Nigeria

EKEREKE, L; *AKPOJARO, J

¹Department of Mathematical Sciences, Faculty of Basic and Applied Sciences, University of Africa, Toru-Orua, Bayelsa State, Nigeria
*Corresponding Author Email: jakpojaro@yahoo.com; ekerekelayefa@rocketmail.com

ABSTRACT: Classification of data in its simplicity is a means of categorizing data into different categories according to rules. In this paper, we reviewed some data mining techniques that are relevant to classification of weather data set. Based on the classification and decision tree rules, we generated a different attributes' table. These attributes were imputed into the WEKA software to produce a decision tree, from which we predicted the appropriate boat users can take based on the weather condition. The results of this study can significantly strengthen decision-making ability of stakeholders who ply the coaster regions of Nigeria, particularly in Bayelsa State. We recommend that bigger boats should be equipped with software that have capabilities for providing artificial intelligence, decision system, sensors, based on weather conditions to facilitate decision-making and also to exhibit intelligence when necessary.

DOI: <https://dx.doi.org/10.4314/jasem.v23i5.16>

Copyright: Copyright © 2019 Ekereke and Akpojaro. This is an open access article distributed under the Creative Commons Attribution License (CCL), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Dates: Received: 26 May 2019; Revised: 24 May 2019; Accepted 27 May 2019

Keywords: Decision tree, Weather, Data Mining, Boat, Water Transportation Sector.

Data mining is the extraction of implicit, previously unknown and rotationally useful information from data set (Chaudhari *et al.*, 2013). Data mining consists of five major elements. First, extract, transform, and load transaction data onto the data warehouse system. Second, store and manage the data in a multidimensional database system. Third, provide data access to business analysts and IT professionals. Fourth, analyze the data using application software. Fifth, present the data in a useful format, such as a graph or table. Many data mining techniques are closely related to some of machine learning techniques (Pabreja, 2012) while others are related to techniques that have been developed in statistics. In this work, we review some techniques related to data mining and data classification techniques that can be used to solve water transportation problems. Water transportation is vital in the sense that it can be useful to carry heavy, large and huge load and this means depends wholly on good weather condition. The technique of data mining will be explored using some data set to predict weather so that water transporters are able to decide whether to use an open boat or a covered boat while conveying large quantity of materials or items.

The objective of this study is to learn a concise representation of weather data set to predict appropriate weather conditions for boat users.

MATERIALS AND METHODS

There are so many techniques available for data classification. The ability of a classifier refers to the ability to correctly classify the unseen data in a class (Shrivastava *et al.*, 2017). In data analysis, it is essential to put the instances in a desired class. For the purpose of this study, we discuss four well known techniques of data classification that are relevant to this study.

Rule based methods: Data mining system learns from examples. It formulates classification rules in order for the prediction of future. Typically, rule are be produced by rule based systems (Viswambari and Selvi, 2014; Aftab *et al.*, 2018).

Neural Network: This can be used for classification purpose. They simulate the human brain. Artificial Neuron can be supervised or unsupervised. They are composed of many units called neuron. Artificial Neuron require long training time and are black box which lacks explanation, but it has high tolerance to noisy data so it can classify untrained data (Chaudhari *et al.*, 2013; Nagalakshmi *et al.*, 2013; Geetha and Nasira, 2014).

KNN (K-Nearest-Neighbor): This classifier uses the training data to assign new features to the class determined by the nearest data point. It uses the

*Corresponding Author Email: jakpojaro@yahoo.com

Euclidean distance measure (not suitable for categorical predictions) to it. (Badhiye *et al.*, 2012; Pabreja, 2012).

Decision Tree: This tries to find an optimal partitioning of the space of possible observations, mainly by the means of subsequent recursive splits (Olaiya, and Adeyemo, 2012; Adeyemo, 2013).

Training a standard decision tree leads to a quadratic optimization problem with bound constraints and one linear equality constraints. Training support vector machines involves a huge optimization problem and many specially designed algorithms have been proposed. We used an algorithm called “Decision Tree Induction” that accelerates the training process by exploiting the distributional properties of the training data. That is, the natural clustering of the training data and the overall layout of these clusters relative to the decision boundary of support vector machines.

A decision tree is a decision-making technique that is commonly used by making a graphical representation of the possible consequences of a number of given cases (Yang and Fong, 2011). It is called a decision tree since the graph used to represent the ramifications of the possible consequences, resemble the branches of a tree. Because of that, a decision tree can be used as a predictive model in a machine learning application. It’s flow-chart is like a tree structure, where each internal node denotes a test on an attribute, each branch denotes an outcome of test, and each leaf node holds a class label. The topmost node in a tree is the root node (Bartoka, Habalab, Bednarc *et al.*, 2012). Given a tuple, X, for which the associated class label is unknown, the attribute values of the tuple are tested against decision tree. A path is traced from the root to a leaf node, which holds the class prediction for that tuple.

In this study, we use the Decision Tree technique because the construction of decision tree classifiers does not require any domain knowledge. It can handle dimensional data. The learning and classification steps of decision tree induction are simple and fast. Their representation of acquired knowledge in tree form is easy to assimilate by users. Decision tree classifiers have good accuracy (Chen *et al.*, 2009).

Classification Rules: Every data classification project is different but the projects have some common features. Data classification requires some rules (Song and Lu, 2015). This classification rules are (i) The data must be available (ii) The data must be relevant, adequate, and clean (iii) There must be a well-defined problem (iv) The problem should not be solvable by

means of ordinary query (v) The result must be actionable

Decision Tree Rules: The decision tree algorithm is a top-down induction algorithm. The aim of this algorithm is to build a tree that has leaves that are homogeneous as possible. The steps involved in this algorithm are as follows: (i) Input: (ii) Data partition, *D*, which is a set of training tuples and their associated class labels. (iii) Attribute list, the set of candidate attributes. (iv) Attribute selection method, a procedure to determine the splitting criterion that “best” partitions the data tuples into individual classes.

Output: the output is a decision tree. The major advantage of converting a decision tree to rules is that it is easier to read and understand by users. The basic rules for decision tree are (Talib *et al.*, 2017): (i) Each path from the root to the leaf of the decision tree therefore consists of attribute tests, finally reaching a leaf that describes the class. (ii) If-then rules may be derived based on the various paths from the root to the leaf nodes. (iii) Rules can often be combined to produce a smaller set of rules. (iv) Once all the rules have been generated, it may be possible to simplify the rules. (v) Rules with only one antecedent cannot be further simplified. So we only consider those with two or more antecedents. (vi) Eliminate unnecessary rule antecedents that have no effect on the conclusion reached by the rule. (vii) In some cases, a number of rules that lead to the same class may be combined.

RESULTS AND DISCUSSION

In order to construct the decision tree for this study, we generated a data set as shown in Table 1. In this Table, there are four attributes (e.g., outlook, temperature, humidity, wind) used to decide if the passenger should board an open boat (OB) or a house boat (HB). For result (board a boat), there are two classes such as OB of HB. These attributes may be increased or decreased. But if the numbers of attributes are more, data classification can be done with more accuracy.

The temperature is measured with a thermometer at different weather conditions; when the temperature is above 70°C, it is said to be hot, when it is between 50°C and 69°C, it is said to be mild, and when it is below 50°C, it is said to be cold.

When the humidity is above 60°C, it is said to be high and when it is below, it is said to be low. Lastly, the wind is measured with a wind vane, when the level of the wind is below 55m/s it is said to be strong and when it is below, it is said to be weak.

Table 1: Shows different attributes

Outlook	Humidity	Temperature	Wind	Board a boat
Sunny	High	Hot	Weak	HB
Sunny	High	Hot	Strong	HB
cool	High	Hot	Weak	OB
Rainy	High	Mild	Weak	OB
Rainy	Normal	Cold	Weak	OB
Rainy	Normal	Cold	Strong	HB
cool	Normal	Cold	Strong	OB
Sunny	High	Mild	Weak	HB
Sunny	Normal	Cold	Weak	OB
Rainy	Normal	Mild	Weak	OB
Sunny	Normal	Mild	Strong	OB
Cool	High	Mild	Strong	OB
Cool	Normal	Hot	Weak	OB
Rainy	High	Mild	Strong	HB

We used the WEKA software based on Table 1 parameters and the output of the decision tree generated is shown in Figure 1.

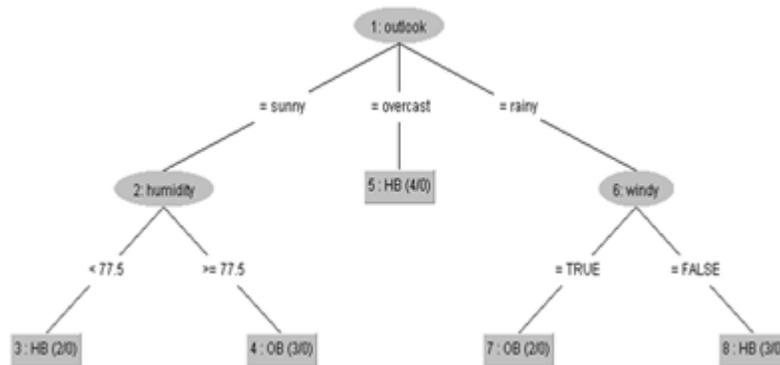


Fig 1: Decision tree using WEKA software

The best decision to make with respect to transportation in the coastal area has always been an issue the weather cannot easily be predicted because of its volatile nature. The inability to make the right decision because of the irregularity in the weather has led to several losses of lives, goods and properties in the process of conveying them to their destination via water. The type of boat used can reduce this menace or risk. Thus this study gives a decision tree that has been pre-trained with a few attribute data set in other to make the right decision both the conveyor and passengers as weather to use an open boat (speed boat) or a closed boat at any point in time.

This study used decision tree induction algorithm to train the data set in the decision tree. This pattern was used because it accelerates the training process by exploiting the distributive properties of the training data.

A training set for the 45 instances was prepared from the dataset Only four parameters (attributes) have been taken into consideration (minimum, maximum and average temperature, minimum, maximum and average humidity, average wind level and finally the outlook of the weather). WEKA data mining tool has

been used for the analysis of predicting an event by decision tree on the basis of above mentioned parameters. Data cleaning was done so that classes of events are not complex. The result shows as output in Figure 1 will determine the type of boat to be used to convey both the goods and humans to their destination at every point in time. From Figure 1 above, there are five possible outcomes that are given under some pre-defined rules. First the outlook is sunny, humidity is less than 77.5, you should use a house boat but if the humidity is greater than or equal to 77.5, an open boat can be used. Condition three: if the outlook is rainy, the weather is windy, use an open boat otherwise use a house boat.

Conclusion: In this paper, we reviewed some data mining techniques that are relevant to classification of weather data set. We imputed the data set into WEKA software to produce a decision tree, from which we predicted the appropriate weather condition for boat users. The results of this study can significantly strengthen decision-making ability of stakeholders who ply the coaster regions of Nigeria, particularly in Bayelsa State. Future work would improve the results by increasing the data set and adding more attributes in the model.

REFERENCES

- Adeyemo, A. (2013). Soft Computing Techniques for Weather and Climate Change Studies. *African Journal of Computing and ICT*, 6(2): 77-90.
- Aftab, S; Ahmad, M; Hameed, N; Bashir, MS; Ali, I; Nawaz, Z. (2018). Rainfall Prediction using Data Mining Techniques: A Systematic Literature Review. *International Journal of Advanced Computer Science and Applications*, 9(5): 143-150.
- Badhiye, SS; Wakode, BV; Chatur, PN. (2012). Analysis of Temperature and Humidity Data for Future value prediction. *International Journal of Computer Science and Information Technologies*, 3(1): 3012-3014.
- Bartoka, J; Habalab, O; Bednarc, P; Gazaka, M; Hluchýb, L. (2012). Data Mining and Integration for Predicting Significant Meteorological Phenomena. *International Conference on Computational Science, ICCS 2010, Elsevier*, pp. 37- 46.
- Chaudhari, AR; Rana, DP; Mehta, RG. (2013). Data Mining with Meteorological Data. *International Journal of Advanced Computer Research*, 3(3): 25-29.
- Chen, GH; Wang, ZQ; Yu, ZZ. (2009). Constructing Decision Tree by Integrating Multiple Information Metrics. *2009 Chinese Conference on Pattern Recognition*, pp. 2-6.
- Geetha, A; Nasira, GM. (2014). Artificial Neural Networks' Application in Weather Forecasting – Using Rapid Miner. *International Journal of Computational Intelligence and Informatics*, 4(3): 177-184.
- Nagalakshmi, R; Usha, M; Ramanathan, AN. (2013). Application of Data Mining Techniques in Maximum Temperature Forecasting: A Comprehensive Literature Review. *International Journal of Advance Research in Computer Science and Management Studies*, Special Issue, pp. 1-9.
- Olaiya, F; Adeyemo, AB. (2012). Application of Data Mining Techniques in Weather Prediction and Climate Change Studies. *Information Engineering and Electronic Business*, 1: 51-59.
- Pabreja, K. (2012). Clustering technique to interpret Numerical Weather Prediction output products for forecast of Cloudburst. *International Journal of Computer Science and Information Technologies*, 3(1): 2996-2999.
- Shrivastava, V; Karmakar, S; Soni, S. (2017). Suitability of Neural Network for Weather Forecasting: A Comprehensive Literature Review. *International Journal of Recent Scientific Research*, 8(12): 22300-22316.
- Song, Y; Lu, Y. (2015). Decision tree methods: Applications for classification and prediction. *Shangahaia archives of psychiatry*, 27(2): 130-135.
- Talib, MR; Ullah, T; Sarwar, MU; Hanif, MK; Ayub, N. (2017). Application of Data Mining Techniques in Weather Data Analysis. *International Journal of Computer Science and Network Security*, 17(6): 22-28.
- Viswambari, M; Selvi, RA. (2014). Data Mining Techniques to Predict Weather: A Survey. *International Journal of Innovative Science, Engineering and Technology*, 1(4): 450-452.
- Yang, HH; Fong, S. (2011). Optimized very Fast Decision Tree with Balanced Classification Accuracy and Compact Tree Size. *The 3rd International Conference on Data Mining and Intelligent Information Technology Applications*, China, pp. 57-64.