Improving Short Term Weather Forecast Accuracy with Machine Learning Models

Jaibhagwan[1], Surender Singh[2]
[1]*Student, Department Of CSE, BRCM CET, Bahal, Bhiwani, Haryana, India*
[2]*Assistant Professor, Department Of CSE, BRCM CET, Bahal, Bhiwani, Haryana, India*
Corresponding Author: jaibhagwancse3950@gmail.com

Abstract— **Traditionally, weather predictions rely on large, complex physical models that incorporate various atmospheric conditions over extended periods. Weather system disturbances frequently cause these conditions to become unstable, which results in forecasts that are off. These models usually require massive amounts of energy, running on hundreds of nodes in a big HPC (High-Performance Computing) environment. In this research, we present a method for weather prediction that trains basic ML (Machine Learning) models using historical data from various weather stations. These models can function in far less resource-intensive contexts and produce useful forecasts for specific weather conditions in the near future in a relatively short amount of time. The evaluation findings show that these models' accuracy is adequate to supplement the most advanced methods available today. We also show the benefits of using weather station data from several nearby locations instead of depending only on data from the area under forecast.**

Index Terms—**Weather forecast, Machine Learning, data preprocessing.**

1. Introduction

Weather conditions change rapidly, making accurate forecasts crucial for various aspects of daily life, from agriculture to commuting. High-performance computing (HPC) systems and intricate physical models are the foundation of current weather prediction models. These systems consume a lot of energy and frequently generate forecasts that are off because of faulty initial data and a lack of knowledge of atmospheric processes.

This research suggests training basic machine learning models for more precise and timely weather predictions using historical weather data from cities adjacent in addition to data from the target city. These models can produce accurate projections even when they operate on systems with fewer resources. A case study in Nashville, Tennessee, demonstrates the model's effectiveness.[1]

Manuscript revised June 11, 2024; accepted June 13, 2024. Date of publication June 15, 2024.
This paper available online at www.ijprse.com
ISSN (Online): 2582-7898; SJIF: 5.59

The paper's key contributions include:

- Using machine learning for quick, resource-efficient weather predictions.
- Implementing automated systems for collecting historical weather data.
- Thoroughly evaluating and comparing various machine learning models for weather forecasting.

2. Background

This section offers a concise overview of the application of machine learning in weather forecasting and discusses relevant literature in this rapidly evolving field.

A. Machine Learning for Weather Forecasting

ML, a data science technique, constructs models from training data. These models, essentially formulas, generate target values based on individual weights assigned to each training variable. The weights, often ranging between 0 and 1, indicate the variable's relationship to the target value in each record. It is essential to have enough training data in order to identify the ideal weights for each variable. A model can correctly anticipate the target value or correct output for a given test data record once it has been trained.[2]

Employing simple machine learning methods offers relief from the complexity and resource demands of traditional weather station models. This approach holds great promise in weather forecasting, with the potential for such models to be easily accessible to the public through web services.

3. Related Work

Studies by various authors explore the integration of ML "into weather forecasting, reflecting a recent trend in the literature.

Holmstrom et al. [3] developed a technique to predict maximum and minimum temperatures for the next 7" days on the basis of the preceding 2 days' data, employing linear and functional linear regression models. In short-term projections, professional forecasting services outperformed their models, but in longer-term forecasts, their models performed better.

To integrate physics into weather forecasting, "Krasnopolsky and Rabinivitz[4] presented a hybrid model that makes use of neural networks.

Support vector machines were used by Radhika et al. [5] to solve a classification problem involving weather" prediction. A different study presented a data mining-based forecasting model that uses k-means clustering and Hidden Markov Models to identify changes in weather patterns.

In order to estimate joint weather-related variable statistics, Grover et al. [6] looked into a hybrid strategy that combined deep neural networks with discriminatively trained predictive models.

Montori et.al.[7] explored crowdsensing, where users share smartphone data for environmental monitoring, introducing the SenSquare architecture for unified data display "from IoT sources and crowdsensing platforms, applicable to smart city environment monitoring. Still, none of" this research took into consideration the possibility of pooling data from nearby sites.

A. Machine Learning Techniques

In this study, focusing on continuous numeric predictions like temperature, we employ regression analysis. Our analysis reveals Random Forest Regression (RFR) as the optimal regressor, leveraging ensemble learning from multiple decision trees. Furthermore, we compare RFR with several other leading ML techniques, including Ridge Regression (Ridge), Multi-layer Perceptron Regression (MLPR), Support Vector Regression (SVR), and Extra-Tree Regression (ETR).

B. Dataset

After gathering data, we divide the raw data into training and test sets, always aiming for Nashville's hourly temperature the following day. 2 months' worth of weather data, starting on July 1, 2018, make up the training set. On the other hand, the seven-day test set spans from September 1 to September 7, 2018. Notably, using test input data from September 1st, the trained model predicts the hourly temperature for September 2nd. Subsequently, predictions for September 3rd rely on September 2nd data, and so forth.

4. Conclusion And Future Work

The paper introduces a technology harnessing machine learning for weather forecasting, offering simpler and less resource-intensive models compared to traditional methods. Evaluation results demonstrate these models' competitive accuracy. Additionally, the study incorporates historical data from neighboring areas to enhance forecasting effectiveness. Future plans involve leveraging low-cost IoT devices to gather weather data from various city locations, aiming to augment training datasets and improve prediction model performance.

References

[1]. Arne Sund. 2015. Using Amazon Machine Learning to Predict the Weather.

[2]. 2018. Weather forecast using the sensor data from your IoT hub in Azure Machine Learning.

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

[4]. Vladimir M Krasnopolsky and Michael S Fox-Rabinovitz. 2006. Complex hybrid models combining deterministic and machine learning components for numerical.

[5]. Y Radhika and M Shashi. 2009. Atmospheric temperature prediction using support vector machines. International journal of computer theory and engineering 1, 1 (2009), 55.

[6]. 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.

[7]. 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).