Analysis On Weather Forecast Prediction Using Sprint Algorithm

Vijay Koraveni
Asst. Prof.
Department of CSE
Siddhartha Institute of Technology and Sciences
Narapally, Hyderabad, Telangana, India

Mamidi Ranjith Reddy
Asso. Prof.
Department of ECE
Siddhartha Institute of Technology and Sciences
Narapally, Hyderabad, Telangana, India

Abstract - Weather forecasting is a new domain that predicts the weather at a specific location at a specific time. Weather forecasting is regarded as the most sensitive scientific topic, with numerous real-time concerns such as erroneous prediction, inability to handle large amounts of data, and a lack of technological innovation. In this study, we present the SPRINT method, which is based on the decision tree principle. The experiments are conducted using a climate dataset and the WEKA tool. The data is classed into sunny, gloomy, and rainy based on climate criteria such as Outlook, Temperature, Humidity, and Windy. The weather is predicted based on the acquired results, demonstrating the proposed approaches' accuracy level. The current approach navie Bayes is used to compare performance, and both findings are presented on a graph. The results show that the SPRINT algorithm is effective and accurate at forecasting weather conditions.

Keywords: Data mining, WEKA tool, SPRINT algorithm, Weather forecasting and climate data

1. INTRODUCTION

The importance of weather forecasting in evaluating and predicting the atmosphere at a specific area at a certain time cannot be overstated. In this approach, climatology is a typical scientific study of weather conditions and climate projections. Weather forecasting has been plagued by technical challenges all across the world in recent years [1]. The main cause of this stumbling block is frequent environmental changes. There are various limits to conducting weather forecasting that have persisted with some efficacy in recent years [4][8]. Weather forecasting is critical in meteorology, especially for meteorologists, and is regarded as the most difficult aspect of accurate prediction. Predicting nature's weather warnings is obviously more vital in ensuring human life and property for the benefit of human life. The weather forecast is critical in farming, with wind, temperature, humidity, and outlook being the most important aspects to consider. The majority of utility businesses use temperature forecasting to assess the future. Climate change has resulted in significant rain, snow, and wind chill, all of which have a negative influence on outdoor activities. Such foresight allows people to plan and design activities based on the events that affect their existence. Failure of exact weather forecasts in hazardous situations places an individual in a risky scenario [1].

The evaluation and processing of large volumes of weather datasets is a fundamental difficulty in weather forecasting. Several data mining approaches are used to do this. Data mining techniques are used to analyze the data, classify it, and summarize the connections found. Classification, Learning, and Prediction are three terms that are used in the data mining process. Classification is a machine learning process used in data mining to forecast and collect information situations. For example, according to the weather on a certain day, whether forecasting datasets classification is "sunny," "rainy," or "cloudy" [1]. Learning is the process of training and mapping a dataset in order to obtain a set of related data. There are two types of learning: supervised and unsupervised learning. The Classifier is used in a supervised learning algorithm to analyse and supply the derived capacity from the training data [17]. Unsupervised learning is concerned with issues of hidden structure in unlabeled data [1]. The learner is provided the unlabeled patterns, which have little chance of yielding a potential solution. Prediction establishes the model's logical linkages on the basis of previous, current, and future data. Finding patterns and data is critical for making quick and accurate predictions [17].

2. RELATED WORK

Weather Forecast Prediction Using an Integrated Approach for Analyzing and Measuring Weather Data was proposed by Munmun Biswas et al [1]. The author uses predictive analysis to forecast the weather in this paper. The Naive Bayes and Chi-square algorithms are used for classification. A web

application is used to forecast the weather, and the user inputs information such as the current forecast, temperature, humidity, and wind condition. The system forecasts the weather based on the input.

Weather Forecasting Using Frequent Pattern Growth Algorithm was proposed by Amruta A. Taksande et al [2]. The major goal of this project is to forecast rainfall. The author used a five-year dataset for Nagpur station from January 2010 to January 2014, which he processed using the Frequent Pattern Growth Algorithm. Temperature, humidity, and wind speed are all factors that go into predicting rainfall. Approximately 90% is achieved based on this observation.

Weather Forecast Using Data Mining Research Based on Cloud Computing was proposed by Zhan Jie Wang and A. B. M. MazharulMujib et al [3]. The author used meteorological data acquired at a specific period to conduct his experiment, which he observed using Artificial Neural Networks and Decision Tree Algorithms. The obtained result demonstrates that the proposed methodology produces good weather predicting results by combining mean weather variables with generating categorization rules.

SushmithaKothapalli et al [4] proposed the Auto-Regressive Integrated Moving Average (ARIMA) model for Real-Time Weather Forecasting and Analysis. The author of this experiment used a dataset that included data on temperature, wind, humidity, and rainfall. CSV, JSON, and XML file formats are used to store the acquired data on the cloud. The ARIMA model, when combined with parameter correlation analysis, yields superior weather forecasting predictions.

Weather forecasting using deep learning techniques was proposed by AfanGalih Salman et al [5]. The major goal of this project is to uncover a hidden hierarchical pattern in the meteorological data. The author used the meteorological dataset provided by BMKG for the experiment (Indonesian Agency for Meteorology, Climatology, and Geophysics). Recurrence Neural Network (RNN), Conditional Restricted Boltzmann Machine (CRBM), and Convolutional Network (CN) models are observed in this dataset. The obtained results accurately forecast weather, which is beneficial to agriculture and tourism in a peaceful manner.

Shyi-Ming Chen et al [6] proposed using fuzzy time series to predict temperature. The main purpose of this solution is to solve predicting issues based on historical data. The author uses a two-factor time-variant fuzzy time series model to solve forecasting difficulties in this research. As a result of this method, good forecasting results were obtained.

For weather forecasting, Imran Maqsood et al [7] developed an ensemble of neural networks. The author gave a comparison of works of an ensemble of artificial neural networks in this paper (ANNs). Temperature, wind speed, and relative humidity data are the main parameters used in this evaluation. Hopfield model (HFM) predictive models, Elman recurrent neural network (ERNN), radial basis function network (RBFN), multi-layered perceptron network (MLPN), and regression approaches were employed in this experiment. The comparative results suggest that RBFN is fair in dealing with weather forecasting problems when compared to others, despite HFM's less accurate performance.

3. PROPOSED METHODOLOGY

Weather forecasting is a broad topic that is vital to people, but there has been little research in this area. As a result, the objective is to present an effective research effort on weather forecasting. Some valuable observations have already been stated in some works. This is our contribution to the research to see how far technology has progressed in weather forecasting. In this study, we propose the SPRINT algorithm for evaluating large climatic data, which is implemented in the Weka tool. We use the dataset from Weka open source, a well-known data source in the research field, for this experiment. The performance of our suggested work is compared to existing work [1] to demonstrate its efficacy. Let's take a closer look at the proposed mechanism.

3.1 Data mining Tools

There are various data mining technologies available in today's technological world. Weka is one of the top open-source tools for research and is widely utilized. Weka is capable of performing standard operations such as data preparation, clustering, classification, regression, visualization, and feature selection. Weka predicts the possible assumption in a better way based on the attributes and their relationships from the data. Weka also supports numerous file formats and numeric or nominal properties, which is a useful feature. Java Database Connectivity is available in Weka, allowing direct access to SQL databases and response to queries. Using different software, a set of linked database tables is merged into a single table for multi-relational data mining. Furthermore, Weka is simple to use and master, making it one of the greatest mining tools in the real-world scenario.

3.2 Data preprocessing

The process of separating raw data into a comprehensible format is known as data preparation. In general, datasets contain both usable and noisy data, and it is important to extract relevant data from them. Preprocessing is the first step in data mining

that saves time and aids in the retrieval of hidden features. Incomplete, inconsistent, and error data are jotted down and eliminated in detail during preprocessing. As a result of this procedure, clear and informative data is obtained, which will be used in subsequent procedures.

3.3 Decision Tree

Decision Trees (DT) are important in data mining. The decision tree lies at the heart of our proposed SPRINT algorithm. With appropriate classification rules, the decision tree plays an important role in weather prediction. In recent years, it has been claimed that DT is the most widely used method for interpreting meteorological data. The neuron model in DT is made up of three parts: summary layout, interfacing links, and consideration based on its own weight/strength. In general, qualities are quite significant in weather forecasting, and what class these attributes will accomplish. The qualities and classes that apply to this study are listed in the table below.

S.no	Attribute	Class
1	Outlook	Sunny, Overcast, Rainy
2	Temperature	High, Mild, Cool
3	Humidity	High, Normal
4	Windy	True, False

3.4 SPRINT algorithm

The SPRINT algorithm is a Decision Tree idea that represents data in a tree structure. The advantage of adopting the suggested SPRINT method is that it allows for early classification. In comparison to other algorithms, the SPRINT algorithm has various distinguishing characteristics as well as improvable measurements that are most closely related to cloud technology. Because cloud computing can handle large amounts of data, it is a viable option. This is the important phrase in this study paper for the evolution of the SPRINT algorithm.

3.5.2 SPRINT workflow

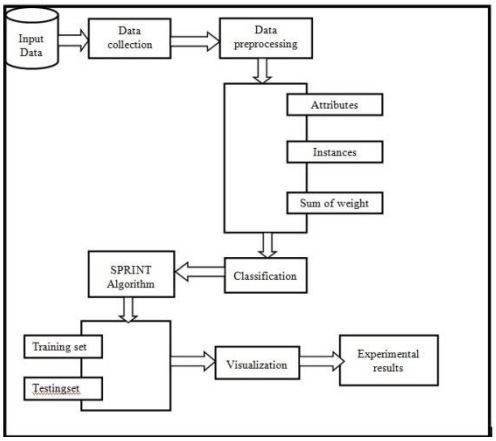

Fig2: Proposed method Workflow

The input data are given first, and then the acquired data is preprocessed. Noisy data is eliminated during preprocessing, and classification rules are used. The SPRINT categorization rule is then applied based on attributes, instances, and the sum of weights. The SPRINT method then starts to build the tree, with each leaf node forming a tree structure with linked data based on the specifications. The rainy-R, overcast-O, and windy-W weather models are used to identify whether it is windy, cloudy, or sunny. The generated data is put in the database if SPRINT correctly checks the weather situation. Finally, sensitivity, specificity, and accuracy are determined using the observation confusion matrix.

4. EXPERIMENTAL RESULTS

WEKA was used to conduct the experiment with a climatic dataset. This observation was based on a dataset obtained from an open-source site. The performance of our suggested system is determined using the WEKA tool. The acquired result is compared with the navie Bayes mechanism [1] to establish the efficiency of our suggested SPRINT method, and the decision table produces the cross-validation summary, and the comparison is done on three metrics such as accuracy, sensitivity, and recall. The parameters used for evaluation, such as TP Rate, FP Rate, Precision, Recall, F-Measure, MCC, ROC Area, and PRC Area, are also observed. The computed values of these parameters utilizing our proposed method are 94 percent more efficient than the others.

5.CONCLUSION

This research examined the issue of weather forecasting in the context of big data analytics. We used the climate dataset from the official WEKA site, which is an open-source WEKA climate dataset, to apply our suggested SPRINT algorithm for precise

weather prediction. It employs the decision tree principle in the development and construction of a tree model based on relevant data. The observation is carried out with WEKA, an innovative data mining tool that allows direct access to SQL databases, to better the research. For comparing historical data, the proposed methodology is ideal. The proposed framework is useful for assessing the relationship between climate variables like temperature, wind speed, humidity, and so on. To demonstrate the efficacy of our suggested mechanism, we compare it to the SPRINT algorithm and Navie Bayes (Existing work [1]). The acquired result demonstrates that the SPRINT algorithm's prediction accuracy is considerably superior than the existing methods.

6. REFERENCES

1) Munmun Biswas, TanniDhoom, and SayantanuBarua "Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data", Article in International Journal of Computer Applications · December 2018 DOI: 10.5120/ijca2018918265
2) Amruta A. Taksande, P. S. Mohod, “Applications of Data Mining in Weather Forecasting Using Frequent Pattern Growth Algorithm”, International Journal of Science and Research (IJSR), Volume 4 Issue 6, June 2015, pp 3048-3051
3) Wang, ZhanJie&Mujib, A B M. (2017). “The Weather Forecast Using Data Mining Research Based on Cloud Computing”. Journal of Physics, pp 1-6
4) SushmithaKothapalli, S. G. Totad, “A Real-Time Weather Forecasting and Analysis”, IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017), pp 1567- 1570
5) A. G. Salman, B. Kanigoro and Y. Heryadi, "Weather forecasting using deep learning techniques," 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2015, pp. 281-285.
6) Chen, S.-M., and J.-R. Hwang. "Temperature prediction using fuzzy time series." Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 30.2 (2000): 263-275.
7) Maqsood, I., M. R. Khan, and A. Abraham. "An ensemble of neural networks for weather forecasting." Neural Computing & Applications 13.2 (2004): pp 112- 122.
8) Almgren, K.; Alshahrani, S.; Lee, J. Weather Data Analysis using Hadoop to Mitigate Event Planning Disasters. Available online: https://scholarworks.bridgeport.edu/xmlui/handle/123456789/1105 (accessed on 1 February 2019).
9) Oury, D.T.M.; Singh, A. Data Analysis of Weather Data Using Hadoop Technology. In Smart Computing and Informatics; Springer: Singapore, 2018; pp. 723–730.
10) Manogaran, G.; Lopez, D. Spatial cumulative sum algorithm with big data analytics for climate change detection. Comput. Electr. Eng. 2018, 65, 207–221.