International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET2310288

Marine Weather Forecasting to Enhance Fisherman's Safety Using Machine Learning

M. Robinson Joel, G. Manikandan, M. Nivetha
Department of Information Technology, Kings Engineering College, Chennai, India

ARTICLEINFO

Article History:

Accepted: 01 April 2023
Published: 19 April 2023

Publication Issue
Volume 10, Issue 2
March-April-2023

Page Number
519-526

ABSTRACT

Through the use of scientific knowledge and weather measurements, weather forecasting is a technique for predicting what the atmosphere will be like in a certain location. In other words, it's a method by which the characteristics of a meteorological state are determined in advance by factors such as temperature, wind, humidity, rainfall, and the quantity of the data set. In an effort to foretell weather conditions now and in the future, meteorologists use a process called weather forecasting. For everyday operations, accurate weather forecasts are required, and this has been one of the most difficult problems to solve globally since the data is multidimensional and nonlinear. According to the survey, supervised and unsupervised machine learning algorithms, artificial neural networks, naive bayes algorithms, and random forest algorithms are some of the different techniques and algorithms utilised for weather prediction in the field of data mining.

Keywords: Datamining, Weather prediction, Weather forecasting, SVM, Navie

I. INTRODUCTION

Traditionally, weather predictions have been made using huge, complicated physics models that take into account a variety of atmospheric circumstances over a lengthy period of time. The weather system's disturbances frequently make these circumstances unstable, which leads the models to provide erroneous forecasts. In a big High-Performance Computing (HPC) environment, the models are often run on hundreds of nodes, which uses a lot of energy. In this project, we describe a method for weather forecasting that uses historical data from a number of weather stations to train basic machine learning models, which can quickly and accurately anticipate specific weather conditions for the near future. The models may be used in contexts with a lot fewer resources. The evaluation's findings demonstrate that the models' accuracy is sufficient to be employed in conjunction with the most cutting-edge methods currently available. Furthermore, we demonstrate that using data from numerous nearby weather stations is preferable than using data from only the area for which weather forecasting is being done.

By gathering real-time meteorological data, the field of weather forecasting uses modern technology to forecast the atmospheric conditions for a specific region. Various weather forecasting instruments, including as satellites, balloons, aeroplanes, buoys, ground stations, and radar systems, are used to update information of the present atmospheric condition. Assimilation is the process by which the state-collected data is transformed into a numerical representation. Climate monitoring, drought detection, agriculture and production, the energy business, the aviation industry, communication, pollution dispersion, etc. all depend on accurate weather predictions. Data mining is a method for predicting the weather. Data mining is the process of extracting a set of data that may be used to make predictions and determine the relationship between various characteristics. Data analysis and rule-based weather prediction are both possibilities provided by data mining. The data are organised and kept in databases.Classification and clustering algorithms are two categories of data mining techniques. A data mining method called classification is used to categorise unidentified samples. Utilising categorization algorithms, rainfall may be simply predicted. A method of grouping items based on information is called clustering. The world's weather is constantly and quickly changing. Today's society relies heavily on accurate projections. We significantly rely on weather forecasts for everything from agriculture to business, transport, and daily commute. In order to maintain simple and smooth movement as well as safe day-to-day operations, it is crucial to anticipate the weather accurately because the entire world is experiencing the consequences of ongoing climate change. The modern weather forecasting methods rely mainly on sophisticated physical models. the application of machine learning to forecast weather in brief time frames using computers with less power is known as "short-term weather prediction." models and need to be run on large computer systems involving hundreds of HPC nodes. Solving the models that describe the climate requires the computing capacity of these big systems. Despite the use of these expensive and sophisticated tools, predictions are sometimes unreliable due to faulty initial observations of the conditions or a lack of knowledge of atmospheric dynamics. Furthermore, solving these kinds of complicated models typically takes a lengthy time. The weather in one location greatly influences the weather in other locations because weather systems may move over great distances and in all directions over lengthy periods of time.

In this research, we suggest a technique to forecast weather by combining historical weather data from nearby cities with local data. We aggregate this data and train basic machine learning models on them so that they can accurately anticipate the weather over the coming days. These straightforward models may be executed on low-cost, less resource-intensive computing platforms while yet producing forecasts that are timely and precise enough for use in our daily lives. We demonstrate in this work that our straightforward model can produce accurate weather forecasts for the city of Nashville in Tennessee, USA, which is renowned for its shifting weather patterns.

This paper's main contributions are as follows:

The use of automated systems to collect historical data from a specialised weather service.

The use of machine learning in the forecast of weather conditions in short periods of time, which may operate on less resource-intensive equipment.

A comprehensive assessment of the suggested method and a comparison of several machine learning models for forecasting future weather conditions.

Artificial intelligence (AI) is the capacity for thought and learning in a computer programme or other system. Making computers "smart" is another goal of this branch of study. As machines' capabilities increase, mental capabilities that were earlier considered to need intelligence are no longer included. The study of artificial intelligence (AI) focuses on building intelligent devices that behave and act like people. Artificially intelligent computers are made for a variety of tasks, such as: face recognition, education, preparation, and decision-making Artificial intelligence is the application of computer science programming to replicate human cognition and behaviour by processing data and the environment, resolving or foreseeing issues, and learning or self-teaching to adapt to a range of activities.

Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for building mathematical models and making predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, Facebook auto-tagging, recommender system, and many more.

Machine Learning is said as a subset of artificial intelligence that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own. The term machine learning was first introduced by Arthur Samuel in 1959. We can define it in a summarized way as: “Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed”.

A Machine Learning system learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it. The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately.

Suppose we have a complex problem, where we need to perform some predictions, so instead of writing a code for it, we just need to feed the data to generic algorithms, and with the help of these algorithms, machine builds the logic as per the data and predict the output. Machine learning has changed our way of thinking about the problem. The below block diagram explains the working of Machine Learning algorithm:

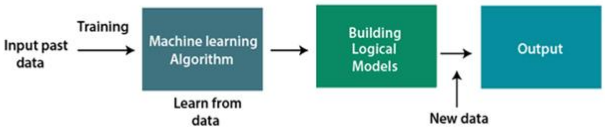


Figure 1. Machine Learning Flow

CLASSIFICATION OF MACHINE LEARNING

At a broad level, machine learning can be classified into three types:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

1) Supervised Learning

A sort of machine learning technique called supervised learning uses sample labelled data to train the machine learning system, which then predicts the outcome.

The system builds a model using labelled data to comprehend the datasets and learn about each one. After training and processing, the model is tested by utilising a sample set of data to see if it accurately predicts the desired outcome.

In supervised learning, mapping input and output data is the main objective. The foundation of supervised learning is monitoring, much like when a pupil is studying under a teacher's supervision. Spam filtering is a prime example of supervised learning.

Supervised learning can be grouped further in two categories of algorithms like Classification and Regression

2) Unsupervised Learning

Unsupervised learning is a type of learning where a computer picks up information without any human intervention. The machine is trained with a collection of unlabeled, unclassified, or uncategorized data, and the algorithm is required to respond independently on that data. Unsupervised learning's objective is to reorganise the input data into fresh features or a collection of objects with related patterns.

There is no predefined outcome in unsupervised learning. The computer searches through the vast volume of data for helpful insights.

It may also be divided into two types of algorithms like Clustering and Association

NAVIES BAYES

The Nave Bayes algorithm is a supervised learning method for classification issues that is based on the Bayes theorem. It is mostly employed in text categorization with a large training set. The Naive Bayes Classifier is one of the most straightforward and efficient classification algorithms available today. It aids in the development of rapid machine learning models capable of making accurate predictions. Being a probabilistic classifier, it makes predictions based on the likelihood that an object will occur. Spam filtration, Sentimental analysis, and article classification are a few examples of Naive Bayes algorithms that are often used. The terms Naive and Bayes, which make up the Nave Bayes algorithm, are defined as follows as Naive because it presumes that the occurrence of one trait is unrelated to the occurrence of other features, it is referred to be Naive. For instance, if the fruit is recognised as an apple based on its red, spherical, and sweet fruit, form, and flavour. So, without relying on one another, each characteristic helps to recognise it as an apple. Because it relies on the Bayes' Theorem concept, it is known as the Bayes principle.

Bayes' theorem is also known as Bayes' Rule or Bayes' law, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.

The formula for Bayes' theorem is given as:

$$P(A|B) = \frac{P(B \setminus A)P(A)}{P(B)}$$

Where, P(A|B) is Posterior probability:
Probability of hypothesis A on the observed event B.

Popular machine learning algorithm Random Forest is a part of the supervised learning methodology. It may be applied to ML issues involving both classification and regression. It is built on the idea of ensemble learning, which is a method of integrating many classifiers to address difficult issues and enhance model performance.

According to what its name implies, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead than depending on a single decision tree, the random forest uses forecasts from each tree and predicts the result based on the votes of the majority of predictions. Higher accuracy and overfitting are prevented by the larger number of trees in the forest. The Random Forest method should be used for the reasons listed below. In comparison to other algorithms, it requires less training time. Even with the enormous dataset, it operates effectively and predicts the outcome with a high degree of accuracy. When a significant amount of the data is absent, accuracy can still be maintained. The stages and graphic below can be used to demonstrate the working process:

- **Step-1:** Select random K data points from the training set.
- **Step-2:** Build the decision trees associated with the selected data points (Subsets).
- **Step-3:** Choose the number N for decision trees that you want to build.
- **Step-4:** Repeat Step 1 & 2.
- **Step-5:** For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

II. LITERATURE REVIEW

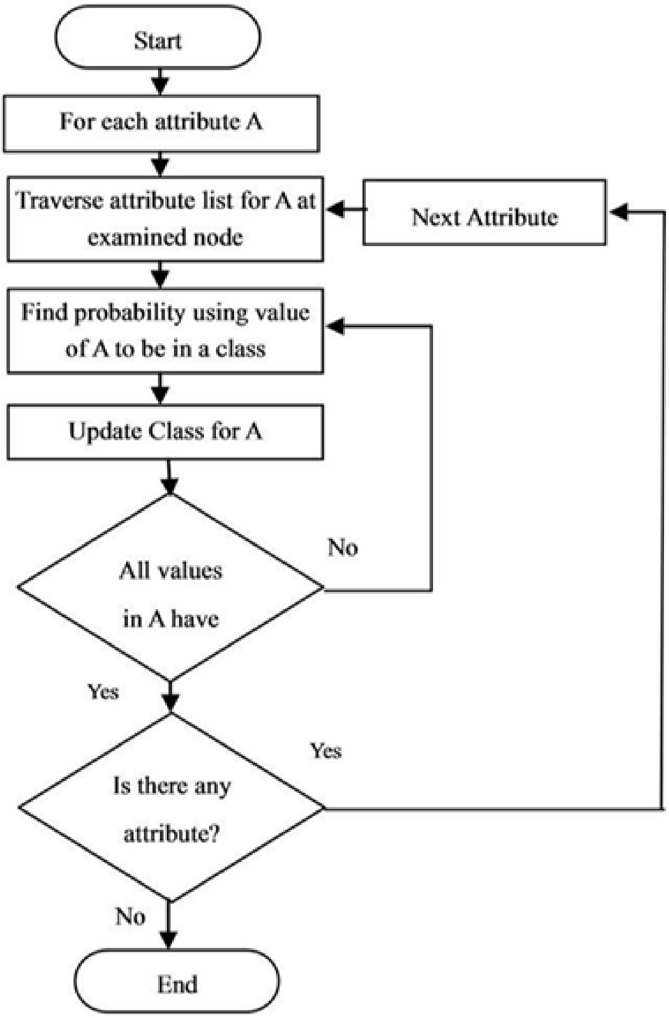
- Support Vector Machine (SVM), neural networks, and other appropriate classification techniques were proposed by Pushpa Mohan et al. [1] for improving classification results. With the use of these methods, it will be possible to anticipate agricultural yields, crop costs, and rainfall.

- The machine learning methods to do weather prediction with data mining approaches for rainfall prediction were evaluated by P. Shivaranjani et al. [2].
- Amruta A. Taksande and colleagues[3] created decision trees and rules using the FP Growth Algorithm to categorise meteorological variables such maximum temperature, lowest temperature, rainfall, humidity, and wind speed in terms of the month and year.
- The National Climatic Center's sensor data was analysed using Hadoop and map reductions by Basvanth Reddy et al. [4] to get improved results for weather prediction in a big data setting.
- K-medoids and the Naive Bayes algorithm were suggested by Prashant Biradar et al. [5] for use in a weather forecasting system that includes variables like temperature, humidity, and wind.The military, navy, marines, agriculture, forestry, and other fields may all use this technology.
- The decision tree algorithm suggested by FolorunshoOlaiya et al[6].Using this algorithm, classification rules and decision trees for meteorological characteristics like maximum and minimum temperatures were created.

III. PROPOSED WORK

- Regression is used in our suggested system since the expected outcomes, in this instance temperature, are continuous numerical values. Given that it ensembles numerous decision trees when making decisions, Random Forest Regression (RFR) is found to be the best regressor. We also compare a number of other cutting-edge ML methods with the Naive Bayes method.
- In this work, we utilise machine learning (ML) techniques to forecast the temperature for the city of Nashville, Tennessee, at any given hour, based on meteorological data from the current day for this city and a few of its neighbouring cities. To construct a single record, we first integrate the weather data at a certain timestamp from all the cities that we take into account. That is, each record in the data will include information on the climate in each city, including the humidity, wind speed and direction, atmospheric pressure, condition, etc. The temperature at the same timestamp the next day is designated as the goal variable for this record.
- As a result, using the weather data from today, we forecast the temperature for the next day. Fortunately, using the same method, we can also forecast every other weather factor, including humidity, rainfall, wind direction and speed, visibility, etc. for the following day and the following several days. In this research, we limit our investigation to just temperature prediction. A model is created using machine learning, a data science approach, from a training dataset.
- Essentially, a model is a formula that generates a goal value based on unique weights and values for each training variable. Each variable's matching weights in each record indicate the model how that variable relates to the goal value (often between 0 and 1). The optimum weights for each variable must be determined using a substantial amount of training data.
- A model may predict the proper output or the target value given a test data record when the weights are learned with the greatest degree of accuracy. We may free ourselves from the intricate and resource-intensive weather models used by conventional weather stations by using straightforward machine learning approaches. It offers a huge amount of potential for weather forecasting. Such a forecasting model is fairly simple to sell to the public as a web service.

IV. SYSTEM ARCHITECTURE



V. IMPLEMENTATION&RESULT

- MODULE 1: Dataset collection
- MODULE 2: Data preprocessing
- MODULE 3: Model Fitting
- MODULE 4: Evaluation

➢ **MODULE 1: Dataset collection**

Source of Data sets:

Datasets are collected from Kaggle website. Once we collect the data, we split the raw data in training and test set. However, the target variable is always the next day hourly temperature for Nashville. Dataset has 9 columns namely

- Summary
- Precis Type
- Temperature
- Apparent Temperature
- Humidity
- Wind Speed (km/h)
- Wind Bearing (degrees)
- Visibility
- Loud Cover.

➢ **MODULE 2: Data preprocessing**

We ensure that each row (record) in the dataset has records for all 10 cities for a given timestamp after receiving the raw data from "wunderground". When building the dataset, we remove any feature with blank or incorrect data. Additionally, we use a method known as "One Hot Encoding" to transform the categorical characteristics in the dataset, such as wind direction and condition, into dummy/indicator variables. Dropping Unwanted Columns:

'Loud Cover': This column has zeros in every row, so this will not impact our model.

'Formatted Date' and 'Daily summary': Summary column existed twice so remove one and no need for Formatted Date column too.

Checking for NULL values: No null values found.

It has been shown that there is a close link between apparent temperature and temperature. So, one column was removed. Almost all columns had outliers, but such numbers are nonetheless conceivable. Let's use the box plot for the Temperature column as an example. Since the category values are of the string form, pre-processing procedures should be used to turn them into numerical values before classifying the Output variable. One of the encoders utilised is the label encoder. We must normalise each column in order to give each variable equal weight.

There are 27 distinct values in the Summary column, which we must forecast, which contains categorical values. Classifiers must and should be used to categorise these category values. Therefore, the two classifiers employed are Random Forest Classifier and Naive Bayes Classifier. There are roughly 12 columns in the dataset, one of which is a dependent variable, while the remaining 11 are independent variables. Some characteristics in the dataset are lowered during pre-processing because they are undesired. After pre-processing, we used the same features to train both

Naive Bayes and Random Forest classifiers. Changed the number of components in Principle Component Analysis (PCA), however the accuracy did not improve. Naive For the sole purpose of supervised classification problems, Bayes is employed to determine the probability for each class of Y. To determine the probability, the Bayes theorem is used. In this case, k is one of the classes of Y and x1, x2, and x3.In the above classifier issue, Xn are the independent variables or the features.

With a 98% accuracy rate, the technology will estimate the soil type. The confusion matrix approach will be used to evaluate the accuracy score.A table known as a confusion matrix is frequently used to illustrate how a classification model (or "classifier") performs on a set of test data for which the real values are known. Accuracy, Recall, Support, F1score, and Precision

VI. CONCLUSION & FUTURE SCOPE

We demonstrated a system in this project that makes use of machine learning methods to generate weather forecasts. Intelligent models may be produced by machine learning technology and are significantly more straightforward than conventional physical models. They are easier to operate on practically any computer, including mobile ones, and are less resource-hungry. Our assessment findings demonstrate that these machine learning models can forecast weather characteristics with sufficient accuracy to rival conventional methods. To forecast the weather in a specific location, we also use historical data from the neighbourhood. We demonstrate that it is superior than focusing simply on the region for which weather forecasts are made.

In the future, we want to collect meteorological data across a city using inexpensive Internet of Things (IoT) devices like temperature and humidity sensors. The training dataset could contain more local information if various sensors are used. The performance of our prediction models will be considerably enhanced by this data and the weather station data.

VII. REFERENCES

[1]. Weather Underground. https://www.wunderground.com/weather/api. 2018

[2]. Weather forecast using the sensor data from your IoT hub in Azure Machine Learning. https://docs.microsoft.com/en-us/azure/iot-hub/ iot-hub-weather-forecast-machine-learning.

[3]. Dan Becker. 2018. Using Categorical Data with One Hot Encoding. https://www. kaggle.com/dansbecker/using-categorical-data-with-one-hot-encoding/.

[4]. Jeffrey Burt. 2017. Machine Learning Storms into Climate Research. https://www. nextplatform.com/2017/04/18/machine-learning-storms-climate-research/.

[5]. Aditya Grover, Ashish Kapoor, and Eric Horvitz. 2015. A deep hybrid model for weather forecasting. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 379–386.

[6]. Mark Holmstrom, Dylan Liu, and Christopher Vo. 2016. Machine Learning Applied to Weather Forecasting.

[7]. Vladimir M Krasnopolsky and Michael S Fox-Rabinovitz. 2006. Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction. Neural Networks 19, 2 (2006), 122–134.

[8]. Federico Montori, Luca Bedogni, and Luciano Bononi. 2017. A Collaborative Internet of Things Architecture for Smart Cities and Environmental Monitoring. IEEE Internet of Things Journal (2017).

[9]. Y Radhika and M Shashi. 2009. Atmospheric temperature prediction using support vector

machines. International journal of computer theory and engineering 1, 1 (2009), 55.

[10].Sund. 2015. Using Amazon Machine Learning to Predict the Weather. https://arnesund.com/2015/05/31/ using-amazon-machine-learning-to-predict-the-weather.

Cite this article as :

M.Robinson Joel, G.Manikandan, M. Nivetha, " Marine Weather Forecasting to Enhance Fisherman's Safety Using Machine Learning, International Journal of Scientific Research in Science, Engineering and Technology(IJSRSET), Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 10, Issue 2, pp.519-526, March-April-2023. Available at doi : https://doi.org/10.32628/IJSRSET2310288
Journal URL : https://ijsrset.com/IJSRSET2310288