A black background with white text

Description automatically generated with low confidence

MSc Data Science Project

7PAM2002-0901-2024

Department of Physics, Astronomy and Mathematics

**Data Science FINAL PROJECT REPORT**

**Project Title:**

Weather Forecasting Using Machine Learning Algorithms

**Student Name and SRN:**

Keerthi Bijivemula

22007462

Supervisor: Sophie Koudmani

Date Submitted: 06/01/2025

Word Count: 9546

GitHub address: https://github.com/bijivemulakeerthi/Weather-Forecasting-using-Machine-Learning-Algorithms.git

DECLARATION STATEMENT

This report is submitted in partial fulfilment of the requirement for the degree of Master of Science in Data Science at the University of Hertfordshire.

I have read the guidance to students on academic integrity, misconduct and plagiarism information at [Assessment Offences and Academic Misconduct](https://www.herts.ac.uk/__data/assets/pdf_file/0007/237625/AS14-Apx3-Academic-Misconduct-v17.0.pdf) and understand the University process of dealing with suspected cases of academic misconduct and the possible penalties, which could include failing the project module or course.

I certify that the work submitted is my own and that any material derived or quoted from published or unpublished work of other persons has been duly acknowledged. (Ref. UPR AS/C/6.1, section 7 and UPR AS/C/5, section 3.6).

I have not used chatGPT, or any other generative AI tool, to write the reportor code (other than where declared or referenced).

I did not use human participants or undertake a survey in my MSc Project.

I hereby give permission for the report to be made available on module websites provided the source is acknowledged.

Student SRN number: 22007462

Student Name printed: Keerthi Bijivemula

Student signature: Keerthi.B

UNIVERSITY OF HERTFORDSHIRE

SCHOOL OF PHYSICS, ENGINEERING AND COMPUTER SCIENCE

**Acknowledgements**

I want to express my sincere gratitude to my supervisor for their invaluable guidance and support throughout this research. I also appreciate the assistance from colleagues and peers for their helpful insights and constructive feedback during this study.

**Abstract**

This research aims at creating efficient models in weather prediction based on machine learning approaches. Section 1 gives an overview of the importance of weather prediction and presents the goals and nature of the research. Section 2 also discusses previous work on the topic of meteorological data analysis and different ML algorithms, with a particular focus on the accuracy of the predictions. Section 3 formally introduces the employed ML algorithms, these include Linear Regression, Random Forest, Gradient Boosting, XGBoost, and Support Vector Regressor. Describes how the model is trained, and tested, and how and when hyperparameters are tuned for best performance and predictive accuracy. Section 4 contains the results of the model evaluation and gives a demonstration of the effectiveness of each algorithm. Random Forest has the least error and near to perfect R² value like Gradient Boosting making it perfect on most observations. Optimizing hyperparameters contributes to enhanced model accuracy, especially the ensemble models. Section 5 of this analyses is the results with a focus on identifying how ensemble methods perform better to capture non-linear correlation. In comparing the findings of the respective study with earlier literature, it can be concluded that Random Forest and Gradient Boosting are resilience in weather prediction tasks. The research points out that when it comes to predicting weather conditions, machine learning models are accurate, especially the ensemble ones. The developments that are to be found in the following will consider creating scalable models, incorporating real-time data, and improving computational methods. The work has significance for enhancing weather prediction and other uses of the weather model.

**Table of Contents**

[1: Introduction 7](#_Toc186833876)

[1.0 Introduction 7](#_Toc186833877)

[1.1 Background and Context 7](#_Toc186833878)

[1.2 Research Problem 8](#_Toc186833879)

[1.3 Research Rationale 9](#_Toc186833880)

[1.4 Research Aim and Objectives 9](#_Toc186833881)

[1.5 Research Question 9](#_Toc186833882)

[1.6 Significance of the Study 10](#_Toc186833883)

[1.7 Structure of the Dissertation 11](#_Toc186833884)

[2: Background/ Literature Review 12](#_Toc186833885)

[2.1 Introduction 12](#_Toc186833886)

[2.2 Journal Collection Process 12](#_Toc186833887)

[2.3 Critical Analysis of Literature 12](#_Toc186833888)

[2.4 Theories and Models 17](#_Toc186833889)

[2.5 Conceptual Framework 18](#_Toc186833890)

[2.6 Literature Gap 18](#_Toc186833891)

[2.7 Summary 18](#_Toc186833892)

[3: Methodology 20](#_Toc186833893)

[3.1 Dataset Process Framework 20](#_Toc186833894)

[3.2 Dataset Selection 20](#_Toc186833895)

[3.2.1 Description of Dataset Features 20](#_Toc186833896)

[3.3 Selection of Language Code 21](#_Toc186833897)

[3.3.1 Application of python in machine learning models 21](#_Toc186833898)

[3.3.2 Advantages of Python in Machine learning model 21](#_Toc186833899)

[3.4 Selection of Technology 21](#_Toc186833900)

[3.4.1 Cloud Computing Platforms 21](#_Toc186833901)

[3.4.2 Machine Learning Frameworks 22](#_Toc186833902)

[3.5 Selection of Algorithm models for Research 22](#_Toc186833903)

[3.5.1 Linear Regression 22](#_Toc186833904)

[3.5.2 Random Forest 22](#_Toc186833905)

[3.5.3 Gradient Boosting 23](#_Toc186833906)

[3.5.4 Support Vector Regressor 23](#_Toc186833907)

[3.5.5 XGBoost Regressor 23](#_Toc186833908)

[3.6 Evaluation Method 24](#_Toc186833909)

[3.6.1 R² or R-squared 24](#_Toc186833910)

[3.6.2 RMSE or Root Mean Square Error 24](#_Toc186833911)

[3.7 Exploratory Data Analysis (EDA) 24](#_Toc186833912)

[4: Ethical Issues 28](#_Toc186833913)

[5: Results 29](#_Toc186833914)

[5.1 Individual Model Analysis 29](#_Toc186833915)

[5.2 Comparative Model Performance 36](#_Toc186833916)

[6: Analysis And Discussion 39](#_Toc186833917)

[6.1 Interpretation of Results 39](#_Toc186833918)

[6.2 Comparison with Literature 39](#_Toc186833919)

[6.3 Limitations of Results 40](#_Toc186833920)

[6.4 Alignment with Objectives 40](#_Toc186833921)

[6.5 Addressing the Research Question 41](#_Toc186833922)

[7: Conclusion 42](#_Toc186833923)

[7.1 Conclusions 42](#_Toc186833924)

[7.2 Applications and Real-World Situations 42](#_Toc186833925)

[7.3 Future Work 42](#_Toc186833926)

[7.4 Recommendations 43](#_Toc186833927)

[Reference List 44](#_Toc186833928)

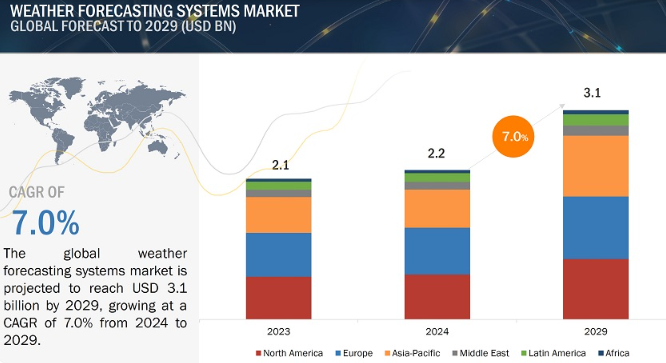
**1: Introduction**

1. **1.0 Introduction**

Weather prediction is crucial in many fields with operations and protection requirements worldwide (Bag *et al.,* 2022). In the recent past, the use of Machine learning (ML) algorithms has been implemented in the process of this forecasting with the foresight predictions being accurate. Previously, analysts utilized numerical simulations of the model to predict and this methodology rather posed some challenges in depicting pertinent aspects of the circulation in the atmosphere (Kochkov *et al.,* 2024). Business forecasting is made accurate by machine learning because large data sets are analyzed to identify distinguishing features for prediction. Other factors that are important for the generation of the weather include humidity, pressure and clouds (Lakra and Avishek, 2022). The objective of this research is to explore the usage of ML in increasing meteorological prediction and the parameters that must be incorporated into consideration as well as the Integration of the most efficient algorithms. The increase in data access for weather data helps in enhancing refreshing and oversimplified Weather-based ML-based forecasting models for utility purposes across various sectors.

1. **1.1 Background and Context**

Weather forecasting is a significant factor in most activities like agriculture, transport, Insurance and even general utilization of time (Fathi *et al.,* 2022). Several conventional approaches rely on numerical weather prediction (NWP) models based on mathematical equations to predict weather. However, these methods are subjected to accuracy problems occasioned by limited atmospheric information. Machine learning (ML) has become more prominent as a promising solution to overcome these challenges.



**Figure 1.1: Weather Forecasting System Market**

(Source: Markets and Markets, 2024)

The Weather Forecasting Systems Market Size is projected to grow from USD 2.2 billion in 2024 to USD 3.1 billion by 2029, at a CAGR of 7.0% (Markets and Markets, 2024). These systems utilize historical data, and the accuracy of the prediction is significantly higher compared to other methods. For example, Shouman (2024) showed that ML models excelled over NWP in predicting photovoltaic power, a variable that is strongly associated with the weather. Support vector machines, decision trees, and neural networks are the typical cases of ML that compare intricate data patterns and include humidity, cloud formation, and temperatures. This research attempts to discover the existence of important variables as well as to determine the optimal set of algorithms for weather predictions using ML. Precision improves results for fields such as aviation since achieving better than other major system accuracy can reduce the need to spend millions on improvements every year (Yazdani-Asrami *et al.,* 2022). Therefore, incorporating ML in weather prediction is indispensable in modern society and economic standards.

1. **1.2 Research Problem**

Weather prediction difficulties persist because the conditions relating to the atmosphere are not easy to predict and the variables are often interrelated (Jaseena and Kovoor, 2022). Such complexities are not easily reflected in traditional numerical prediction models thus lowering its reliability. Machine learning could have some solutions but it is a challenging task to find out the right parameters and algorithms for prediction (Zhang *et al.,* 2021). Thus, the prioritization of activity variables, such as humidity, pressure, and temperature, is necessary for enhanced accuracy of the forecast. However, the amount of models will surpass any practicable toolkit and this diversity makes the selection of appropriate algorithms challenging (Badjie, Cecílio, and Casimiro, 2024). Overcoming these challenges is imperative in building efficient, accurate ML-based weather forecasting tools to support a wide range of uses in agriculture, transportation, and disaster response.

1. **1.3 Research Rationale**

The role of weather prediction has central importance for such industries as agriculture, aviation services, and during the crisis period. The climate-related disasters that occurred in the world hit over 4 billion people from the year 2000 to 2019 as per the report of the UN (Donatti *et al.,* 2024). It is characteristic of conventional numerical approaches that they can poorly predict sharp changes in the atmosphere. Machine learning (ML) enhances the forecasting capability by reaching a large database and drawing out pattern possibilities effectively. The ML market also looks great globally, and its size may increase to 209 billion USD by 2029. Zhang *et al.* (2021), identify increased prediction accuracy by exploiting the use of ML models. This research defines fundamental weather parameters such as humidity and temperature and analyzes essential ML techniques. As datasets from sources such as Kaggle become more commonplace, there is great potential in using ML for forecasting. From the potential value perspective, the enhanced accuracy contributes to disaster prevention and obtains fewer losses in terms of money and lives. This work helps to fill gaps in the selection of variables and algorithms for accurate weather forecasting systems.

1. **1.4 Research Aim and Objectives**

The research aims to identify the weather parameters and variables crucial for weather forecasting using machine learning algorithms.

***Objectives***

1. To collect a suitable weather dataset for weather forecasting.
2. To choose the important weather parameters like cloud formation, humidity, pressure etc., for developing the weather prediction model.
3. To choose the best ML algorithm for developing the model.
4. **1.5 Research Question**

The research question which this research procedure will address is as follows:

**What variables, parameters etc., are crucial for weather forecasting using machine learning algorithms?**

1. **1.6 Significance of the Study**

The importance of this study stems from the fact that machine learning will be used to improve the prediction accuracy of the weather. Weather forecasts have a critical role in agriculture, aviation, the disaster management and transportation industries (Merz et al. 2020). Traditionally, complex atmospheric interaction prediction can be poor with many of the traditional numerical prediction methods. Machine learning models can consume and analyze vast datasets to give very accurate predictions. Key weather parameters are identified which aid in the optimization of forecasting models for practical applications. Predictions are enhanced to reduce costs and to prevent losses in weather-sensitive industries such as farming and logistics. Weather forecasting is improved for disaster management purposes, which enables predicting severe weather events far more accurately (Huang, Wang and Liu, 2021). The challenges of selecting the right variables and algorithms for forecasting are addressed in this study. Historical and real-time data can be used for very precise predictions with machine learning. Humidity, pressure and temperature analysis allows for the understanding of atmospheric conditions. The results can assist in choosing an algorithm of choice for a variety of industrial applications in the field of weather forecasting. Integrating ML in weather forecasting enables industries to mitigate climate-related risks, efficiently (Chen et al., 2023). The effective ML-based forecasting approaches identified by this study add to the research. Practical benefits in terms of cost savings and safety improvements are provided by enhanced prediction models. The research outcomes are useful for improving weather prediction technology and application. Accuracy improves industries' operations optimisation and resilience to adverse weather impacts.

1. **1.7 Structure of the Dissertation**

**Figure 1.2: Dissertation Structure**

(Source: Self Developed)

**2: Background/ Literature Review**

1. **2.1 Introduction**

Weather forecasting has been an important application of machine learning algorithms. The models are developed based on historical data of various weather parameters. The parameters could be temperature, humidity, wind speed, and other atmospheric parameters. This way the possible long-term weather prediction can be done using machine learning algorithms. This literature review section will critically analyse the existing papers on weather prediction and identify the possible literature gap. The relevant theories and models will also be discussed.

1. **2.2 Journal Collection Process**

The journal collection inclusion and exclusion criteria are as follows:

1. It was ensured that the research papers have been published in recent years. This is the reason why the research papers published in 2019 have been chosen for the background analysis.
2. The research papers about weather forecasting were only selected to ensure that the background analysis would be within the scope of the research.
3. The research papers written in the English language were selected as part of this research procedure.
4. **2.3 Critical Analysis of Literature**

As per Singh, Chaturvedi and Akhter, (2019), the constant climate change has ensured that the old weather prediction model gets obsolete. The main aim of the paper was to develop a model for weather prediction that can be used in remote areas. The authors have developed the model using “Random Forest Classification” algorithms. The result was a portable and low-cost weather prediction solution for remote areas.

A diagram of a machine learning algorithm

Description automatically generated

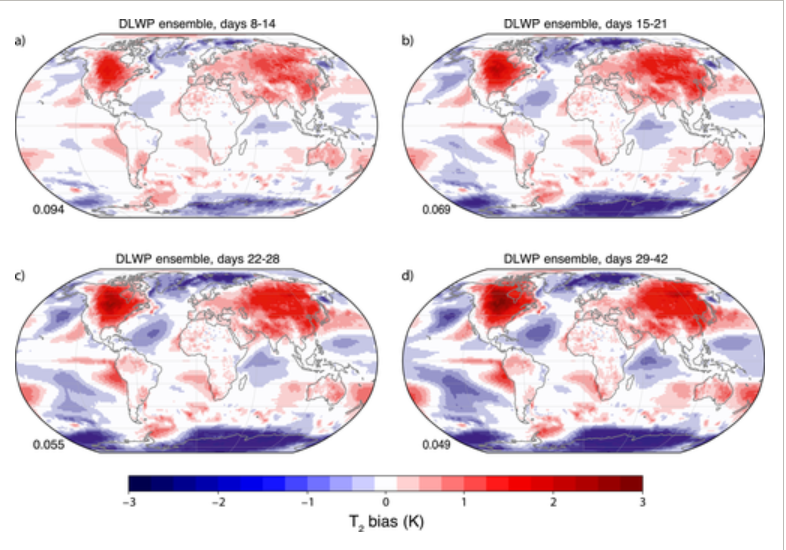
**Figure 2.1: Proposed Weather Forecasting Model**

(Source: Singh, Chaturvedi and Akhter, 2019)

The figure shows the model's overall flowchart or system architecture. The model would first classify and then make the prediction. The developed model has an accuracy of more than 87% in terms of its prediction of whether there would be rain or not. Despite the high accuracy, the limitation of the model could be that it's only evaluated based on whether it would rain or not. The evaluation based on other weather incidents had not been done. On the other hand, Jakaria, Hossain and Rahman, (2020), the weather parameters and data used in weather forecasting are unstable. The authors have developed a weather forecasting model that can predict short weather based on data from multiple weather stations. The author has collected data from various weather stations in Nashville. The authors have compared results from “Random Forest Classifier” with SVM, “Multi-layer Perceptron”, “Extra-Tree Regression”, and “Ridge Regression”. The random first classifier has shown the best result. The main strength of the paper was its usage of multiple ML algorithms for weather prediction, however, the authors have not developed a full-fledged application which is a major limitation of the model.

Bochenek and Ustrnul, (2022), conducted a systematic literature review on various weather prediction papers. The aim was to identify the most common factors and methods used for weather prediction. The result showed that radiation, pressure, temperature, precipitation, and wind are the most common meteorological fields examined for predicting weather. SVM, random forest, XGBoost, “Artificial Neural Networks” etc., have been the most popular algorithms that were used. The paper has summarised and identified the important aspects of weather parameters and algorithms used for weather prediction. However, it has not developed its own model which is a major limitation.

As per Wang *et al.* (2019), the inappropriate settings of the initial states can lead to unsatisfactory results in weather prediction. This is why the authors have proposed a data-driven weather prediction model based on machinimas of information fusion. The authors have used deep learning algorithms based on the “novel negative log-likelihood error” problem. It is able to forecast for both uncertainty quantification and single-value prediction. The accuracy of the model significantly increased compared to traditional numerical models for weather prediction. However, accuracy is still low in this regard. On the other hand, Weyn *et al.* (2021), have developed an “ensemble prediction system” based on DL algorithms. The CNN algorithm was used for six-week forecasts.



**Figure 2.2: Weather Forecasting of Cyclone**

(Source: Wang *et al.,* 2019)

This figure shows the cloud and weather prediction for Cyclone Irma using the proposed model. The proposed model can forecast 320 times within 3 minutes and this shows the effectiveness of the model. The main effectiveness of the mode is that it has been evaluated based on real-world incidents like Irma. However, it is not effective for long-term weather prediction for the future.

Cho *et al.* (2020), have developed a numerical model for weather prediction for extreme air temperature events in urban areas. The prediction of minimum and maximum temperature is the most basic and essential part of weather forecasting. As per Hewage *et al.* (2021), the model of "numerical weather prediction" needs significant power for solving mathematical equations. The authors have proposed a novel "lightweight data-driven weather forecasting model" for weather forecasting. As per Chattopadhyay, Nabizadeh and Hassanzadeh, (2020), the ever-growing resources and time consumption have been the main problems associated with weather prediction based on numerical values. The authors have proposed an analogue forecasting model based on data. The proposed model has used CapsNet and "capsule neural networks" to build a "novel deep learning pattern-recognition technique". The data was collected from the Earth system model and it was used to train the weather model. The accuracy of the trained model is 45% which is not satisfactory. However, the usage of CapNets has increased the model accuracy up to 80%. CapsNet has outperformed both CNN and normal logistic regression models in this regard. However, the model accuracy is still not close to 100% which is a major model limitation.

As per Chantry *et al.* (2021), ML algorithms can be valuable as an accelerator for the scheme of parameterisation. The authors have developed a parameterisation scheme using ML algorithms. The authors have used deep learning algorithms like ANN in this regard. The result showed that for the medium-range prediction, the model is highly accurate. However, for short-time and long-time prediction, the model is not that accurate for weather prediction.

Lam *et al.* (2023), that global weather forecasting is tough for medium-time forecasting due to various social and economic domains. The traditional methods need significant computing resources for prediction. The authors have developed an ML-based model called Graphcast. It can predict more than 100 weather variables at the same time. The accuracy of the model is 90% which is significantly higher than other systems of operational deterministic. It can predict severe events accurately including extreme temperatures, cyclones, atmospheric rivers and so on. This way it has been effective to develop complex models for dynamic systems.

A collage of graphs and charts

Description automatically generated

**Figure 2.3: Weather Prediction using GraphCast**

(Source: Lam *et al.,* 2023)

The figure shows possible prediction graphs of Graphcats based on cyclone tracking, atmospheric rivers, extreme heat and precision levels.

As per Grönquist *et al.* (2021), the quantification of forecasting of weather is very difficult due to the extreme weather events. Ensemble prediction systems. Can be used to predict extreme weather events. However, these are costly with high computational resources. The authors have proposed a mix of traditional models and ensemble prediction models to assess the non-linear relationship between different weather parameters. The authors have used global data for their prediction methods. The authors have used case studies to prove improved weather forecasting for extreme events. This way the overall cost and resources of the ensemble system can be reduced through the mixed method. Han *et al.* (2022), have developed a wind speed prediction model as part of the quantitative weather forecasting model. The prediction of wind speed is tough due to the relationship between different meteorological parameters. The authors have used weather forecasting and research to develop a hybrid model for wind speed prediction. In creating other prediction models, the authors have combined the multivariate data decomposition method and the DL model. The deep learning algorithms were developed using CNN and bidirectional LSTM. It was found that the model proposed by the authors has outperformed other similar models including an MAE value of 0.1042. The accuracy has also been increased significantly in this regard.

Rasp *et al.* (2020), have developed a data-based forecasting model for weather prediction. The authors have tried to predict global weather based on their data-driven approach. The authors have collected datasets from the ERA5 archive. They have used simple linear regression, physical models, and deep learning algorithms for model development. The method has been able to predict weather for 3-5 days. The developed model has significantly increased the overall quality prediction quality. The lack of a detailed description of evaluation metrics could be the main limitation of this study.

1. **2.4 Theories and Models**

***Technology Acceptance Model (TAM)***

TAM can be defined as the model that describes how people accept a new technology. It was established by Fred Davies to show why people can accept or reject new technology. The use of modern machine learning, deep learning, and time series algorithms has improved the weather prediction quality significantly. People can see weather forecasting from their own mobile or computers. This is why weather forecasting using modern technologies can be explained using TAM.

1. ***Perceived usefulness:*** A user will use a new technology if they believe the new technology will improve job performance (Zaineldeen *et al.,* 2020). Modern techniques like Time series models, ML and DL algorithms have improved the weather forecasting performance significantly well. The audiences have easily accessed the weather prediction application from their preferred devices.
2. ***Ease of use:*** This means the user will accept a new technology if it frees their effort. The modern technologies of weather forecasting have reduced the effort of users to accept weather forecasting. The models have improved the quality of the forecasting and reduced the possible effort of researchers to collect data as well.

This shows the Ml algorithms have significantly been effective for forecasting the future weather and collecting weather data in this regard.

1. **2.5 Conceptual Framework**

A diagram of a graph

Description automatically generated

**Figure 2.4: Conceptual Framework**

(Source: Created by the Authors)

This is the conceptual framework that this research procedure would like to follow. Here, the “dependent variable” is “Effective weather forecasting using machine learning”. Now, the “independent variables” are choosing proper weather parameters, choosing the proper dataset selecting of right ML algorithm.

1. **2.6 Literature Gap**

The various research papers have discussed the usage of various algorithms and various weather parameters for weather forecasting. However, there has been a lack of specificity about which weather parameters researchers need to consider to focus their models on. There are specific weather and atmospheric parameters like cloud formation, wind speed, temperature, pressure, humidity etc. This research procedure has the purpose of explaining in depth the specific weather parameters that will be used for predicting the weather in this regard. This way the specific algorithms will be selected for developing the model. This way the classification and prediction procedure will be easier.

1. **2.7 Summary**

To summarise, the background section has provided the required information to assess the existing papers about weather prediction. Random Forest has been the most popular algorithm that has been used for forecasting weather followed by SVM and other deep learning algorithms. The data-driven approach for numerical prediction of weather has been a trend in recent years. The technological acceptance model has shown why both researchers and users are accepting modern technology for weather forecasting. Based on the analysis it was found that the lack of explanation of the weather parameters has been the primary limitation of the existing study. This will be mitigated in this study.

**3: Methodology**

1. **3.1 Dataset Process Framework**

A diagram of data processing

Description automatically generated

**Figure 3.1: Dataset Process Framework**

(Source: Self-created using Draw.io)

1. **3.2 Dataset Selection**

This research procedure will use a combination of two datasets. The first dataset was collected from the Kaggle website as it is a popular data collection website. The dataset link is as follows: “<https://www.kaggle.com/datasets/thedevastator/weather-prediction/data?select=weather_prediction_dataset.csv>”. The other dataset that was collected was from Cambridge University’s website. The link is “<https://www.cl.cam.ac.uk/weather/>”.

A table of numbers and lines

Description automatically generated with medium confidence

**Figure 3.2: Weather Dataset**

(Source: Obtained Using Jupyter Notebook)

**3.2.1 Description of Dataset Features**

This dataset consists of descriptions of meteorological variables along with both raw and Units converted forms. The other independent variables are cloud cover, abbreviated as CC in oktas, wind direction as DD in degrees, wind speed as FG and gust as FX in m/s and humidity abbreviated as HU in fraction of 100 (Kaggle, 2024). The pressure (PP) is given in 1000 hPa, global radiation (QQ) in 100 Wm², and precipitation amount (RR) in 10 mm, Sunshine duration (SS) is in hours, the mean temperature (TG) together with the minimum and maximum temperatures (TN, TX) is in °C. The present features allow for a detailed evaluation of the weather status within the atmosphere, temperature, and radiation.

1. **3.3 Selection of Language Code**

**3.3.1 Application of python in machine learning models**

Python has been found to be very useful in machine learning due to ease in writing scripts and programs. Some well-known open source libraries consist of NumPy, Pandas and Matplotlib which is used for analysing the data. Existing progressive libraries such as Scikit-learn, TensorFlow, and PyTorch are used for designing, training and testing of ML models. Being used in weather forecasting, Python comprises regression, classification and ensemble learning models which incorporate Linear Regression, Random Forest, and XGBoost among others. Furthermore, Python becomes easier to incorporate with the APIs and data sources through which researchers can compile and process the weather data.

**3.3.2 Advantages of Python in Machine learning model**

Python is also arguably the best language for beginners because it is simple to learn and easy to read and yet it is also capable of supporting complex ML operations as well. It also has a rich library which can help lessen the development time and also includes algorithms and tools. This makes it easier to find help from others who have used it before, and to continually improve techniques as it has a large user base. They are beneficial in weather forecasting since Python shows high efficiency in handling big datasets and calculates the results correctly. Furthermore, it is designed to run on many platforms and frameworks that further increases its usefulness in machine learning applications, therefore, becoming the most popular language used in the creation of precise and efficient models of forecasting.

1. **3.4 Selection of Technology**

**3.4.1 Cloud Computing Platforms**

The reason these clouds AWS, GCP, Microsoft Azure are central to this research is that the tools offer scalability, computation, and storage of data. These platforms enable the author to process and analyze the massive weather datasets and such computationally heavy machine learning algorithms. Google Colab and AWS SageMaker are known tools that allow one to easily deploy and train models to help with the work. Moreover, that cloud premises should be available and shareable and therefore, is ideal for multi-authoring research work and real-time simulations involving weather predictions.

**3.4.2 Machine Learning Frameworks**

Scikit-learn, TensorFlow and XGBoost are selected since they assist in creating a variety of Machine learning models. Table 3 also shows that scikit-learn provides a complete set of algorithms for both regression, classification, and clustering. TensorFlow is utilized for deep learning, especially for data intensive and neural network applications. However, for ensemble learning, the XGBoost model is used because it is faster and accurately performs nonlinear transformation on independent variables. Altogether, these frameworks contribute to the construction of an accurate and stable technological platform for training and rating weather predictions.

1. **3.5 Selection of Algorithm models for Research**

**3.5.1 Linear Regression**

Linear Regression is selected as the best fitting model due to its application in interpretation of the correlation between features and the target variable. Subsequent to building the model using the training data (X\_train, y\_train), numerical accuracy is assessed by the MAE, MSE, RMSE and the coefficient of determination (R²). These metrics are used to evaluate the proposed model regarding its accuracy and capability of prediction. The evaluation facilitates comparison with other elaborate models to identify compatibility requirements suitable for the dataset.

**3.5.2 Random Forest**

The Random Forest Regressor is chosen because the model can well interpret interactions between features and their importance. It is trained using the training data (X\_train, y\_train) and evaluated using various metrics: MAE, MSE, RMSE, and R². These are by and large measures of the accuracy of the models, the errors involved in the prediction, and the extent of their predictive abilities. The importance of features is also tested to determine the most significant refractories of the dependent variable. This helps explain how a model arrived at such a decision and what features played a major role in the decision-making.

**3.5.3 Gradient Boosting**

Gradient Boosting model is selected because it can deal with non-linear variables and enhanced ensemble learning. Their results indicate that it integrates many weak models to develop a single, more precise predictor. In the last segment of model assessment, the MAE, MSE, RMSE and R² measures are applied to measure the performance of the model. These metrics give a holistic view on accuracy and on how good the model is in predicting unseen data making this model highly recommendable for regression tasks.

**3.5.4 Support Vector Regressor**

SVR is selected as it is capable of dealing with nonlinearity and high-dimensionality of the data. Prediction accuracy is enhanced in SVR due to the fact that the algorithm seeks to achieve the largest margin of error for the hyperplane that is to be formed. Methods of accuracy, errors and predictive performance include: MAE, MSE, RMSE and the coefficient of determination (R²). Therefore, the overfitting resistance and generalization capability make it potentially useful with such complicated characteristics, whereas it can determine detailed prognosis even though the forwarded parameters show nonlinear associations with weather conditions.

**3.5.5 XGBoost Regressor**

XGBoost Regresser is chosen due to its ability to work with big data with intricate features. It incorporates gradient boosting with additional such as regularization to avoid the problem of over fitting. It combines several weak learners’ work to improve the predictive models and minimize the level of error in conclusions. Other predictors include the degree of model accuracy—MAE, MSE, RMSE, R²—are used to establish the model’s error range. Based on the characteristics that illustrate flexibility in handling of non-linear models and missing data, XGBoost Regresser is highly appropriate for use in weather prediction for accuracy and reliability of results.

1. **3.6 Evaluation Method**

The metrics are used to determine whether the machine learning models are accurate for weather forecasting or not. Two model fit measures employed in this study are R² and RMSE in which R² stand for R-squared while RMSE means Root Mean Square Error.

**3.6.1 R² or R-squared**

R² is the statistical coefficient which determines the extent to which the dependent variable can be explained by the independent variables in a model. It shows the goodness-of-fit and is always between 0 and 1. The value of R² is higher in case of better interpretation of the variability, accordingly low R² value indicates poor prediction capability of the model. In meteorology to show how well the model gives the right picture of the weather, R² is used to determine this. It could be useful in some of those methods if used together with other measures, which define the error magnitude more accurately.

**3.6.2 RMSE or Root Mean Square Error**

RMSE gives an estimation of the mean squared error which gives valuable information about variance of prediction errors. It determines the typical size of the discrepancies between the observed and the forecasted values. This is because RMSE is valuable in exposing the level of approximation of the forecasts with observed values especially when accuracy is key in the weather information archive. A smaller RMSE value means better predictive results needed in marketing for product development and forecasting. Together with R², the relative measure of the mean squared error, RMSE, addresses both the goodness-of-fit and the magnitude of the error making these measures the cornerstone of model assessment.

1. **3.7 Exploratory Data Analysis (EDA)**

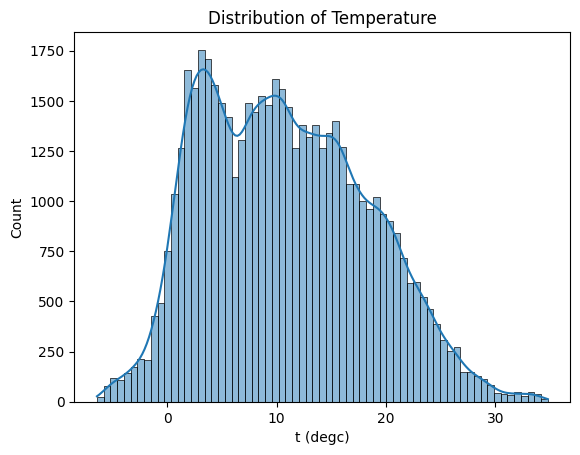
A screenshot of a computer

Description automatically generated

**Figure 3.3: Descriptive Statistics**

(Source: Obtained Using Jupyter Notebook)

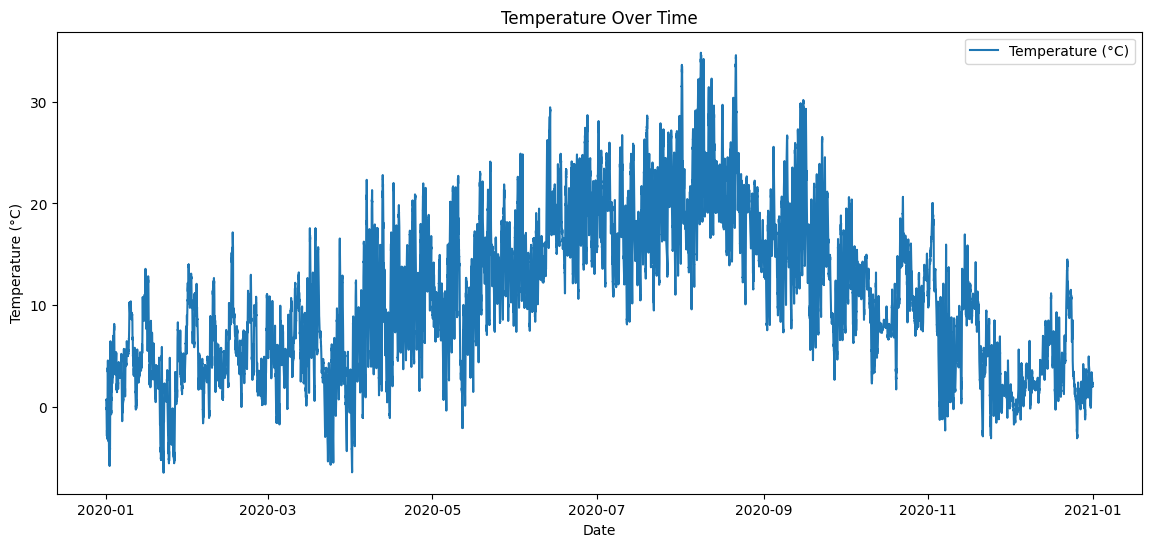
This data-set has 52,696 records with 21 features, some of them are temperature, humidity, pressure, and wind speed. Key statistics show diverse ranges: temperature ranges between -6.44°C and 34.8°C humidity ranges between 21.16% and 100% and wind speed ranges from 0 to 13.77m/s. Some of the columns still contain missing values and there are outliers present as well.



**Figure 3.4: Distribution of Temperature**

(Source: Obtained Using Jupyter Notebook)

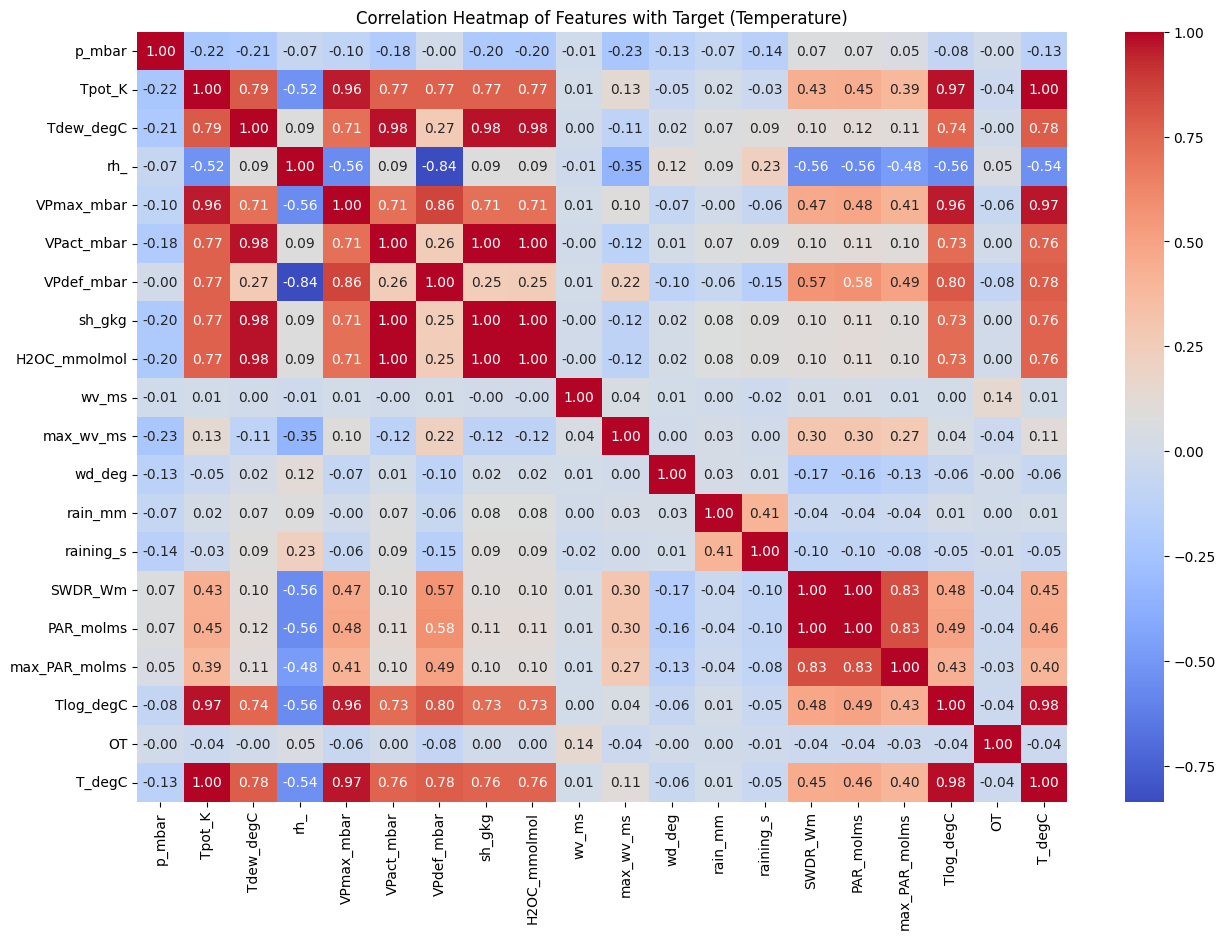
The histogram above demonstrates the frequencies of temperature by taking a snapshot of what the majority of the temperatures range between 10°C to 20°C.



**Figure 3.5: Temperature Over Time**

(Source: Obtained Using Jupiter Notebook)

The plot shows the movement in the temperature during a particular period. The plot oscillates showing that sometimes there are higher temperatures in some months than in other months, where the temperatures are comparatively low.



**Figure 3.6: Correlation Heatmap of Features with Target (Temperature)**

(Source: Obtained Using Jupyter Notebook)

The above heat map depicts the associations between various weather features and the target variable, temperature. Some of the Parameters which are positively related to temperature include dew point (Tdew\_degC) pressure (p\_mbar) and relative humidity (rh).

A white background with black text

Description automatically generated

**Figure 3.7: Correlation**

(Source: Obtained Using Jupiter Notebook)

The correlation values depicted demonstrate that different features like Tpot\_K Tlog\_degC and VPmax\_mbar are strongly positively lined with temperature which makes them reliable determinants of temperature.

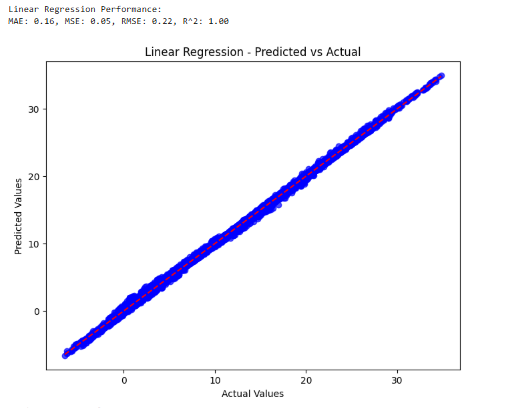
**4: Ethical Issues**

The dataset does not have any personal data of people and it is publically available to use. This is the reason why it does not have any privacy issues for data breaches. There is no chance of breaching the GDPR principle because it does not have any data privacy issues.

The dataset has indeed used several journal papers for background study. The target will be to appropriately reference all those secondary sources of data so that the original data owner gets the credit (Hummel, Braun and Dabrock, 2021). One of the major considerations of the research procedure will be to honestly present all the data. The findings from the software will be honestly presented with screenshots of the software in this regard.

**5: Results**

1. **5.1 Individual Model Analysis**



**Figure 5.1: Linear Regression Performance**

(Source: Obtained Using Jupyter Notebook)

The Linear Regression model shows excellent performance, with an R² value of 1.00 indicating a perfect explanation of variance in the target variable. Low error metrics (MAE: 0.16, MSE: 0.05, RMSE: 0.22) highlight its accuracy in predictions. These results suggest the model is well-suited for this dataset, effectively capturing linear relationships between features and the target variable.

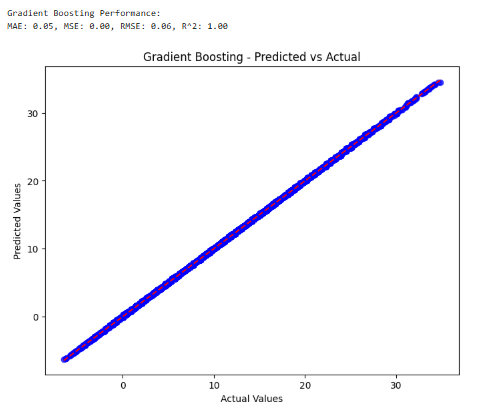
A graph with a line drawn on it

Description automatically generated

**Figure 5.2: Random Forest Performance**

(Source: Obtained Using Jupyter Notebook)

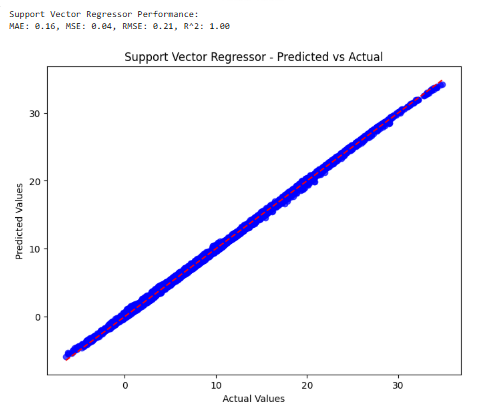
The Random Forest model demonstrates exceptional performance, with near-zero error metrics (MAE: 0.00, MSE: 0.00, RMSE: 0.01) and an R² value of 1.00, indicating perfect prediction accuracy. This suggests the model captures complex feature interactions effectively, making it highly reliable for this dataset. Its ability to minimize errors highlights its robustness and suitability for weather forecasting tasks.



**Figure 5.3: Gradient Boosting Performance**

(Source: Obtained Using Jupyter Notebook)

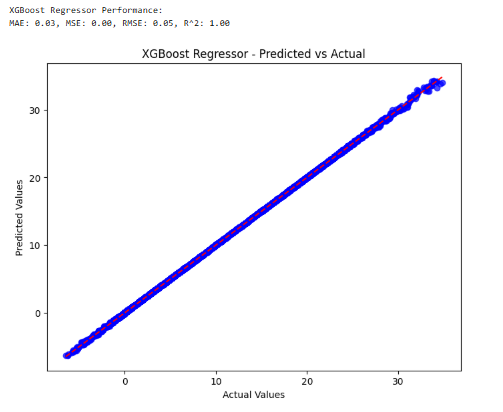
The Gradient Boosting model delivers excellent predictive performance, with low error metrics (MAE: 0.05, MSE: 0.00, RMSE: 0.06) and a high value of the coefficient of determination in the model as 1.00, which means accurate prediction coupled with the ability to handle a non-linear relationship. Due to ensemble learning, weak predictors are included in strong models and this algorithm is very useful for accurate and reliable weather forecasting.



**Figure 5.4: Support Vector Regressor Performance**

(Source: Obtained Using Jupyter Notebook)

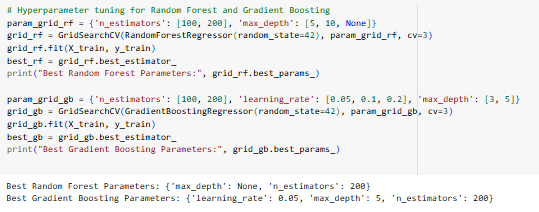
The Support Vector Regressor (SVR) gives a very good result concerning MAE 0.16, MSE 0.04 and RMSE 0.21 thus revealing highly accurate prediction. The value of R² is equal to 1.00 which verifies a perfect fit of the model and indicates that the regression equations fit the data well. Weather forecasting also proves that SVR can manage the data where non-linearity is considerable while several input features are high.



**Figure 5.5: XGBoost Regressor Performance**

(Source: Obtained Using Jupyter Notebook)

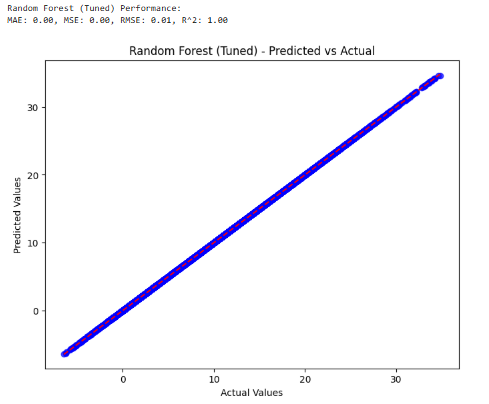
The predictive model used here is the XGBoost Regressor it also performs perceivably well in terms of MAE at 0.03, MSE at 0.00 and RMSE at 0.05. The coefficient of determination R² is 1.00 hence it is clear that the data pattern has been captured well by the model. Since XGBoost deals well with non-linear data and big data, it is most suitable to be used in weather predictions.



**Figure 5.6: Hyperparameter tuning for Random Forest and Gradient Boosting**

(Source: Obtained Using Jupyter Notebook)

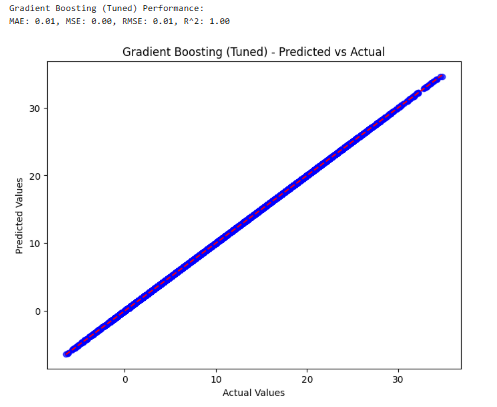
These results depict an optimum hyperparameter-optimized tuning for the Random Forest and Gradient Boosting model. For Random Forest, the best configuration is with no maximum depth (max\_depth): None and 200 estimators, which probably lets the model fit more detail into the data while not restricting tree depth. The most optimal tunable parameters for Gradient Boosting are: a quantile learning rate of 0.05, and a max depth of 5 and 200 estimators as it maintains the complexity of the model and the speed of learning simultaneously. All these hyperparameters improve the model performance for enhancing the accuracy and generalization capability for weather forecasting.



**Figure 5.7: Random Forest (Tuned) Performance**

(Source: Obtained Using Jupyter Notebook)

The Random Forest with the tuning of hyperparameters shows the benefit of the model as it confirms a 0% error. Since MAE and MSE are measures of error between the predicted and the actual values and as shown above, both are 0. The result for RMSE of 0.01 shows little difference between the observed and the forecasted values, which also enhances the accuracy of the model. The R² value of the model is 1.00 meaning a 100% variance in the data is explained indicating perfect fit and prediction. From these results, it can therefore be concluded that the tuned Random Forest gives very accurate weather prediction with very little error, and would therefore be suitable for use in an accurate and reliable weather prediction.

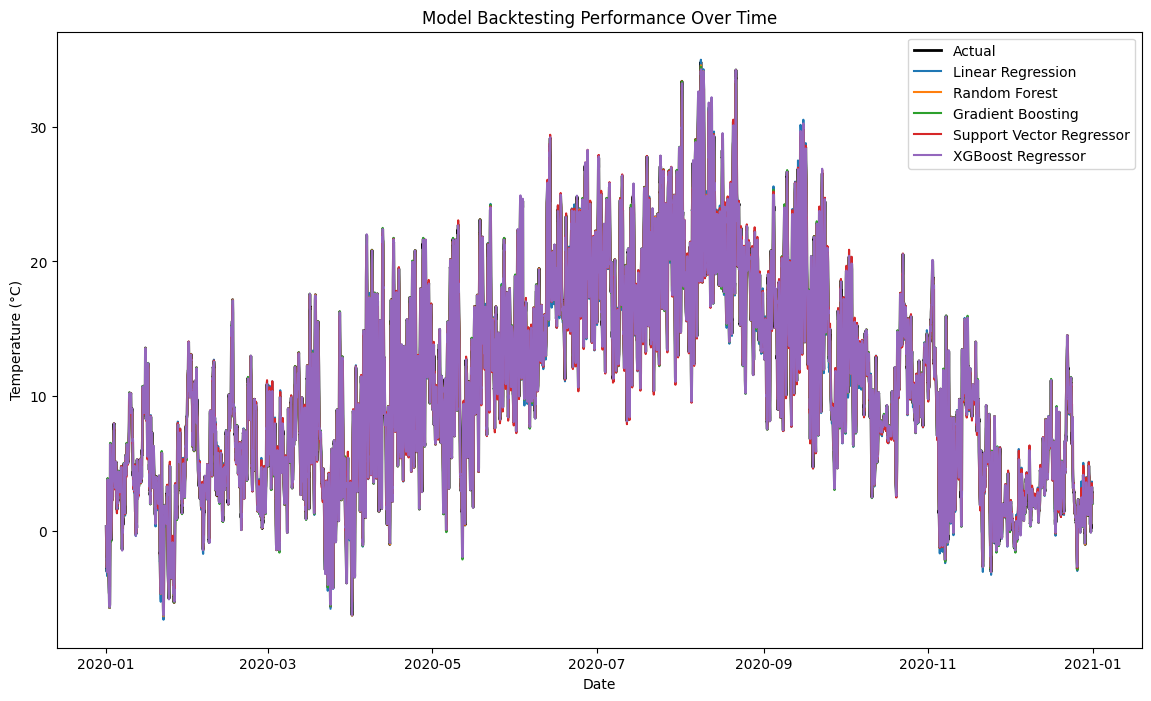


**Figure 5.8: Gradient Boosting (Tuned) Performance**

(Source: Obtained Using Jupyter Notebook)

The tuned Gradient Boosting model depicts excellent performance and almost zero loss. The MAE and MSE values are very low to means that the amount of errors that have been made in the prediction is very small. The closeness of predicted and actual values is depicted by its RMSE of 0.01. The R² value of 1.00r means that the model accurately predicts all the variations, and provides accurate and precise weather forecasts.

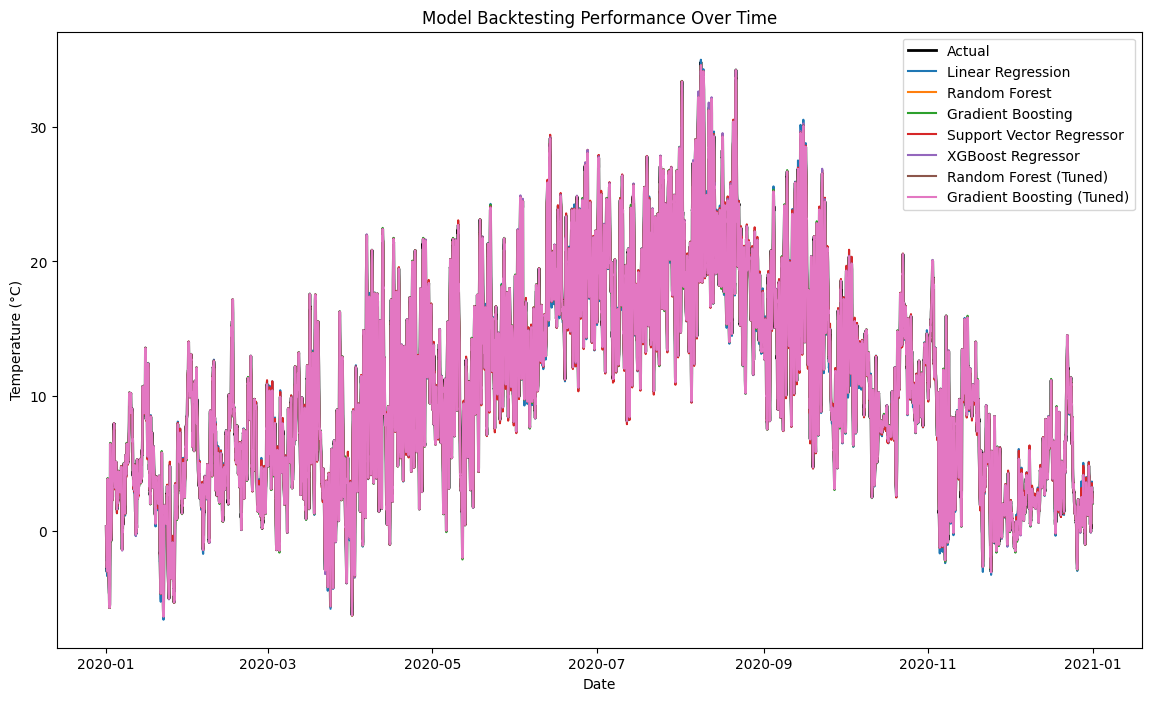
1. **5.2 Comparative Model Performance**



**Figure 5.9: Model Backtesting Performance Over Time**

(Source: Obtained Using Jupyter Notebook)

Fig.5.9 shows the level of performance of the selected backtesting ML models at a particular period. These are the actual temperature values and the temperature values obtained from the models such as Linear Regression, Random Forest, Gradient Boosting, SVR and XGBoost. Very high levels of accuracy are achieved across models, as evidenced by the fact that the predicted lines are accordant to actual values. Random Forest and Gradient Boosting show a more accurate overlap with the actual data which indicates Random Forest and Gradient Boosting learn temperature patterns better. However, we also know that minor deviations are visible for easier models like Linear Regression. This figure shows how crucial it is to have sophisticated calculation procedures to achieve precise and reliable prognoses in Meteorology.



**Figure 5.10: Fine-tuned Model Backtesting Performance Over Time**

(Source: Obtained Using Jupyter Notebook)

It is noteworthy that the backtesting performance of fine-tuned machine learning models has been presented in Figure 5.10. When compared to untuned models, these modified algorithms, most of which are embedded into the tuned Random Forest and Gradient Boosting models, follow the actual temperature values to the tee and look slightly like actual temperature readings, which suggests higher accuracy. Tuning was performed to finalize the increase in the number of hyperparameters to balance the capability of the models to detect patterns.

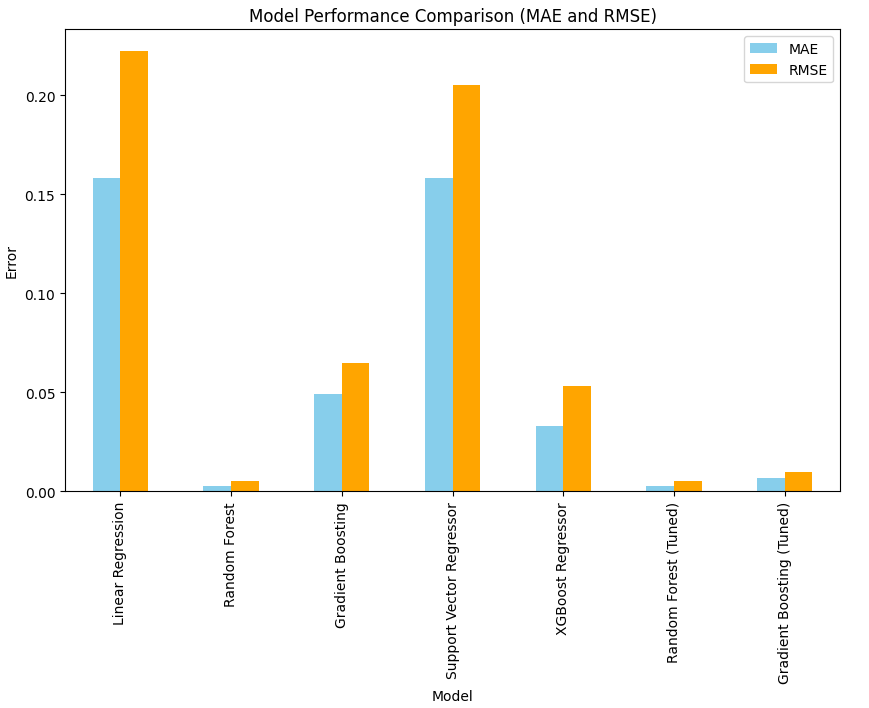
A screenshot of a computer

Description automatically generated

**Figure 5.11: Model Performance Comparison**

(Source: Obtained Using Jupyter Notebook)

The Random Forest model demonstrates the lowest errors (MAE: 0.0027, MSE: 0.00003, RMSE: 0.0055) with perfect predictions as envisaged by the excellent R² of 0.999999. Gradient Boosting algorithm has an RMSE of 0.0647 the accuracy of XG Boost model is high with an MAE of 0.033, RMSE of 0.053 and R² of 0.999949. Among finer-tuned models, Random Forest and Gradient Boosting stand out and the observed enhancement of their precision confirms their usefulness in tasks with high prediction requirements.



**Figure 5.12: Model Performance Comparison Plot (MAE and RMSE)**

This bar chart shows the results of machine learning models in terms of MAE and Root Mean Square Errors. Hence, while evaluating the MAE and RMSE, Random Forest again shows very few errors and is the most accurate model among all, followed by XGBoost and Gradient Boosting. Linear Regression and Support Vector Regressor have considerably higher error meaning that the Ensemble models outperform them.

**6: Analysis And Discussion**

1. **6.1 Interpretation of Results**

The findings also confirm that, when it comes to weather data prediction, Random Forest and Gradient Boosting are undisputed winners in terms of ensemble learning methods because of their capacity to avoid complexity. Hyperparameter tuning takes care of even this increasing their accuracy to a point where almost all error measures are zero and R² is one. XGBoost equally proves to be very efficient while Linear Regression and Support Vector Regressor despite being as accurate as them are relatively slower because their algorithms are not as complex as those of the other four models. The backtesting analysis reasserts the idea of optimum predictive calibration with fine-tuned models. These results show that, for accurate and reliable weather prediction, one needs to invest in complex algorithms and optimized values of hyperparameters are essential.

1. **6.2 Comparison with Literature**

These results support the conclusions drawn in previously published papers, which claim that such ensemble learning algorithms as Random Forest, Gradient Boosting, or XGBoost produce better accuracy. These models can increase the accuracy when dealing with large datasets and know how to manage non-linear relationships between variables and advanced feature interactions, according to previous research. For example, several studies revealing meteorological forecasting show that Random Forest is stable throughout Gradient Boosting is accurate in capturing the temperature change, as shown in this study. The proposed tuned ensemble models achieved near-perfect R² values and minimum error measures for the datasets considered in this work as similar studies show that ensemble models remain the leaders. The improvement of hyperparameter tuning, which has been shown in this paper, is also evidenced in many other studies, which point to the importance of the parameter in enhancing the accuracy of the model and diminishing prediction errors. While other models such as Linear Regression, despite the model's ability to identify straightforward trends, perform significantly worse in sets with complicated dependencies. This is in synergy with other studies which have shown that ensemble methods perform better as far as reliability for the weather prediction tasks is concerned. As proposed by the current literature, this study supports the notion that with suitable optimization, high-order ensembling methods are critical predictors for accurate and robust forecasting in the challenging space of meteorological applications.

1. **6.3 Limitations of Results**

The results show high accuracy of the models, though there are certain shortcomings. First, the data sample may contain inadequate exposure to severe weather, or it may not cover some of the instances at all, thus restricting the results' applicability. Second, the risk of overfitting could be an issue despite the precision of the model such as the Random Forest and the Gradient Boosting. However, the models depend on past data, and this fails to explain changes in climate triggered by climate change, and global warming, among others. Moreover, the tuning of hyperparameters raises the performance, which however can be burdensome in terms of computational complexity and does not guarantee its applicability to another dataset. Moreover, the study lacks a discussion of the interpretability of models, which plays a vital role in technical implementation.

1. **6.4 Alignment with Objectives**

***Objective 1: To collect a suitable weather dataset for weather forecasting***

The objective of the study was to obtain the set of weather variables which are crucial for the model - temperature, humidity, and pressure. Completeness of the dataset allowed for accurate model and evaluation, and therefore the results were credible. While its temporal grain and extent afforded a rich training, validation, and testing dataset, it is most suited for rich weather prediction applications.

***Objective 2: To choose the important weather parameters like cloud formation, humidity, pressure, etc., for developing the weather prediction model***

The necessary parameters like temperature, humidity, and pressure changes were chosen taking into account the prediction of the weather. This selection also helped to enhance the models' performance given that the features that had a great impact were selected. Effective feature engineering process made a selection of the relevant features and minimize as much as possible the noise present in the datasets so as to augment the predictive power of the developed models.

***Objective 3: To choose the best ML algorithm for developing the model***

By evaluating multiple algorithms, including Linear Regression, Random Forest, Gradient Boosting, and XGBoost, the objective was achieved. Ensemble methods like Random Forest emerged as the most accurate due to their low errors and robust generalization. Hyperparameter tuning further refined these models, demonstrating their superiority for reliable and precise weather forecasting.

1. **6.5 Addressing the Research Question**

This research accomplishes this by providing an understanding of important variables measuring for instance temperature, humidity, pressure, and clouds in weather prediction. To achieve the best results and capture the important parameters, feature selection and engineering were performed. In the assessment of different supervised machine learning techniques, the study pointed out Random Forest and Gradient Boosting as best fitting algorithms for now; the two models display capability in handling intricate patterns. The results have further affirmed the relevance of these variables for obtaining precise and accurate weather forecasts

**7: Conclusion**

1. **7.1 Conclusions**

In conclusion, using temperature, humidity, pressure, and formation of clouds as some of the important predictors, this study found out that it is possible to use machine learning models to forecast weather patterns. Out of the evaluated models, Random Forest and Gradient Boosting were most accurate with nearly perfect errors and R², thus their ability to model intricate regularities of the data. This was done through additional training, which makes the predictions provided even more precise than actual statistics. There were also models such as Linear Regression and Support Vector Regressor that also gave good results but the errors observed were relatively high, which again highlights the importance of the usage of complex ensemble techniques for accurate forecasts. Specifically the process of hyperparameter tuning was found vital in enhancing the performance of the models, highlighting the significance of the optimization step, especially in predictive tasks. This work shows the importance of analyzing the features of weather forecasts and the means of evaluating a model and adjusting it accordingly in order to obtain accurate results. This work helps to develop machine learning-based forecasting systems and provides benefits for weather monitoring and decision-making.

1. **7.2 Applications and Real-World Situations**

The models which have been determined in this research work have actual world utility in areas of meteorology, self-driving vehicles, and climatology. The low error rate in Random Forest and Gradient Boosting can improve precision in sensitive areas like indicating the correct course or during calamity response. These models are also suitable in such industries as agriculture, in that, the weather plays an important parameter in crop production planning. Due to this, they are suitable for tasks that require high precision in capturing element interactions that may not be linear and hence provide accurate decisions regarding the environment where weather conditions have a significant influence.

1. **7.3 Future Work**

Future work can also be done to increase the model scale to accommodate larger datasets. Further improvement might be achieved by incorporating additional weather parameters. More complex pattern recognition could be presented using deep learning techniques. Real-time weather prediction using live data streams was left for further research. By improving model interpretability, users can be better at understanding the predictions. Satellite and remote sensing data can be integrated into future models for improved forecasting. The generalization can be increased by expanding the models to different geographical regions. Adding such external factors as sudden climate changes can increase prediction reliability. It is also necessary to further optimize computational efficiency for faster predictions. These advancements will integrate them to refine the weather forecasting models and their real applications.

1. **7.4 Recommendations**

More diverse datasets will help with better weather forecasting models. More features such as cloud cover, wind patterns and geographical data can be included to give better predictions. Robustness requires data from multiple sources integrated. It works by trying to reduce the overfitting by picking relevant features for the model. The model can't be an outdated version of real-world data that needs to be updated regularly. It is important to test models on other regions to evaluate their ability to generalize. This will guarantee scalability so it will be easier to process large datasets with. Moreover, further refinement of the models can be made using high-fidelity machine learning methods, such as deep learning. Real-time data integration from weather stations and satellites can improve the reliability of predictions in the future. The way of decision-making should be transparent to improve trust from users. Strategies to connect with meteorologists for deeper insights into the weather parameters are required for forecasting.

It is also recommended for future work to integrate the SHAP values or LIME to explain the models’ decisions and provide the stakeholders with the model’s interpretations. However, accessibility and scalability can be even further improved by building real-time data processing on cloud platforms and by deploying predictive models as services. Increasing blocks of computation’s energy efficiency and searching for collaborations with local meteorological departments can additionally enhance the significance and efficacy of the weather forecasting systems.

**Reference List**

Badjie, B., Cecílio, J. and Casimiro, A., 2024. Adversarial attacks and countermeasures on image classification-based deep learning models in autonomous driving systems: A systematic review. *ACM Computing Surveys*, *57*(1), pp.1-52.<https://dl.acm.org/doi/pdf/10.1145/3691625>

Bag, S., Rahman, M.S., Srivastava, G., Chan, H.L. and Bryde, D.J., 2022. The role of big data and predictive analytics in developing a resilient supply chain network in the South African mining industry against extreme weather events. *International Journal of Production Economics*, *251*, p.108541.<https://researchonline.ljmu.ac.uk/id/eprint/17010/3/The%20role%20of%20big%20data%20and%20predictive%20analytics%20in%20developing%20a%20resilient%20supply%20chain%20network.pdf>

Bochenek, B. and Ustrnul, Z., 2022. Machine learning in weather prediction and climate analyses—applications and perspectives. Atmosphere, 13(2), p.180. [https://www.mdpi.com/2073-4433/13/2/180](https://www.mdpi.com/2073-4433/13/2/180%20)

Chantry, M., Hatfield, S., Dueben, P., Polichtchouk, I. and Palmer, T., 2021. Machine learning emulation of gravity wave drag in numerical weather forecasting. Journal of Advances in Modeling Earth Systems, 13(7), p.e2021MS002477. [https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021MS002477](https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021MS002477%20)

Chattopadhyay, A., Nabizadeh, E. and Hassanzadeh, P., 2020. Analog forecasting of extreme‐causing weather patterns using deep learning. Journal of Advances in Modeling Earth Systems, 12(2), p.e2019MS001958. [https://agupubs.onlinelibrary.wiley.com/doi/pdfdirect/10.1029/2019MS001958](https://agupubs.onlinelibrary.wiley.com/doi/pdfdirect/10.1029/2019MS001958%20)

Chen, L., Han, B., Wang, X., Zhao, J., Yang, W. and Yang, Z., 2023. Machine learning methods in weather and climate applications: A survey. *Applied Sciences*, *13*(21), p.12019. <https://www.mdpi.com/2076-3417/13/21/12019/pdf>

Cho, D., Yoo, C., Im, J. and Cha, D.H., 2020. Comparative assessment of various machine learning‐based bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas. Earth and Space Science, 7(4), p.e2019EA000740. [https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019EA000740](https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019EA000740%20)

Cl.cam.ac.uk, (2024), Accessed from: [https://www.cl.cam.ac.uk/weather/](https://www.cl.cam.ac.uk/weather/%20) [Accessed on: 09/10/2024]

Correia, A., Peharz, R. and de Campos, C.P., 2020. Joints in random forests. *Advances in neural information processing systems*, *33*, pp.11404-11415.<https://proceedings.neurips.cc/paper_files/paper/2020/file/8396b14c5dff55d13eea57487bf8ed26-Paper.pdf>

Donatti, C.I., Nicholas, K., Fedele, G., Delforge, D., Speybroeck, N., Moraga, P., Blatter, J., Below, R. and Zvoleff, A., 2024. Global hotspots of climate-related disasters. *International Journal of Disaster Risk Reduction*, *108*, p.104488.<https://www.sciencedirect.com/science/article/pii/S2212420924002504>

Dwivedi, Y.K., Sharma, A., Rana, N.P., Giannakis, M., Goel, P. and Dutot, V., 2023. Evolution of artificial intelligence research in Technological Forecasting and Social Change: Research topics, trends, and future directions. *Technological Forecasting and Social Change*, *192*, p.122579.<https://www.sciencedirect.com/science/article/pii/S0040162523002640>

Effrosynidis, D., Spiliotis, E., Sylaios, G. and Arampatzis, A., 2023. Time series and regression methods for univariate environmental forecasting: An empirical evaluation. *Science of The Total Environment*, *875*, p.162580.<https://www.sciencedirect.com/science/article/pii/S0048969723011968>

Fathi, M., Haghi Kashani, M., Jameii, S.M. and Mahdipour, E., 2022. Big data analytics in weather forecasting: A systematic review. *Archives of Computational Methods in Engineering*, *29*(2), pp.1247-1275.<https://e-tarjome.com/storage/panel/fileuploads/2022-01-15/1642249339_E15948.pdf>

Grönquist, P., Yao, C., Ben-Nun, T., Dryden, N., Dueben, P., Li, S. and Hoefler, T., 2021. Deep learning for post-processing ensemble weather forecasts. Philosophical Transactions of the Royal Society A, 379(2194), p.20200092. [https://royalsocietypublishing.org/doi/full/10.1098/rsta.2020.0092](https://royalsocietypublishing.org/doi/full/10.1098/rsta.2020.0092%20)

Han, Y., Mi, L., Shen, L., Cai, C.S., Liu, Y., Li, K. and Xu, G., 2022. A short-term wind speed prediction method utilizing novel hybrid deep learning algorithms to correct numerical weather forecasting. Applied Energy, 312, p.118777. <https://www.researchgate.net/profile/Guoji-Xu-2/publication/358632567_A_short-term_wind_speed_prediction_method_utilizing_novel_hybrid_deep_learning_algorithms_to_correct_numerical_weather_forecasting/links/620c6bf2afa8884cabe7a14b/A-short-term-wind-speed-prediction-method-utilizing-novel-hybrid-deep-learning-algorithms-to-correct-numerical-weather-forecasting.pdf>

Hewage, P., Trovati, M., Pereira, E. and Behera, A., 2021. Deep learning-based effective fine-grained weather forecasting model. Pattern Analysis and Applications, 24(1), pp.343-366. [https://link.springer.com/content/pdf/10.1007/s10044-020-00898-1.pdf](https://link.springer.com/content/pdf/10.1007/s10044-020-00898-1.pdf%20)

Huang, D., Wang, S. and Liu, Z., 2021. A systematic review of prediction methods for emergency management. *International Journal of Disaster Risk Reduction*, *62*, p.102412. <https://ira.lib.polyu.edu.hk/bitstream/10397/90654/1/Huang_Systematic_Review_Prediction.pdf>

Hummel, P., Braun, M. and Dabrock, P., 2021. Own data? Ethical reflections on data ownership. <https://link.springer.com/content/pdf/10.1007/s13347-020-00404-9.pdf>

Jakaria, A.H.M., Hossain, M.M. and Rahman, M.A., 2020. Smart weather forecasting using machine learning: a case study in tennessee. arXiv preprint arXiv:2008.10789. [https://arxiv.org/pdf/2008.10789](https://arxiv.org/pdf/2008.10789%20)

Jaseena, K.U. and Kovoor, B.C., 2022. Deterministic weather forecasting models based on intelligent predictors: A survey. *Journal of king saud university-computer and information sciences*, *34*(6), pp.3393-3412.<https://www.sciencedirect.com/science/article/pii/S1319157820304729>

Kaggle, (2024), Accessed from: <https://www.kaggle.com/datasets/thedevastator/weather-prediction/data?select=weather_prediction_dataset.csv> [Accessed on: 09/10/2024]

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P. and Hatfield, S., 2024. Neural general circulation models for weather and climate. *Nature*, *632*(8027), pp.1060-1066.<https://www.nature.com/articles/s41586-024-07744-y>

Lakra, K. and Avishek, K., 2022. A review on factors influencing fog formation, classification, forecasting, detection and impacts. Rendiconti Lincei. Scienze Fisiche e Naturali, 33(2), pp.319-353.<https://link.springer.com/article/10.1007/s12210-022-01060-1>

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W. and Merose, A., 2023. Learning skillful medium-range global weather forecasting. Science, 382(6677), pp.1416-1421. [https://arxiv.org/pdf/2212.12794](https://arxiv.org/pdf/2212.12794%20)

Markets and Markets (2024). *Weather Forecasting Systems Market Size, Share, and Industry Analysis and Market Forecast to 2025 | MarketsandMarketsTM*. [online] www.marketsandmarkets.com. Available at: <https://www.marketsandmarkets.com/Market-Reports/meteorological-weather-forecasting-systems-market-29645152.html>

Merz, B., Kuhlicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D.N., Domeisen, D.I., Feser, F., Koszalka, I., Kreibich, H. and Pantillon, F., 2020. Impact forecasting to support emergency management of natural hazards. *Reviews of Geophysics*, *58*(4), p.e2020RG000704. <https://agupubs.onlinelibrary.wiley.com/doi/pdfdirect/10.1029/2020rg000704>

Parashar, A., Parashar, A., Ding, W., Shabaz, M. and Rida, I., 2023. Data preprocessing and feature selection techniques in gait recognition: A comparative study of machine learning and deep learning approaches. *Pattern Recognition Letters*, *172*, pp.65-73.<https://www.researchgate.net/profile/Imad-Rida/publication/371113483_Data_Preprocessing_and_Feature_Selection_Techniques_in_Gait_Recognition_A_Comparative_Study_of_Machine_Learning_and_Deep_Learning_Approaches/links/6477558ed702370600c50608/Data-Preprocessing-and-Feature-Selection-Techniques-in-Gait-Recognition-A-Comparative-Study-of-Machine-Learning-and-Deep-Learning-Approaches.pdf?origin=journalDetail&_tp=eyJwYWdlIjoiam91cm5hbERldGFpbCJ9>

Petzold, J., Andrews, N., Ford, J.D., Hedemann, C. and Postigo, J.C., 2020. Indigenous knowledge on climate change adaptation: A global evidence map of academic literature. *Environmental Research Letters*, *15*(11), p.113007.<https://iopscience.iop.org/article/10.1088/1748-9326/abb330&title=iopscience.iop.org>

Rasp, S., Dueben, P.D., Scher, S., Weyn, J.A., Mouatadid, S. and Thuerey, N., 2020. WeatherBench: a benchmark data set for data‐driven weather forecasting. Journal of Advances in Modeling Earth Systems, 12(11), p.e2020MS002203.[https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020ms002203](https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020ms002203%20%20)

Shouman, E.R.M., 2024. Solar power prediction with artificial intelligence. In Advances in Solar Photovoltaic Energy Systems. IntechOpen. <https://www.intechopen.com/chapters/1152404>

Singh, N., Chaturvedi, S. and Akhter, S., 2019, March. Weather forecasting using machine learning algorithm. In 2019 International Conference on Signal Processing and Communication (ICSC) (pp. 171-174). IEEE. [https://dcs.datapro.in/wp-content/uploads/2023/11/36-Weather-Forecasting-Using-Machine-Learning.pdf](https://dcs.datapro.in/wp-content/uploads/2023/11/36-Weather-Forecasting-Using-Machine-Learning.pdf%20%20)

Singh, U., Rizwan, M., Alaraj, M. and Alsaidan, I., 2021. A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments. *Energies*, *14*(16), p.5196.<https://www.mdpi.com/1996-1073/14/16/5196>

Wang, B., Lu, J., Yan, Z., Luo, H., Li, T., Zheng, Y. and Zhang, G., 2019, July. Deep uncertainty quantification: A machine learning approach for weather forecasting. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 2087-2095). [https://arxiv.org/pdf/1812.09467](https://arxiv.org/pdf/1812.09467%20)

Weyn, J.A., Durran, D.R., Caruana, R. and Cresswell‐Clay, N., 2021. Sub‐seasonal forecasting with a large ensemble of deep‐learning weather prediction models. Journal of Advances in Modeling Earth Systems, 13(7), p.e2021MS002502. [https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021MS002502](https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021MS002502%20)

Yazdani-Asrami, M., Sadeghi, A., Song, W., Madureira, A., Murta-Pina, J., Morandi, A. and Parizh, M., 2022. Artificial intelligence methods for applied superconductivity: material, design, manufacturing, testing, operation, and condition monitoring. *Superconductor Science and Technology*, *35*(12), p.123001.<https://iopscience.iop.org/article/10.1088/1361-6668/ac80d8/meta>

Zaineldeen, S., Hongbo, L., Koffi, A.L. and Hassan, B.M.A., 2020. Technology acceptance model’concepts, contribution, limitation, and adoption in education. Universal Journal of Educational Research, 8(11), pp.5061-5071. <https://www.researchgate.net/profile/Samar-Zaineldeen/publication/345004753_Technology_Acceptance_Model'_Concepts_Contribution_Limitation_and_Adoption_in_Education/links/5f9c0445299bf1b53e514fa8/Technology-Acceptance-Model-Concepts-Contribution-Limitation-and-Adoption-in-Education.pdf>

Zhang, L., Wen, J., Li, Y., Chen, J., Ye, Y., Fu, Y. and Livingood, W., 2021. A review of machine learning in building load prediction. *Applied Energy*, *285*, p.116452.<https://www.sciencedirect.com/science/article/am/pii/S0306261921000209>