iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 11 **Issue:** I **Month of publication:** January 2023

**DOI:** https://doi.org/10.22214/ijraset.2023.48718

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 11 Issue I Jan 2023- Available at www.ijraset.com*

# Machine Learning Methods to Weather Forecasting to Predict Apparent Temperature: A Review

Saurabh Nayak[1], Priyanka Dubey[2]
[1, 2]*Research Scholar SKU, Chhatarpur*

***Abstract:*** ***Predicting the atmosphere's state for a future date and specific place is the problem of weather forecasting. Traditionally, the atmosphere has been treated as a fluid in physical simulations to accomplish this. The equations for fluid dynamics and thermodynamics are numerically solved to determine the current state of the atmosphere and its future state. However, the system of ordinary differential equations that governs this physical model is unstable under perturbations, uncertainties in the initial measurements of the atmospheric conditions, and a lack of understanding of complex atmospheric processes govern the range of accurate weather forecasting to a 10 day period, beyond which weather forecasts are significantly unreliable.***
***Contrarily, machine learning doesn't need a thorough understanding of the physical processes that govern the atmosphere and is more resilient to perturbations. As a result, machine learning could be a good substitute for physical models in weather forecasting.***

## I. INTRODUCTION

Big Data is made up of massive amounts of organized, semi-structured, and unstructured data. This is why such data is so difficult to manage, track, and store. To cope with large data, there are currently a plethora of strategies, tactics, and procedures available. In this study, we track weather-related data and use data mining, including machine learning, to make predictions about the path and likely repercussions of the weather.

We propose employing data mining and machine learning to systematically retrieve data before making climatological and meteorological predictions in this way. Extreme weather, pollution, and the consequences have lately escalated in India. Ranchers frequently face issues in the horticulture business due to unexpected weather patterns.

The sole component utilized to anticipate the weather is the distribution of naturally occurring air particles such as ozone, nitrogen dioxide, carbon dioxide, sulfur dioxide, and others. There are a variety of ways and computations that we can use to forecast the climate based on the information, which can help to lessen these responses to some level. Machine learning-based data mining is employed in the weather prediction process.

Weather is vitally crucial at all stages of human growth. As a result, climate prediction is becoming increasingly important in many disciplines, including research, agriculture, and the management of food security problems. There was never a reliable source for weather information in the past. There were various issues at the time in the agriculture, industrial, and food supply chains. However, because we live in a highly advanced civilization, there are several ways to learn about the weather. To determine current weather conditions, scientists are using Big Data and its Eco-System, as well as machine learning approaches such as support vector machines and linear regression.

## II. EVALUATION OF SOLAR FORECASTING

Historically, A majority all power grid variability is caused by swings in demand because conventional power production technologies like fossil and nuclear energy were built to operate in steady output modes. However, as seen in Figure, the solar resource occasionally displays a high level of unpredictability.

According to recent studies, ISOs require precise projections of solar irradiance over a range of time horizons in order to promote greater market or grid penetration for solar electricity. Although this work describes a vast number of forecasting techniques, the comparison of outcomes and assessment of relative.

©IJRASET: All Rights are Reserved 

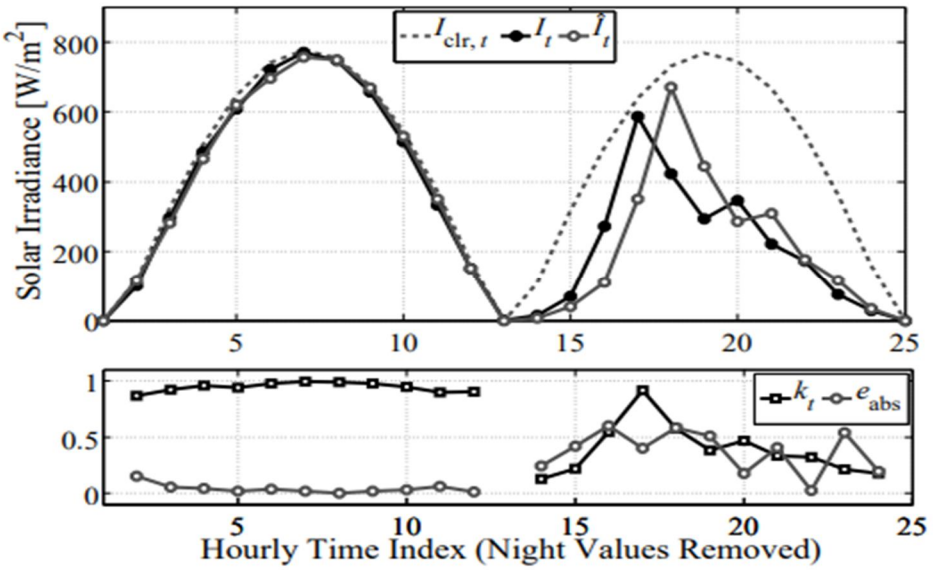



Figure 1 Example of persistence model performance for a clear and cloudy day.

On the clear day, the persistent model performs reasonably well. However, during the gloomy day, the persistence model exhibits significant errors that result from 'time delays' and sudden changes in observed irradiance.Clear sky index kt and absolute error eabs = |I − ˆI|/I of The identical time steps for the persistent model are displayed in the graph's lower section.

Comparative model advantages have remained elusive. These issues result from the fact of solar irradiance is innately dependent on place, season, and climate, as well as the numerous evaluation methods that different authors have used to rate the accuracy of their models. The coefficient of determination, which compares the variance of an error to the variance of the data to be modelled, is one of the traditional statistical measures to assess the accuracy of a model.

$$R^2 = 1 - \frac{\mathrm{Var}(\hat{I} - I)}{\mathrm{Var}(I)},$$

## III. LITERATURE REVIEW

Gil, D et al 2019.the impact of weather & air pollution on patients with respiratory diseases who attended emergency departments. Methods. The National Emergency Department Information Systems (NEDIS) dataset was used in this investigation. The National Temperature Data Service provided the weather information. Each weather element represented the data acquired over 4 days: the day of the patient's visit and the 3 days prior. For data analysis, we used scikit-RandomForestRegressor. There were 525,579 participants in the research.

Koc, M., & Acar, A. (2021).based on the Arduino Micro 33 BLE Sense as such an edge device, which could also run Tensor Flow Lite to accomplish identification on the device, offer a unique real-time wearable system of finger wind recognition in 3 (3D) space. This system employs deep learning to provide users the flexibility to create characters (10 numbers and 26 lowercase English letters) in open space. the motion data collected by inertial measuring system (IMUs) & processed by a microprocessor, both included in a Nano Nano 33 BLE Sense, using an algorithm to identify 36 characters. In order to train a convolutional neural network (CNN), we collected 63000 air-writing strokes data samples from 35 people, including 18 men and 17 women. We were successful in achieving a high recognition accuracy of 97.95%.

Liu, Y. (2022). allows us to think about several attack techniques while developing our manoeuvre plan. In order to produce the manoeuvre strategy for an air combat chase, a different frozen game framework that utilizes deep deep learning is suggested in this research. The one-on-one air warfare scenario and flight levels with fixed velocity are used in the creation of the manoeuvre plan agents for both sides' aircraft guiding. A reinforced learning environment of agent training is constructed using middleware that links the agents to air combat simulation software.

By using a reward shaping strategy, the training rate is accelerated and the produced trajectory's performance is enhanced. To deal with nonstationarity, agents are trained using alternate freeze games and a deep reinforcement algorithm.(Z. Wang et al., 2020)
Guo 2020 et al.While most prediction techniques place greater emphasis on model selection, there are explanations for changes in air pollution concentration. Recent deep learning frameworks are quite adaptable, therefore the model may need to be deep and complicated in order to match the dataset. Consequently, when the neural network based model's weights are numerous, over fitting issues might arise in just one deep neural network model. Additionally, stochastic gradient descent (SGD) produces a local optimum solution by treating all parameters equally during the learning process. The inherent association between PM2.5 and other auxiliary data, including meteorological data, season data, including time stamp data, which are employed to cluster for improving performance, is examined in this study using the Pearson correlation coefficient. In order to estimate the PM2.5 concentration for the upcoming hour, a deep ensemble network (EN) model combining a recurrent neural network (RNN), a long short-term memory (LSTM) network, or a gated recurrent unit (GRU) network is used.(C. Guo et al., 2020)
(Z. 2020 A. Wang et al., 2020). environmental and public health effects. The use of convolutional neural network (CNNs) to predict PM10 concentrations depending on atmospheric data is discussed in this article. In this case study, deep neural nets (both 1D & 2D) were investigated to see if they might be used for prediction problems. Additionally, in this contribution, the prediction model's accuracy is increased by the use of an ensemble technique known as Bagging (BEM).
Vennila, C.2022 et al. In this work, an ensembles of machine learning algorithms were employed to increase that accuracy of the recommended model. The simulation's findings demonstrate that, when compared to current approaches, the suggested method has a lower placement cost. The recommended ensemble model beat the traditional individual models when performance of such an ensemble classifier it include all of the combinations techniques was compared to that of the ensemble model. The results showed that a blended model that used machine learning & statistics performed better than a model that only used machine learning.(Vennila et al., 2022).

## IV. MACHINE LEARNING MODELS

Solar power forecasting is built on machine learning and artificial intelligence. Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), & Neural Networks (NNs) make up the majority of modern forecasting models (SVM). These models are typically data-driven and don't require the same level of power engineering understanding as meteorological models. Machine learning forecast algorithms can be used to anticipate PV power directly, eliminating the need to first estimate solar irradiance and then convert it to power generation. Another benefit of this category of techniques is their adjustable forecasting horizon. Most of the statistical & meteorological models mentioned in Section are suitable for short- or extremely short-term forecasting techniques, often intra-day forecasts.Machine models' forecasting horizon, in comparison, might be more flexible, ranging from inter-prognostications to next-day projections.

### *A. SVM*

The purpose of a support vector machine (SVM) is to act as a discriminator between two groups. To do this, we provide valuable data that has been feature-labeled and develop a classifier that is excellent at finding hidden data. A measuring classification approach is the most fundamental component of support vector machines. The major problem with classification is frequently overcome by binary classification of training instances that can be divided linearly.

### *B. ANN*

An artificial neural network is a computational model of biological neurons (ANN). Another term for neural networks is artificial neural networks. The idea behind ANN was primarily inspired by the biological science, particularly its neural branch, which is vital to and essential to the functioning of the human body.On human bodies, neural network training exercises are carried out. A neural network is a collection of connected input-output units have unique weights.

### C. *Decision Tree*

The majority of algorithms used in the classifications space are decision tree algorithms, by far. Another advantage is that it aids in categorization. The DT method offers a straightforward approach to modelling. People can quickly analyse a spanning tree to understand the decision-making process by using the straightforward tool known as a decision tree.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

### *D. KNN*

By focusing on the nearest neighbour whose value is already known, neural networks can find the missing information point. Find the closest place you can. There are two distinct parts to a neural network's mechanism. Despite the fact that structure & function approaches are used less frequently, neural networks are helpful for categorization. The K-NN approach is one of the "least" techniques according to the system. A neural network is used in the KNN technique to determine the value of k, which in turn establishes the number of neural networks that must be tested against with a sample set of data with an in-class description. There are two primary categories of methodologies for researching neural network models: those that use KNN and those that do not.

### *E. Naïve Bayes*

When there are persistent decision independence, naive Bayes, a kind of basic probabilistic classifier, helps with the application of the Bayes theorems. The naïve Bayes method was developed by George Bayes. It is more likely to occur in the future since it is more likely to occur after a continuous previous occurrence because Bayes' classifiers are based on dependent quantities.

### *F. Bayesian Network*

Another name of BN is belief networks. A BN is a probabilistic visual-spatial dispersion. There are two distinct parts to this BN. The initial component is built upon an Acyclic Graph, which refers to graph knot as stochastic processes and displays probabilistic compulsion worrying on the edges that are perceived between nodes or random variables.

## V. CONCLUSION

Weather forecasts have grown increasingly significant in recent years, since they may help us save time, money, property, or even our lives. Although India has a large number of weather stations, the most of them are located in densely populated regions such as cities, suburbs, or towns. This makes weather forecasts in remote regions less accurate, which can be troublesome for individuals who rely significantly on weather predictions for their daily activities, such as farmers. In this work, we assess temperature, apparent temperature, humidity, wind speed, wind direction, visibility, cloud cover, and forecast the weather using machine learning techniques such as Random Forest, Decision Tree, MLP classifier, Linear regression, and Gaussian naive Bayes. Based on the results, an accuracy comparison study is carried out.

## REFERENCES

[1] Vennila, C., Titus, A., Sudha, T. S., Sreenivasulu, U., Reddy, N. P. R., Jamal, K., Lakshmaiah, D., Jagadeesh, P., & Belay, A. (2022). Forecasting Solar Energy Production Using Machine Learning. International Journal of Photoenergy, 2022. https://doi.org/10.1155/2022/7797488.

[2] Li, M., & Wang, Y. (2020). An energy-efficient silicon photonic-assisted deep learning accelerator for big data. Wireless Communications and Mobile Computing, 2020. https://doi.org/10.1155/2020/6661022.

[3] Guo, C., Liu, G., & Chen, C. H. (2020). Air Pollution Concentration Forecast Method Based on the Deep Ensemble Neural Network. Wireless Communications and Mobile Computing, 2020. https://doi.org/10.1155/2020/8854649.

[4] Liu, Y. (2022). Short-Term Prediction Method of Solar Photovoltaic Power Generation Based on Machine Learning in Smart Grid. Mathematical Problems in Engineering, 2022. https://doi.org/10.1155/2022/8478790.

[5] Koc, M., & Acar, A. (2021). Investigation of urban climates and built environment relations by using machine learning. Urban Climate, 37(23), 100820. https://doi.org/10.1016/j.uclim.2021.100820.

[6] Gil, D., Johnsson, M., Mora, H., & Szymanski, J. (2019). Advances in architectures, big data, and machine learning techniques for complex internet of things systems. Complexity, 2019. https://doi.org/10.1155/2019/4184708.

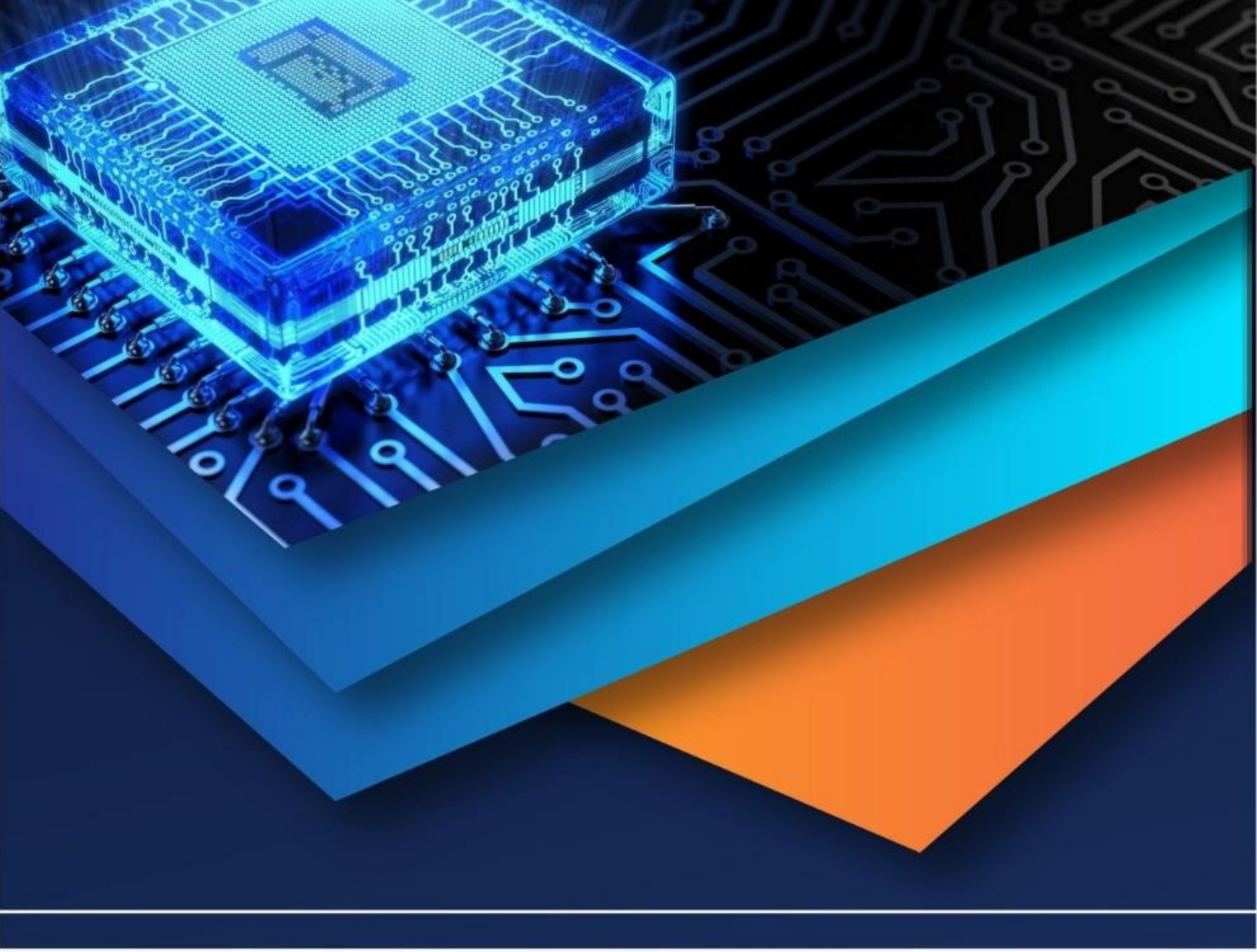
[7] Do, T. H., Tsiligianni, E., Qin, X., Hofman, J., La Manna, V. P., Philips, W., & Deligiannis, N. (2020). Graph-Deep-Learning-Based Inference of Fine-Grained Air Quality from Mobile IoT Sensors. IEEE Internet of Things Journal, 7(9), 8943–8955. https://doi.org/10.1109/JIOT.2020.2999446

[8] Drewil, G. I., & Al-Bahadili, R. J. (2022). Air pollution prediction using LSTM deep learning and metaheuristics algorithms. Measurement: Sensors, 24(October), 100546. https://doi.org/10.1016/j.measen.2022.100546

[9] Koc, M., & Acar, A. (2021). Investigation of urban climates and built environment relations by using machine learning. Urban Climate, 37(23), 100820. https://doi.org/10.1016/j.uclim.2021.100820

[10] Korunoski, M., Stojkoska, B. R., & Trivodaliev, K. (2019). Internet of Things Solution for Intelligent Air Pollution Prediction and Visualization. EUROCON 2019 - 18th International Conference on Smart Technologies, 1–6. https://doi.org/10.1109/EUROCON.2019.8861609

[11] Kothandaraman, D., Praveena, N., Varadarajkumar, K., Madhav Rao, B., Dhabliya, D., Satla, S., & Abera, W. (2022). Intelligent Forecasting of Air Quality and Pollution Prediction Using Machine Learning. Adsorption Science and Technology, 2022. https://doi.org/10.1155/2022/5086622



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)