International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Optimized K Nearest Neighbor Classification Algorithm for Weather Prediction

A Zakiuddin Ahmed, T.Abdul Razak

Abstract—Weather has a lot of blow in our daily life and also gained researchers concentration due to its enormous effect in the human life. To defend ourselves from weather, we need to predict the weather such as rainfall, humidity and temperature etc. Using classification algorithms, we can predict the weather by using the past datasets. In this research paper, WEKA tool is used to implement classification algorithms for weather forecasting. Machine Learning is an internal part of artificial intelligence, which is used to design algorithms based on the relationships between data and data trends.

Keywords: Machine Learning, KNN, Random Forest, Decision Tree, WEKA, Filtering.

I. INTRODUCTION

Machine Learning Technique is an inner part of artificial intelligence. In ML, computer learns routinely from data and information using different computer algorithm. Computer does not need to unequivocally programmed. These can be improved and change algorithm by themselves. There is a growing need of Machine Learning among companies for professionals and it is used all over the world. It will help us to understand the ins and outs of Machine Learning.

Clustering and classification techniques are mostly used in various applications, such as pattern recognition, image processing, market research and data analysis. In marketing fields customers purchase pattern can be discovered with the help of clustering and classification techniques. These techniques can be supportive in classifying documents available on the Web for information discovery purposes.

Conventional weather forecasting relies on a grouping of weather observations and data models. Meteorologists generate weather forecasts by gathering as much data as promising and then dealing out it all the way through weather prediction models. Meteorologists develop forecasts based on models from weather agencies, as well as models formed by private weather forecasters.

Manuscript published on November 30, 2019.
* Correspondence Author
A Zakiuddin Ahmed*, Research Scholar, Department of Computer Science, Jamal Mohamed College(Autonomous), Tiruchirapalli, Affiliated to Bharathidasan University, Tiruchirappalli, Tamilnadu, India. (Email: azaki2020@gmail.com)
Dr.T.Abdul Razak, Associate Professor, Dept. of Computer Science, Jamal Mohamed College(Autonomous), Tiruchirappalli, Affiliated to Bharathidasan University, Tiruchirappalli, Tamilnadu, India. (Email: abdul1964@gmail.com)

Steps involved in Machine Learning

1. Collecting the data
2. Preprocessing the data
3. Selecting a model
4. Training the data
5. Evaluation
6. Tuning the parameter and
7. Result prediction.

Algorithms in Machine Learning

1. Logistic Regression
2. KNN Algorithm
3. Random Forest Algorithm
4. Linear Regression Algorithm
5. Support Vector Machine Algorithm
6. Decision Tree Algorithm
7. Gradient Boosting Algorithms
8. K-Means Algorithm
9. Naive Bays Algorithm and
10. Dimensionality Reduction Algorithms.

Applications of Machine Learning

1) Social Media Services
2) Virtual Personal Assistants
3) Email Spam and Malware Filtering
4) Predictions while Commuting
5) Videos Surveillance
6) Online Customer Support

Weather Forecasting

Weather dataset for the prediction models comes from outside observations provided by millions of computerized weather stations around the world, as well as from radars and satellites. Surface surveillance includes atmospheric data such as temperature, wind speed, humidity, and precipitation. Meteorologists then put the observed data into weather models to create weather forecasts. Advancements in computing power, such as the use of supercomputers, which means that weather forecasts have continued to become even more accurate.

Making the most accurate weather predictions, particularly regarding precipitation, requires more than surface observations. This is where radar and satellite come into the picture. Weather radars work by transmitting radio waves into the atmosphere, and the waves bounce off of objects like rain and snow, thereby informing radars the exact location of precipitation.

II. RELATED WORK

Qing Yi Feng1, et.al [1], in this research paper, they used machine learning toolbox that is purely based on climate data gathered for analysis and reconstruction of complex networks. It can also handle data containing multiple variables from these networks.

S.Siddharth et.al [2], they have used Decision Tree algorithm as a means to classify weather parameters like maximum temperature, minimum temperature in terms of day, month and year for the purpose of predicting the weather condition of a specific region.

Sanyam Gupta, et.al [3], they proposed an efficient method for weather prediction using linear regression concepts and normal equation model. The normal equation is a very efficient weather prediction model and using the entities such as temperature, humidity and dew-point, it can be used to make reliable weather predictions. It can yield better results when applied to cleaner and larger datasets.

Muthulakshmi A, ME (SE), et.al [4] in their work they proposed a methodology that aims to provide the models to predict rainfall from the previous weather datasets. In their previous work, the parameters of weather were recorded only for the present time. But in their proposed model, they make a working model of selection that can be helpful for classifying the framework that is used in continuous monitoring of the climatic attributes.

Ashish kapoor, et.al [5] in their work made a weather prediction model that predicts by considering the combined influence of key weather variables. They also made a kernel and showed that interpolation of space can be made by using GPS with such a kernel, taking into account various weather phenomena. They also performed time related analysis within a learner based on gradient tree and augmented the system using deep neural network.

John Williams et.al [6] has shown in their work that by using the Random Forest Machine Learning Algorithm a set of skillful predictors for thunderstorm initiation can be identified. To identify “regimes” the random forest method used, in which they can improve the skill of the application by using forecast logic.

Intelligent Weather Predictions

Technological advancements in the 21st century have brought many improvements to weather forecasting. The growth of smart phones has brought on-the-go weather forecasting to billions of people around the world, while the location data of the devices improves the accuracy of forecasting. Another recent development, the AI revolution, has not spared weather prediction either. Developments in Machine Learning mean that AI can be incorporated into existing weather models to produce even more accurate forecasts. Machine Learning Models for weather forecasting quickly process large amounts of weather data, and they can compare data from weather stations and satellites with traditional forecasts to make highly accurate predictions.

III. PROPOSED MODEL

Optimized K Nearest Neighbor classification algorithm is the proposed model in this research paper which helps us to accurately predict the temperatures, i.e. both low and high temperatures. Therefore initially in our work, only eight parameters are selected for use which are max. temperature, min. temperature, humidity, atmospheric etc. The second algorithm looks for the previous weather patterns which are similar to the present day weather patterns, then it predicts the future weather patterns based on the data of the trained dataset which was collected from Kaggle.

proposed algorithm

- Dataset: Weather
- Instances: 3000
- Attributers: 8

Existing Methods

- Decision Tree
- Random Forest

Proposed Method and Findings

- Optimized KNN

Stage 1: Preprocessing with filter
Stage 2: Optimized KNN through Clustering approach

- Optimized KNN provides better accuracy to predict weather.

Step 1.Training the model with x(i) where x(i)∈ W. W={ mxT, mnT, meanH, meanAP} from Kaggle dataset.

Step 2.Evaluate and optimize the model with test dataset by changing kernel and Gaussian function values.

Step 3.a) 7-fold cross validate the model with blind dataset.

Step 3.b) Minimize the cost function.

Step 4.Input past two weeks mxT and mnT (after the model is ready).

Step 5.Predict the mxT and mnT for the future.

IV. RESULT AND DISCUSSION

In this research paper weather dataset collected from Kaggle repository is implemented with help of machine learning tool WEKA. In this paper decision tree, random forest and proposed KNN algorithms are used. Decision tree and random forest algorithms are used directly, but the KNN algorithms is modified as Optimized KNN Algorithms for the purpose of predicting the accurate weather forecasting. The Accuracy, Precision and Recall are calculated using the following formulae and the results are furnished in the below given table.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \ldots\ldots\ldots (1)$$

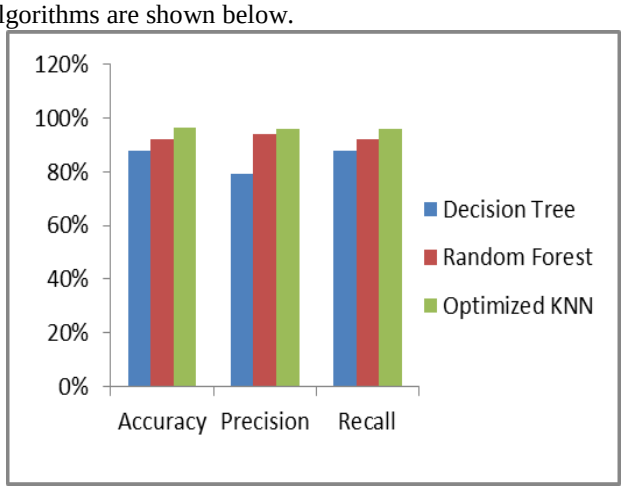
$$Precision = \frac{TP}{TP+FP} \ldots\ldots\ldots (2)$$

$$Recall = \frac{TP}{TP+FN} \ldots\ldots\ldots (3)$$

S. No	Machine Learning Algorithms	Accuracy	Precision	Recall
1	Decision Tree	88%	0.79	0.88
2	Random Forest	92%	0.94	0.92
3	**Optimized KNN**	**96.48%**	**0.96**	**0.96**

The classification Algorithms namely Decision Tree and Random Forest are the existing Algorithms for this research paper and the Optimized K Nearest Neighbor is the proposed Algorithm.

The comparison charts for the existing and proposed Algorithms are shown below.



Comparison of Algorithms for Weather prediction

V. CONCLUSION

Weather prediction for opportunity based on the duration with additional parameters involves the use of artificial Neural Network. To make sure the justification of prediction model for weather conditions, mutually systems are compared to check the suitability of applicability. In this research work, Decision Tree, Random Forest and optimized KNN Techniques are used to predict the weather parameters. The same dataset is used in both the Classification Algorithms, i.e., in the existing Decision Tree, Random Forest Algorithms and proposed optimized K Nearest Neighbor Algorithm, So that a comparative analysis could be made easily. Finally the proposed optimized K Nearest Neighbor classification Algorithm achieves superior results compared to other Classification Algorithms for enhanced weather prediction.

REFERENCES

1. Qing Yi Feng1, Marc Segond , AviGozolchiani, Yang Wang , Markus Abel ,ShilomoHavlin, Armin Bunde, RuggeroVasile and Henk A. Dijkstra1 "ClimateLearn: Machine Learning approach for climate prediction using network measures ", Germany, 2017.
2. S.Siddharth, Bhatkande1 ,Roopa G. Hubballi "Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques", Belgaum India: Inter. Journal of Advanced Research in Computer and Communication Engineering (IJARCE), 2017.
3. Sanyam Gupta, Govind Singhal and Indumathy "Using Normal Equation Method and Linear regression Techniques for weather prediction" , International Journal of Computer Science and Information Technology (IJCSIT), 2016.
4. Muthulakshmi A, ME (SE), Dr.S.Baghavathi Priya " A survey on weather forecasting to predict rainfall using big data analytics", IJISET , 2016.
5. Ashish Kapoor , Eric Horvitz and Adiyta Grover "A Hybrid Model for Weather prediction ",:MicrosoftResearch, Redmond.
6. John Williams and D. A. Ahijevych, C. J. Kessinger, Steiner. M, Saxen .T.R , S. Dettling, National Center for Atmospheric Research, Boulder, Colorado " Finding Weather Regimes And Skillful Predictor Combinations for Short-Term storm forecasting", Colorado.