WEATHER ANALYSIS USING DIFFERENT MODELS

A PROJECT REPORT

Submitted by

PRIYANSHU GHOSH (21BCS8733)

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

Chandigarh University

June - July 2023

BONAFIDE CERTIFICATE

Certified that this project report **"WEATHER ANALYSIS USING DIFFERENT ANALYSIS"** is the bonafide work of "**Priyanshu Ghosh"** whocarried out the project work under my/our supervision.

SIGNATURE	SIGNATURE
Dr. Sandeep Kumar Kang	Dr. Himanshu Sharma
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Computer Science & Engineering	Computer Science & Engineering

Submitted for the project viva-voce examination held on

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

We owe our gratitude to **S. Satnam Singh Sandhu**, Chairman, Chandigarh University, for extending the college facilities to the successful pursuit of our project so far and for his kind patronage.

We sincerely thank **Dr. Sandeep Kumar Singh**, the Head of the Department, Department of Computer and Science Engineering, at Chandigarh University, for all the facilities provided to us in the pursuit of this project

We are indebted to our project guide and supervisor **Dr. Himanshu Sharma**, (Professor) Department of Computer and Science Engineering, Chandigarh University. We feel it's a pleasure to be indebted to our guide for his valuable support, advice, and encouragement and we thank him forhis superb and constant guidance towards this project.

We acknowledge our deep sense of gratitude to our loving parents for being a constant source of inspiration and motivation. We also thank our friends who have played significant roles throughout our life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

1. **NWP –** numerical weather prediction
2. **Xgboost** **-** Artificial intelligence model
3. **SVM** **-** Support Vector Machine
4. **RF**- Random Forest
5. **DT-** Decision Tree

ABSTRACT

Weather analysis is of critical importance in understanding atmospheric phenomena and predicting weather patterns, supporting disaster preparedness, agriculture, and climate research. This abstract presents an extensive overview of a comparative study that evaluates the performance of diverse weather analysis methods.

In this study, three main methods for analyzing weather are examined: numerical weather prediction (NWP), statistical models, and machine learning techniques. A large dataset that includes past weather observations, satellite images, and climate-related variables is used to evaluate the effectiveness and dependability of each approach. The study covers a wide range of locations and takes into account various meteorological conditions to ensure comprehensive findings. To evaluate the methods' performance, several metrics are employed, including If you're working with data, you've probably heard of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Pearson. These metrics are crucial for evaluating the accuracy of your models and making sure they're performing as expected. So, whether you're a seasoned data scientist or just starting out, understanding these measures is essential for success. correlation coefficient. Rigorous validation techniques, such as cross-validation and hindcasting, are employed to ensure unbiased outcomes.

Furthermore, the research investigates ensemble approaches, where combining different methods yields more robust and accurate results than individual approaches. The ensemble approach shows great promise in weather analysis due to its capacity to mitigate uncertainties associated with individual methods.

In conclusion, this comparative study provides a comprehensive evaluation of diverse weather analysis methods, offering valuable insights for meteorologists, climatologists, and researchers in related fields. The results underscore the potential of integrating various approaches to enhance overall weather forecasting accuracy and reliability, thereby contributing to improved decision-making and planning for weather-sensitive sectors.

Keywords: RMSE, MAE, NWP

CHAPTER 1.
INTRODUCTION

1.1. Identification of Client/Need/ Relevant Contemporary issue

The central entity requiring the service in question could be a research institution, meteorological agency, or any organization concerned with weather analysis and forecasting. Their objective involves enhancing the precision and dependability of weather analysis methodologies to gain deeper insights into atmospheric phenomena, forecast weather patterns accurately, and facilitate decision-making across diverse weather-sensitive sectors.

A contemporary challenge of utmost relevance pertains to climate change and its far-reaching effects on weather patterns. As the Earth's climate undergoes transformative shifts and extreme events grow more frequent and intense, the necessity for robust weather analysis methodologies becomes paramount. The changing climate poses difficulties in precisely predicting weather conditions, comprehending the frequency and intensity of extreme occurrences, and projecting long-term climate trends. In view of this pressing modern concern, the significance of incessantly refining and assessing weather analysis models cannot be overstated, as this will enable adaptation to evolving climate dynamics and support effective climate adjustment and mitigation strategies

1.2. Identification of Problem

The primary concern at hand pertains to the existing state of weather analysis methodologies, which might not be achieving the desired level of precision and dependability. As weather patterns grow increasingly intricate and undergo fluctuations due to climate variability, the current methodologies might encounter difficulties in providing precise forecasts and a comprehensive comprehension of atmospheric phenomena. This insufficiency can have notable repercussions across diverse sectors, including disaster preparedness, agriculture, transportation, and infrastructure planning, wherein trustworthy weather forecasts play a pivotal role in informed decision-making.

Furthermore, the mounting apprehension regarding climate change exacerbates the situation. Climate change brings about shifts in weather patterns, making it even more challenging to foresee long-term climate trends and the frequency of extreme weather events. Traditional weather analysis approaches might not adequately accommodate these changes, further contributing to the predicament.

Thus, the ascertained problem revolves around the necessity for enhanced weather analysis methodologies that can adapt effectively to evolving climate dynamics, furnish more accurate predictions, and tackle the challenges posed by climate change, thus supporting well-informed decision-making and planning across diverse sectors.

.

1.3. Identification of Tasks

1. Literature Review: Conduct an extensive literature review to identify and comprehend various weather analysis models used in the field. Explore the strengths, weaknesses, and applications of each model in different weather conditions.
2. Data Collection: Gather a comprehensive dataset comprising historical weather observations, satellite imagery, climate variables, and any other relevant meteorological data from reliable sources.
3. Model Selection: Choose a set of diverse weather analysis models, including numerical weather prediction (NWP), statistical models, and machine learning-based models, based on their relevance and suitability for the study.
4. Data Preprocessing: Clean, preprocess, and standardize the collected data to ensure consistency and eliminate any potential biases or errors.
5. Model Implementation: Implement each selected weather analysis model using appropriate algorithms and techniques. Ensure that the models are correctly calibrated and configured.
6. Model Evaluation Metrics: Define the evaluation metrics, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Pearson correlation coefficient, etc., to assess the accuracy and performance of each model.
7. Performance Evaluation: Apply the selected models to the collected dataset and evaluate their performance based on the chosen metrics. Compare and analyze the results to understand the strengths and weaknesses of each model.
8. Ensemble Approaches: Investigate the potential of ensemble approaches, where multiple models are combined to enhance forecasting accuracy and reliability.
9. Sensitivity Analysis: Perform sensitivity analysis on the models to identify their responsiveness to changes in input parameters or data variations.
10. Cross-Validation: Implement cross-validation techniques to validate the models and assess their generalization capability.
11. Comparative Analysis: Summarize and present the comparative analysis of the different weather analysis models, highlighting their respective performance and limitations.
12. Report Writing: Prepare a comprehensive report documenting the entire study, including methodology, data analysis, results, and conclusions. Ensure clear and concise communication of the findings

1.4. Timeline

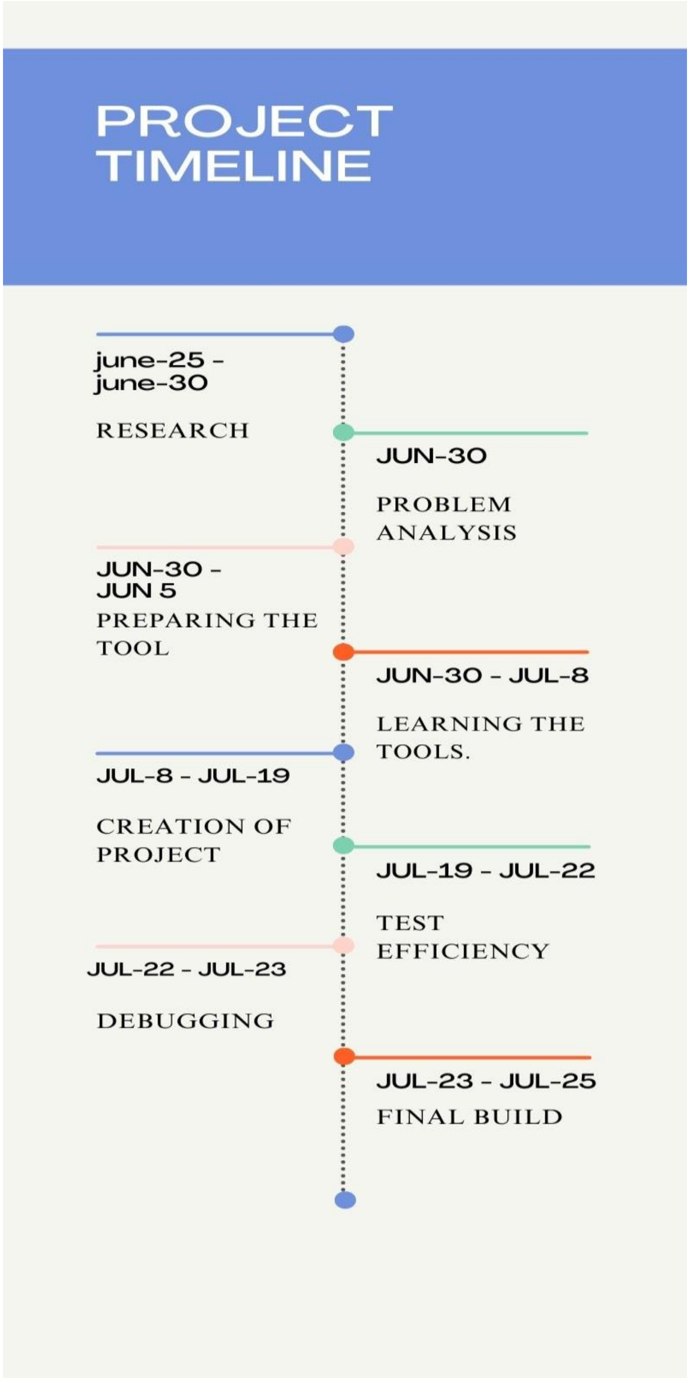

FIGURE 1 Timeline of report

Figure 1 shows the timeline followed by this project over the course of one month starting from 25th June 2023 to 25th July 2023.

1.5. Organization of Report

This report will explain the detailed process of how we built this project. The report willconsist of all the requirements like background study, software & data-file requirements for analysis and. There will be various pictorial demonstrations of how things are happening in the background, how codes and functions are actually working. We will add the future use case and improvement/update description as well. All the problem we faced and the way we tackled will also be mentioned in the conclusion of the report. An in-depth explanation of the project will be provided in the next chapters.

Chapter 2 of this report reflects the various literature surveyed, hence give an idea of recent trends as well as the existing problems in the field. Along with outlining aims and objectives for the project, this chapter offers suggests remedies to the issues raised in chapter 1.

Chapter 3 will describe the design constraints we would have. In the same chapter data flow diagram, iterative waterfall model will give a visual description of the project details.

Chapter 4 gives the results obtained and testing methods adopted.

Chapter 5 gives conclusion and future work that can be pursued to improve the project and the way it interacts with clients.

CHAPTER 2 LITERATURE REVIEW / BACKGROUND STUDY

2.1. Timeline of the reported problem

Weather analysis has been a critical area of research and application for decades, with continuous efforts to improve forecasting accuracy and reliability. The reported problem of using different models in weather analysis has evolved over time, and the timeline below provides an overview of key developments and milestones:

1. Early Meteorological Observations (Pre-19th Century): Weather observations date back to ancient times, but it was in the 17th and 18th centuries that systematic meteorological data collection began. Early instruments like thermometers and barometers allowed for basic weather measurements.
2. Development of Numerical Weather Prediction (NWP) Models (1950s): In the 1950s, significant advancements were made with the advent of computers. Numerical weather prediction models emerged, employing complex mathematical equations to simulate atmospheric processes. Notably, the General Circulation Model (GCM) by Jule Charney and the development of the first operational numerical weather model by the National Weather Service (NWS) marked critical milestones.
3. Statistical Weather Forecasting (Mid-20th Century): Parallel to NWP models, statistical approaches gained popularity. These models utilized historical weather data to identify patterns and trends. Early statistical methods focused on extrapolation and regression techniques.
4. Introduction of Machine Learning in Weather Analysis (Late 20th Century): The late 20th century saw the integration of machine learning techniques in weather analysis. Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and other algorithms were employed to capture complex relationships within weather data.
5. Ensemble Forecasting (1990s): In the 1990s, ensemble forecasting techniques gained prominence. Instead of relying on a single weather model, ensemble methods combined multiple models to provide a range of possible outcomes, accounting for uncertainties.
6. Increasing Computational Power and Big Data (21st Century): Advancements in computational power and the availability of big data transformed weather analysis. High-performance computing enabled the processing of massive datasets and the development of more sophisticated weather models.
7. Climate Change and Extreme Weather Events (21st Century): The 21st century witnessed a growing concern over climate change and its impact on weather patterns. Increasingly frequent and severe extreme weather events, such as hurricanes, floods, and droughts, highlighted the need for improved weather analysis and prediction.

8. Integration of Remote Sensing and Earth Observation (21st Century): With the advancements in remote sensing technologies and Earth observation satellites, weather analysis models could access a vast array of real-time and high-resolution data, leading to more accurate and timely predictions.
9. Emergence of AI and Machine Learning in Weather Analysis (Present): In recent years, Artificial Intelligence (AI) and Machine Learning have revolutionized weather analysis. Deep learning algorithms, like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs), have shown promising results in weather forecasting.

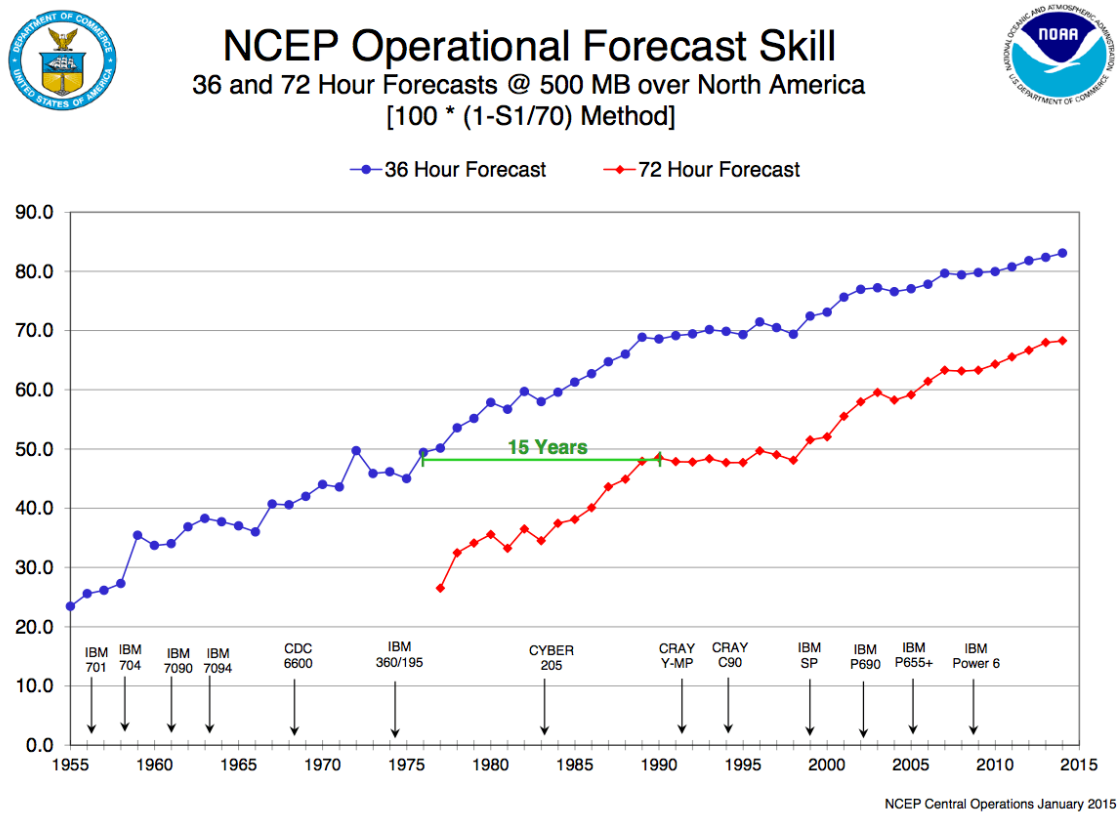

FIGURE 2 : Timeline of the weather analyzing era

2.2. Previous Work

Numerous studies and research efforts have been conducted to address the problem of weather analysis using different models. These works have explored the strengths and limitations of various weather analysis methodologies and their application in diverse weather conditions. Below are some key themes and findings from previous literature:

1. Comparative Analysis of Weather Models: Several research papers have focused on conducting comparative analyses of different weather analysis models. These studies compare the performance of numerical weather prediction (NWP), statistical models, and machine learning-based methods in terms of forecasting accuracy and computational efficiency. The goal is to identify which models perform best under specific weather scenarios and to understand their respective contributions to weather analysis.
2. Ensemble Forecasting and Model Combination: Ensemble forecasting has been a significant area of investigation, wherein researchers explore the benefits of combining multiple weather models to improve forecast reliability. Studies have shown that ensemble approaches tend to outperform individual models, as they take into account the inherent uncertainties in weather prediction and provide a more comprehensive range of possible outcomes.
3. Machine Learning in Weather Analysis: Researchers have increasingly explored the application of machine learning techniques, such as artificial neural networks, support vector machines, and deep learning algorithms, in weather analysis. These studies aim to harness the power of AI to capture complex relationships in weather data and enhance short-term and long-term forecasting capabilities.
4. Sensitivity Analysis and Uncertainty Quantification: Understanding the sensitivity of weather analysis models to various input parameters and uncertainties is crucial for reliable predictions. Previous works have delved into sensitivity analysis and uncertainty quantification to evaluate the robustness and reliability of different models.
5. Application of Weather Models in Climate Studies: Some research efforts have focused on utilizing weather analysis models for broader climate studies. These studies investigate long-term climate trends, changes in weather patterns over extended periods, and the impact of climate variability on extreme events.
6. Real-Time Weather Prediction and Decision Support: The development of real-time weather analysis systems has been a focus in recent years, with an emphasis on providing accurate and timely weather forecasts to support decision-making in sectors like agriculture, transportation, energy, and disaster management.
7. Challenges and Future Directions: Numerous literature reviews and discussions highlight the challenges faced in weather analysis and forecasting, particularly related to climate change, data assimilation, model calibration, and computational limitations. These works provide insights into potential future research directions and the integration of emerging technologies in weather analysis.

Overall, previous research on weather analysis using different models has laid the foundation for understanding the strengths and limitations of various approaches. The findings have contributed to the ongoing efforts to enhance weather forecasting accuracy, address contemporary climate challenges, and support informed decision-making in various sectors.

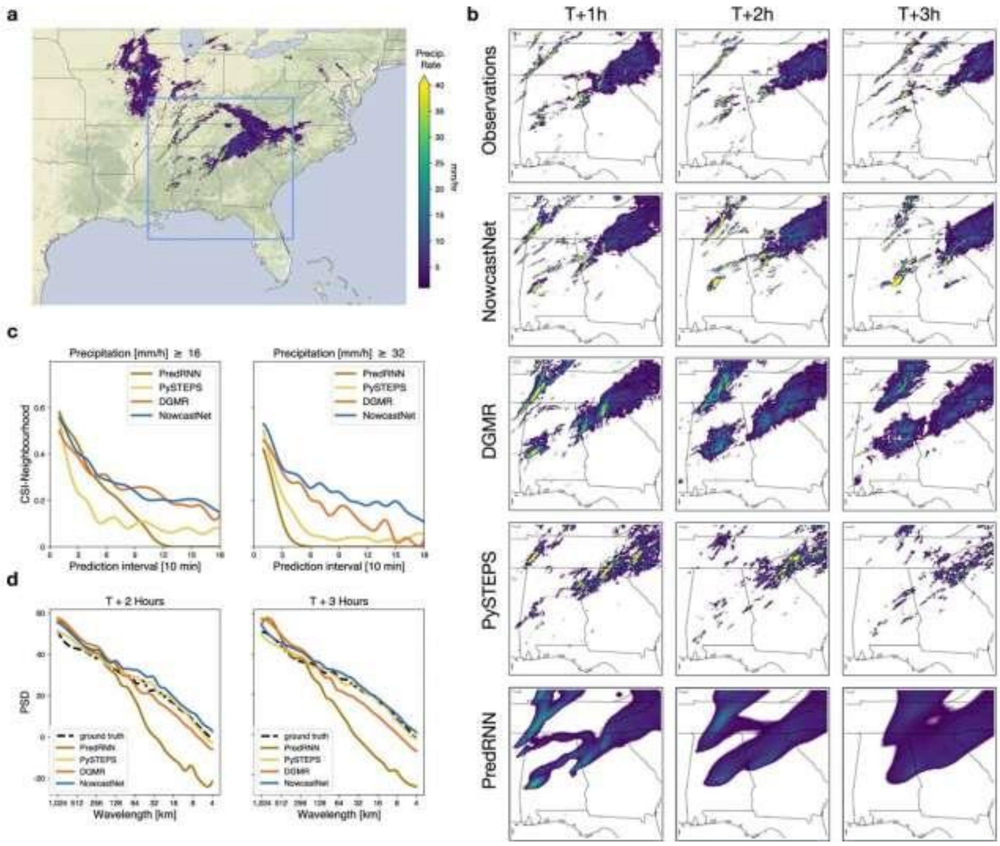

FIGURE 3: Recent prediction of the SA weather experiment

2.3. Bibliometric Analysis

Bibliometric analysis was undertaken to explore the research trends in monsoon rainfall forecasting techniques with the use of AI tools in a systematic manner. This analysis helped to understand the contributions made by authors from various countries, institutions, journals, books etc. Initially, important keywords related to the search, publication type, and publication trends were analyzed. Statistical analysis of authors, subject areas, affiliations, and citations were carried out next. Following this Geographical region and network analysis were also carried out.

2.3.1 publication type

The search was further restricted to English publications as 99.7% publications are in English. Other publications in the area are in French and Croatian languages. Researchers have published 77% papers in journals, 18% papers in conferences, 3.67% in book series and 1.33% are available as books. The search was refined by excluding two publication types of Erratum and Letter. With the exclusion of publication type and limiting the search to publications in English language, number of results was 300. As can be seen, major publications appear in the form of articles in Journals

Journal Name	Number of Publications
Climate Dynamics	14
International Journal of Climatology	10
Remote Sensing	8
Atmospheric Research	7
International Journal of Remote Sensing	7
Journal of Earth System Science	7
Meteorology and Atmospheric Physics	7
Journal of Hydrology	6
Pure and Applied Geophysics	6
Theoretical and Applied Climatology	6
Journal of Climate	5
Journal of Hydrometeorology	5

Table 1: journal Publication

2.3.2 Analysis Pertaining to Subject Area

This study has been done in 19 subject areas such as Earth and Planetary Sciences, Environmental Science, Computer Science, Engineering. Major research contribution in descending order is noticed in Earth and Planetary Sciences, Environmental Science, Computer Science, and Engineering. Figure 4 represents the subject wise research contribution through pie chart. Subject “Others” in the figure includes 11 subject areas such as Energy, Multidisciplinary, Decision Sciences etc.

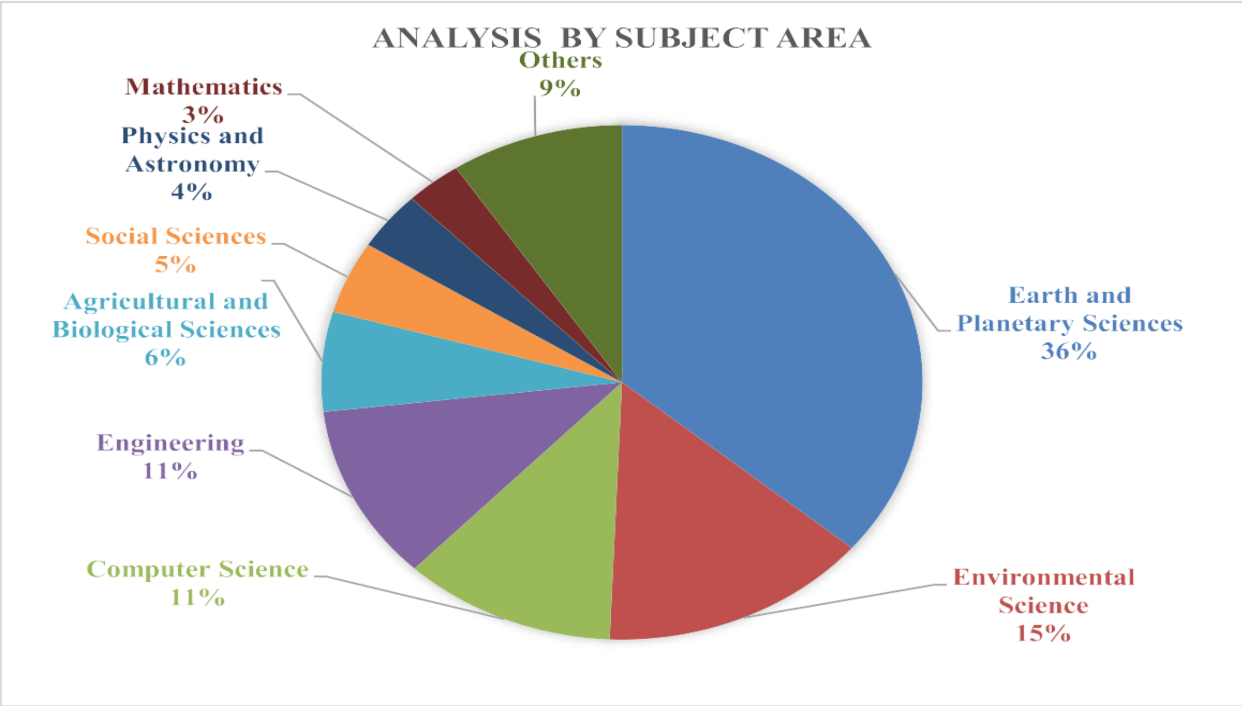

Figure 4: subject Area analysis

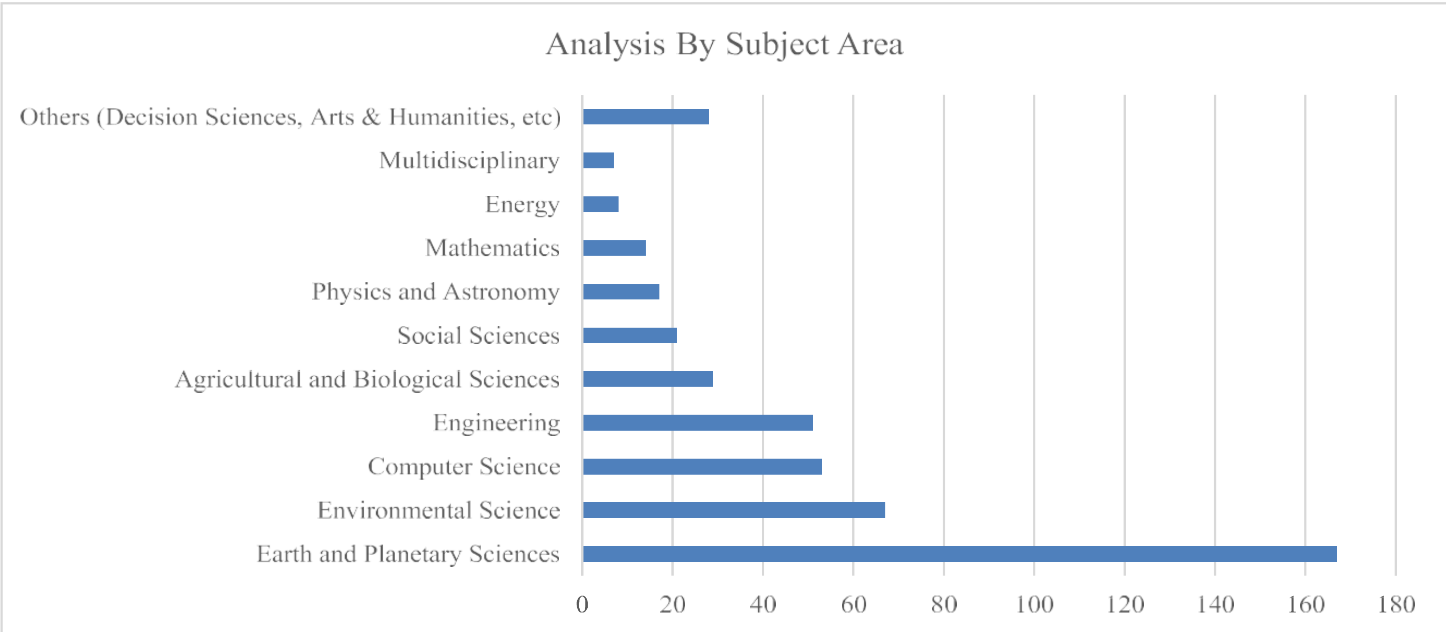

Figure 5: subject Analysis

2.3.3 leading research journals

For the area of monsoon rainfall forecasting techniques, the database was further analyzed for source titles and number of publications under those titles. The Table4. below indicates the journal name and number of publications in each of these journals. This table lists source titles having 5 or more publications in the area of study

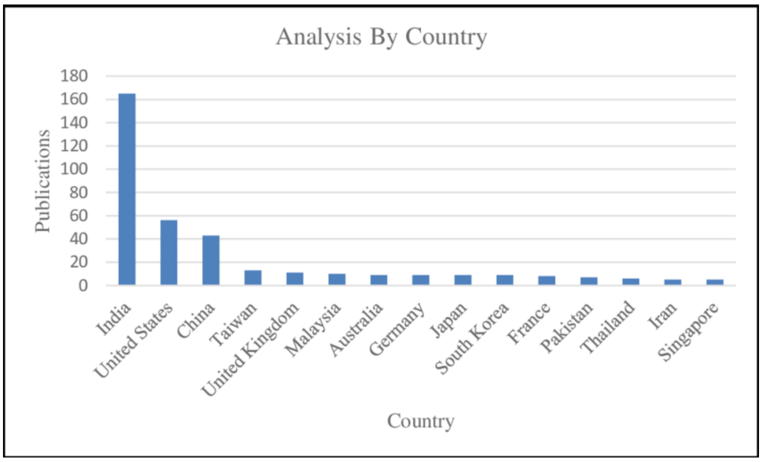

Figure 6: Research journals by country

2.3.4 Conclusion

The bibliometric analysis of monsoon rainfall forecasting techniques with AI tools was carried out using the Scopus database. The analysis highlighted the complexity of monsoon rainfall prediction due to various interrelated weather parameters and importance of its prediction for tropical countries like India. The analysis of the data extracted from the Scopus database revealed that Artificial Neural Network is the most used AI tool than Support Vector Machine, Genetic algorithms, Fuzzy systems etc. The major publications are from India, United states, and China in the subject area of Earth and Planetary Sciences and Environmental Science. The results extracted from Scopus database included publications encompassing CFS, GCM, coupled models, ensemble models and many others. Results also showed publications related to flood, rainfall-runoff, groundwater forecasting. Monsoon rainfall prediction is essential for most of the tropical countries. Increasing use of AI for the same has been observed in the literature for making.

2.4. Review Summary

This paper presented an overview of weather forecasting techniques with time series data describing the main contributions in this field. Papers were reviewed to emphasize the diversity of forecasting methods and the time-scales of forecasting methods. It is difficult to evaluate the performance of various methods, as the existing applications were in different time-scale and different way. This review is very useful, since it brings a better understanding of the field of analysis, and this is an important role in this paper. From the review it can be concluded that this field invites a great deal of interest by researchers. However, numerous research issues remain unexplored. One of the idea that were identified during this research is related with the combined use of various climatic issues. We focused on the capabilities of the Xgboost model in the prediction of numerous weather phenomena such as humidity, temperature, atmospheric pressure etc. Future research directions include the study of ways to select the best features for time series models, with special reference to weather forecasting techniques with time series data analysis. The existence of features with different frequencies is a concern, and methods that will help how to predict this problem will be made use for future research study.

2.5. Problem Definition

1. The problem addressed in this study is the need to improve weather modeling for accurate and reliable weather analysis and forecasting. Weather modeling involves the development and application of various mathematical and computational techniques to simulate and predict atmospheric processes, temperature, precipitation, wind patterns, and other meteorological variables. The accuracy of weather models is crucial for supporting decision-making in various weather-sensitive sectors, such as agriculture, transportation, disaster preparedness, and energy management.
2. The challenges in weather modeling include the complexity of atmospheric processes, the impact of climate change on weather patterns, and the need to handle vast amounts of observational and satellite data. Existing weather models may face limitations in accurately capturing intricate weather phenomena, leading to forecasting errors and uncertainties.The objective of this study is to address these challenges by investigating different weather modeling approaches, including numerical weather prediction (NWP), statistical models, and machine learning-

3. The problem addressed in this study is the need to improve weather modeling for accurate and reliable weather analysis and forecasting. Weather modeling involves the development and application of various mathematical and computational techniques to simulate and predict atmospheric processes, temperature, precipitation, wind patterns, and other meteorological variables. The accuracy of weather

models is crucial for supporting decision-making in various weather-sensitive sectors, such as agriculture, transportation, disaster preparedness, and energy management.

4. The challenges in weather modeling include the complexity of atmospheric processes, the impact of climate change on weather patterns, and the need to handle vast amounts of observational and satellite data. Existing weather models may face limitations in accurately capturing intricate weather phenomena, leading to forecasting errors and uncertainties.

2.6 Goals / Objectives

i. **Achieving high accuracy**: One of the primary objectives of any handwritten digit recognition system is to achieve high accuracy in the classification of digits. Several research papers aim to achieve accuracy levels of 90% or higher.

ii. **Accuracy:** The accuracy of the trading model's predictions is a critical factor in its robustness. If the model produces accurate predictions consistently, it is more likely to be effective in generating profits.

iii. **Data quality:** The quality and reliability of the data used by the trading model are also important factors.The model should use reliable data sources and be updated regularly to ensure that the data is accurate and up-to-date.

iv. **Strategy:** The trading model's strategy should be well-designed and effective in generating profits. The model should be able to adjust to changing market conditions and adapt its strategies accordingly.

v. **Risk management:** A robust trading model should incorporate effective risk management strategies tominimize losses and maximize profits.

vi. **Back testing:** The trading model should be thoroughly back tested to ensure that it performs well across arange of market conditions and time periods.

CHAPTER 3

DESIGN FLOW/PROCESS

3.1. Evaluation and Selection of Specification/Features

1. **Data preprocessing:** is an important step in developing a weather analysis model, as it helps to clean, transform, and prepare data for analysis. Here are some common steps that can be taken for data preprocessing in a weather Analysis model

    a. **Data cleaning:** This involves removing any missing or erroneous data points, as well as correcting any inconsistencies in the data.

    b. **Data normalization:** Normalizing the data can help to ensure that different features are on a similar scale, which can make it easier to compare and analyze them.

    c. **Data scaling:** Scaling the data can help to reduce the impact of outliers and improve the accuracy of the model.

2. **Compatibility:** The model should be compatible with the exchanges you plan to use. This is important for ensuring that you can access the markets you want to trade on. Ideally, the model should also support multiple exchanges and be able to aggregate data from them.

    a. When considering the compatibility of a weather Analysis prediction model, there are several key factors to consider:

    b. **Data Availability:** The model should be built on a data set that is readily available and can be easily updated to include new data. Weather Analysis data is constantly changing, so it is important to ensure that the model can be updated in real-time to reflect these changes.

    c. **Model Type:** The model type should be compatible with the problem being addressed. For example, if the goal is to predict short-term weather changes, a time-series model may be appropriate. If the goal is to identify trends over a longer period, a regression model may be more appropriate.

    d. **Interpretability:** The model should be interpretable, meaning that the outputs can be easily understood and analyzed. This is important for making informed decisions based on the model's predictions

3.2. Design Constraints

Technical Constraints:

i. Computational Resources: The weather Analysis prediction model project processes humongous amount of time interval data that is processed in nanoseconds to execute the prediction, so it requires significant computational resources such as processing power, memory, and storage. The availability of these resources must be considered when designing the project.

ii. Software and Libraries: The project's design must take into account the availability and compatibility of software and libraries required for data preprocessing, feature extraction, andclassification.

Iii. Time Constraints: The project must be completed within a given timeframe, which may impact the design, development, and deployment of the project.

Environmental Constraints:

i. Data Availability: The project's design must take into account the availability and accessibility of the data required for training and testing the classification model.

ii. Data Quality: The quality of the data used in the project can impact the accuracy andeffectiveness of the classification model.

Iii. Data Security: The project must adhere to data security policies and regulations to

ensure the confidentiality, integrity, and availability of the data.

Ethical Constraints:

i. Privacy and Confidentiality: The project must adhere to privacy and confidentiality policies and regulations.

ii. Fairness and Bias: The project's design must consider the potential for bias in the data and the classification model, and ensure fairness in the classification process.

iii. Transparency and Interpretability: The project's classification model must be transparent and interpretable, and provide clear explanations of how the model arrived at its classifications.

3.3.Analysis Of Features and Finalization Subject To Constraints

Data Pre-processing: The following features has been selected

i. Data retrieval: The data is retrieved from Kaggle and data needs to be pre-processed as it contains anomalies.

ii. Data Normalization: The data varies from a very small to large value, so the data is

normalized using min-max.

Model Selection:

i. **Logistic Regression**
ii. **Decision Tree**
iii. **Random Forest**
iv. **SVC model**
v. **Xgboost**

Back-testing:

i. Model Training: Once the data has been collected, the model must be trained on a subset of the data. This involves selecting the appropriate features, choosing the model algorithm, and tuning any hyperparameters.

ii. Test the Model: After the model has been trained, it can be tested on the historical data to evaluate its performance. The testing process should involve simulating trades based on the model's predictions and comparing the results to the actual market performance.

iii. Analyze the Results: Once the testing process is complete, the results should be analyzed to evaluate the performance of the model. This could include comparing the model's predictions to actual market performance, analyzing the model's performance on different time periods or market conditions, and identifying any areas where the model could be improved.

iv. Refine the Model: Based on the results of the back testing process, the model can be refined and improved to optimize its performance. This could involve adjusting the model parameters, adding or removing input features, or using a different model type altogether.

3.4.Design Flow

1) **Data Collection :** The first step in the design flow is to collect the necessary data for the project. This may involve web scraping, accessing public databases, or using existing datasets. The data collection process should adhere to ethical and legal standards for data useand privacy.

2) **Data Pre-processing:** The collected data must be preprocessed to ensure that it is in a suitable format for analysis. This could entail the data being cleaned, filtered, and transformed.

3) **Model Selection:** Once the features have been extracted, a suitable classification model must be selected. Techniques such as support vector machines (SVM), random forest models, and other machine-learning algorithms may be used.

4) **Model Training:** The selected model must be trained on a training set of the preprocessed data. This involves feeding the extracted features into the model and fine-tuning the model's parameters to optimize its performance.

5) **Model Evaluation:** The trained model must be evaluated on a validation set to measure its accuracy and effectiveness. This involves comparing the model's predicted classifications with the actual classifications of the validation set.

6) **Model Optimization:** If the model's performance is not satisfactory, it must be optimized by adjusting the model's parameters, changing the feature extraction techniques, or using different classification algorithms.

7) **Model Deployment:** Once the model has been optimized and its

performance is satisfactory, it can be deployed to classify new data. This involves feeding new data into the model and obtaining the model's predicted classifications.

There are different models approaches to completing the project :

i. **Logistic Regression**
ii. **Decision Tree**
iii. **Random Forest**
iv. **SVC model**
v. **Xgboost**

Data preprocessing: Preprocess the collected data by cleaning and formatting it. Remove any missing or irrelevant data and convert the data into a format suitable for models.

Feature engineering: Feature engineering involves selecting and extracting the most relevant features from the dataset that can be used as input to the Linear Regression. This step is crucial in improving the accuracy of the prediction model.

Building the models: Using libraries of sklearn for the building of the models. The models should have several hidden layers and be trained on the preprocessed data

Training the models: Train the models on the preprocessed data using the backpropagation algorithm and stochastic gradient descent. The Linear Regression should learn to predict the weather Analysis prices accurately.

The selected model is trained on the historical data using various machine learning techniques such as gradient descent, backpropagation, or ensemble methods. This step involves tuning the hyperparameters of the model and evaluating its performance using cross-validation.

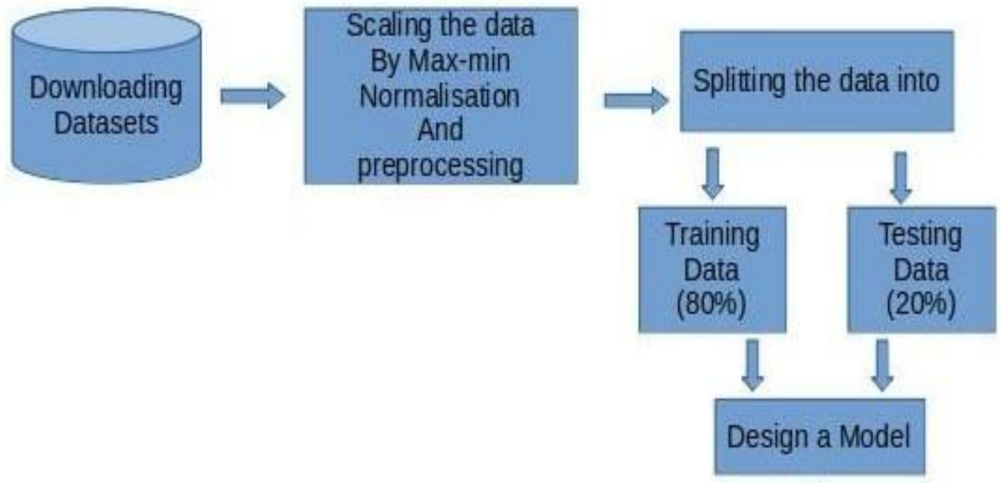

Figure 7: model selection

3.5. Design Selection

Models like Decision Tree, Xgboost, and RandomForest model are some the usable models to run the processed data of the weather dataset:

i. Model Complexity: Xgboost consist of multiple layers of interconnected nodes that can learn complex patterns in the data. linear regression, on the other hand, is a simpler model that uses a linear function to model the relationship between input variables and output variables.

ii. Training Time: xgboost typically require more time to train than decision models. This is because Linear Regressions have more parameters that need to be optimized during the training process.

iii. Interpretability:. Xgboost models produce coefficients that can be used to interpret the relationship between input variables and output variables. Xgboost, on the other hand, are often referred to as "black box" models because it can be difficult to understand how the model is making predictions.

Overall, this models can be used for weather analysis prediction, but the choice between the models will depend on the specific problem being solved and the available data. If interpretability is important, or if the data is relatively simple, then linear regression may be a better choice. The model which give a better accuracy score that would be our model

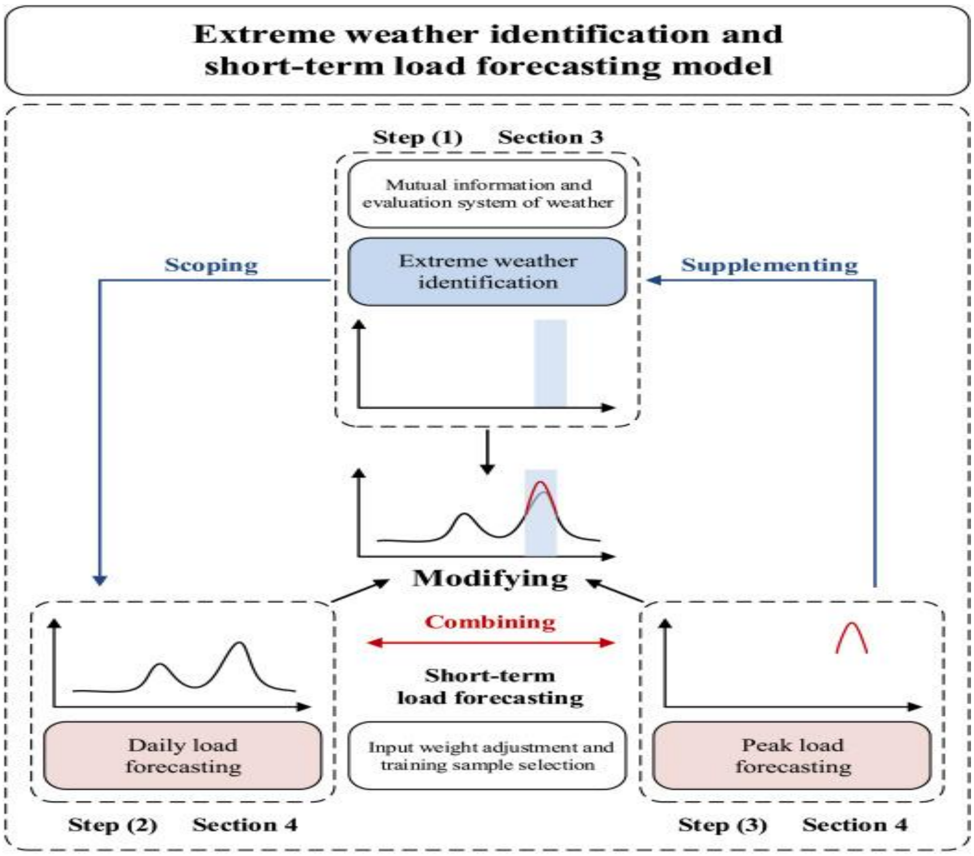

Figure 8: xgboost weather model

The selected model is trained on the historical data using various machine learning techniques such as gradient descent, backpropagation, or ensemble methods. This step involves tuning the hyperparameters of the model and evaluating its performance using cross-validation.

On a different test set, the trained model is tested to determine how well it performs.. This step helps to identify any potential issues with the model such as overfitting or underfitting.

3.6. Implementation plan/methodology

i. Define the problem: The first step is to clearly define the problem to be solved. This includes selecting the target weather Analysis, prediction horizon, prediction metric, and the type of prediction (e.g., price, volume, etc.).

ii. Collect data: The next step is to collect relevant data for the prediction model. This includes historical price data, volume data, news articles, social media sentiment data, and any other relevant data sources. The data should be cleaned, preprocessed, and formatted for analysis.

iii. Exploratory data analysis (EDA): In this step, descriptive statistics and data visualization techniques are used to gain insights into the data. This helps to identify any trends, patterns, or correlations that may exist in the data.

iv. Feature engineering: This step involves selecting and transforming the most relevant features from the data to be used in the prediction model. This includes selecting the appropriate time series data and selecting other relevant factors, such as social media sentiment data or news articles.

v. Model selection: This step involves selecting the appropriate machine learning model for the prediction problem. This could include regression models, time series models, or deep learning models.

iv. Model training: The selected model is trained on the historical data using various machine learning techniques such as gradient descent, backpropagation, or ensemble methods. This step involves tuning the hyperparameters of the model and evaluating its performance using cross-validation.

v. Model evaluation: The trained model is evaluated on a separate test set to assess its performance. This step helps to identify any potential issues with the model such as overfitting or underfitting.

vi. Model refinement: Based on the results of the evaluation, the model may need to be refined by adjusting the hyperparameters, selecting different features, or selecting a different model.

vii. Model deployment: Once the model is refined, it can be deployed in a production environment. This involves integrating the model into a prediction system. .

x. Model monitoring and maintenance: The final step involves monitoring the performanceof the deployed model and making necessary adjustments to maintain its accuracy andeffectiveness.

CHAPTER 4

RESULTS ANALYSIS AND VALIDATION

4.1. Implementation of solution

Model Used :

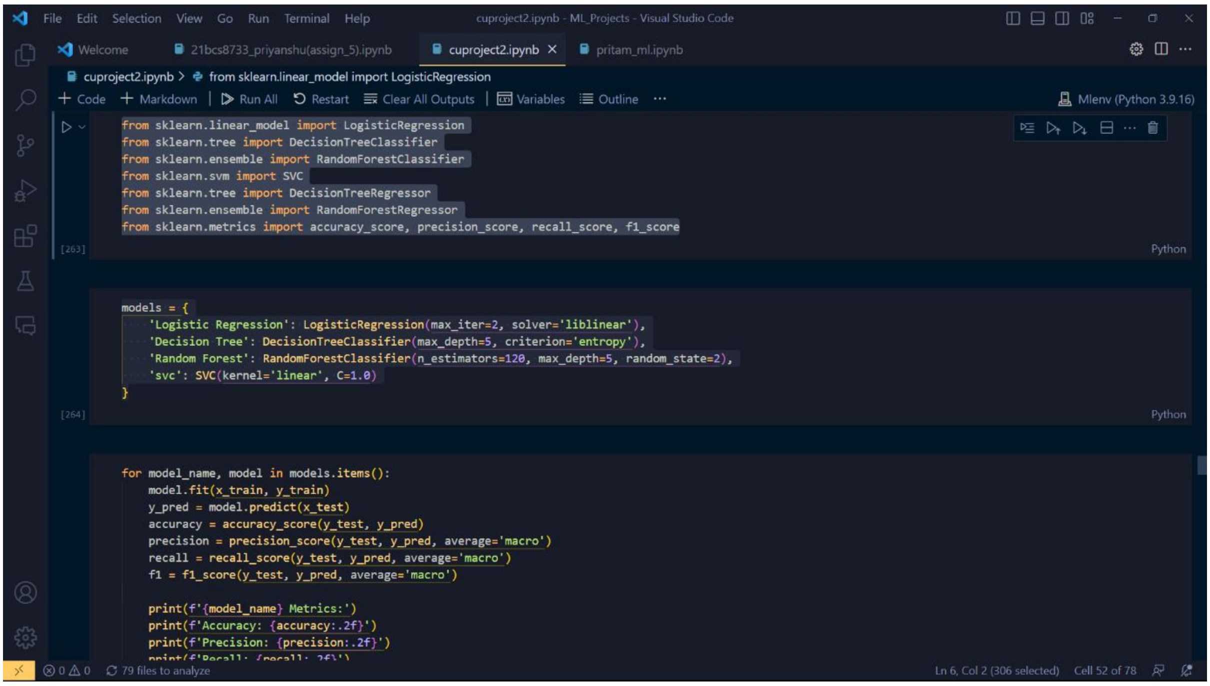

Figure 9: 4 models on vscode

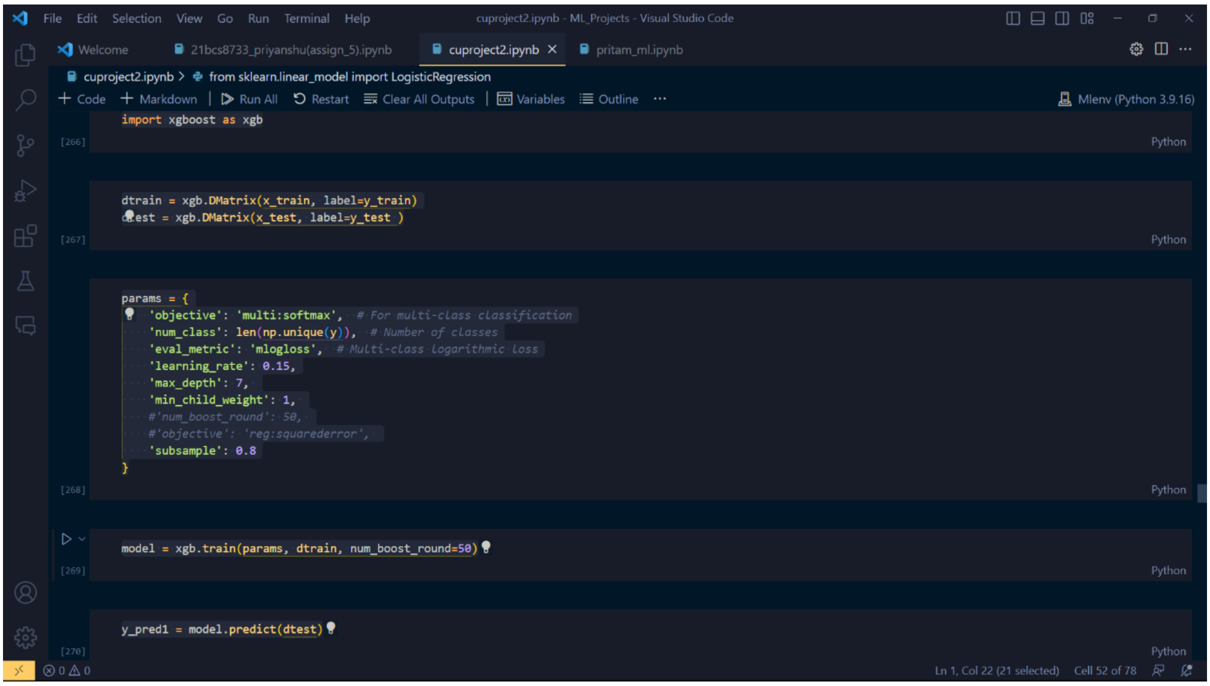

Figure 9: 5th model xgboost

1. Logistic Regression
2. DecisionTreeClassifier
3. RandonForestClassifier
4. SVC
5. Xgboost

Precision and Accuracy Records

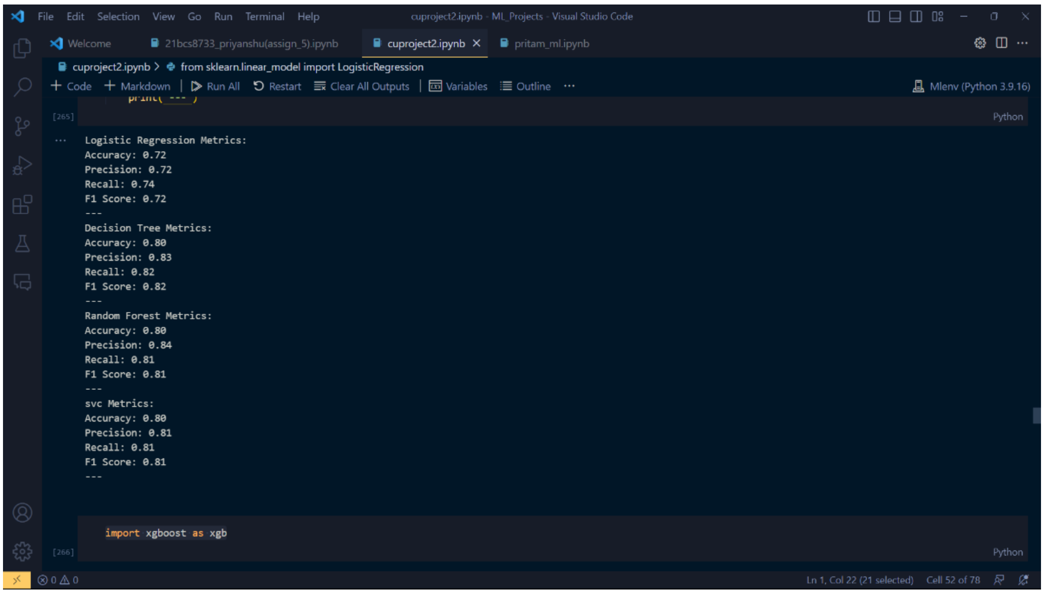

Figure 10: precision and accuracy of the 4 models

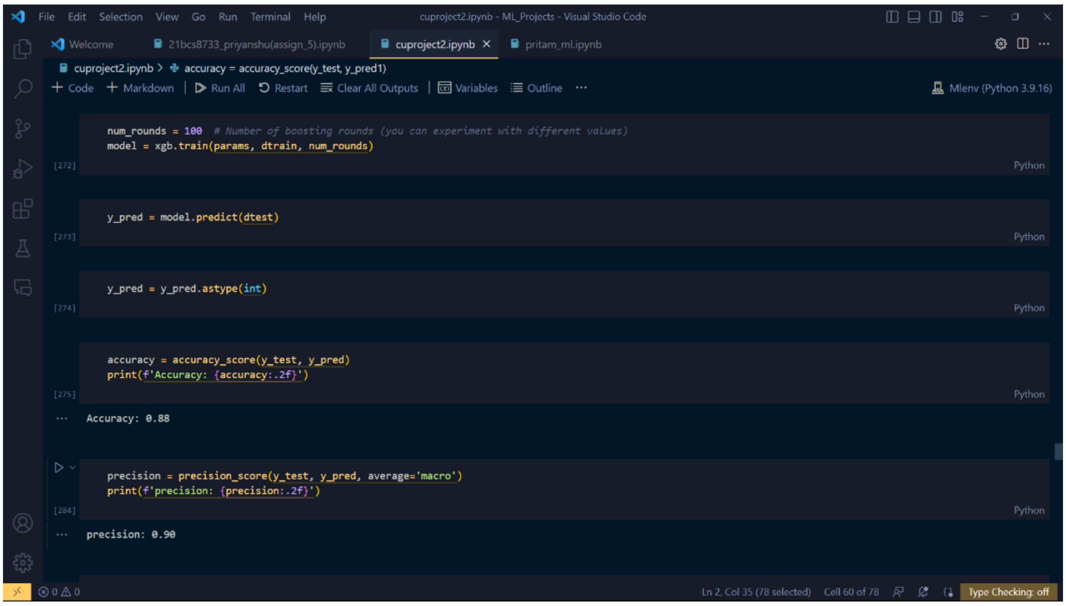

Figure 11: precision and Accuracy of the Xgboost model

The models and their accuracy :

```
Logistic Regression Metrics:
Accuracy: 0.72
Precision: 0.72
Recall: 0.74
F1 Score: 0.72
---
Decision Tree Metrics:
Accuracy: 0.80
Precision: 0.83
Recall: 0.82
F1 Score: 0.82
---
Random Forest Metrics:
Accuracy: 0.80
Precision: 0.84
Recall: 0.81
F1 Score: 0.81
---
svc Metrics:
Accuracy: 0.80
Precision: 0.81
Recall: 0.81
F1 Score: 0.81
```

Where as Accuracy and prediction of the Xgboost is:

```
Accuracy: 0.88
precision: 0.90
```

hence the model which can predict the changes in weather more accurately would be the **Xgboost model** as compare to the other model used in the Aiml processes.

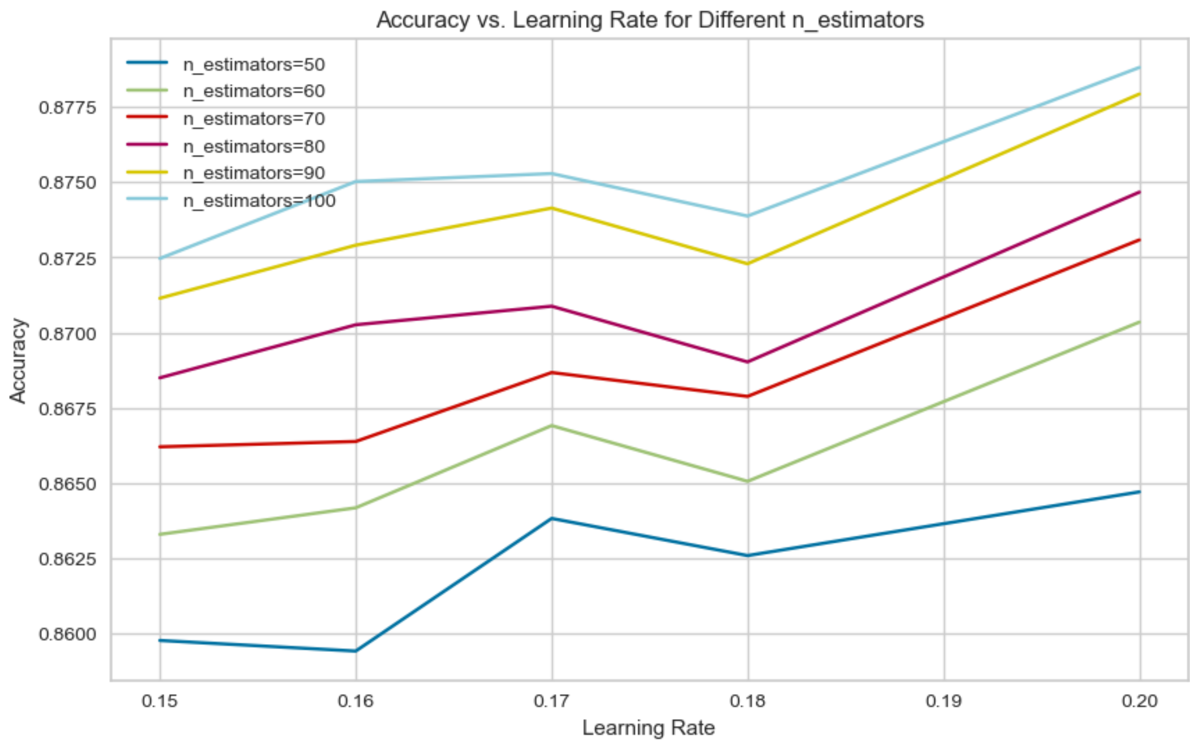

Figure 12: Accuracy vs learning Rates of different n_estimators

DATA VISUALIZATION

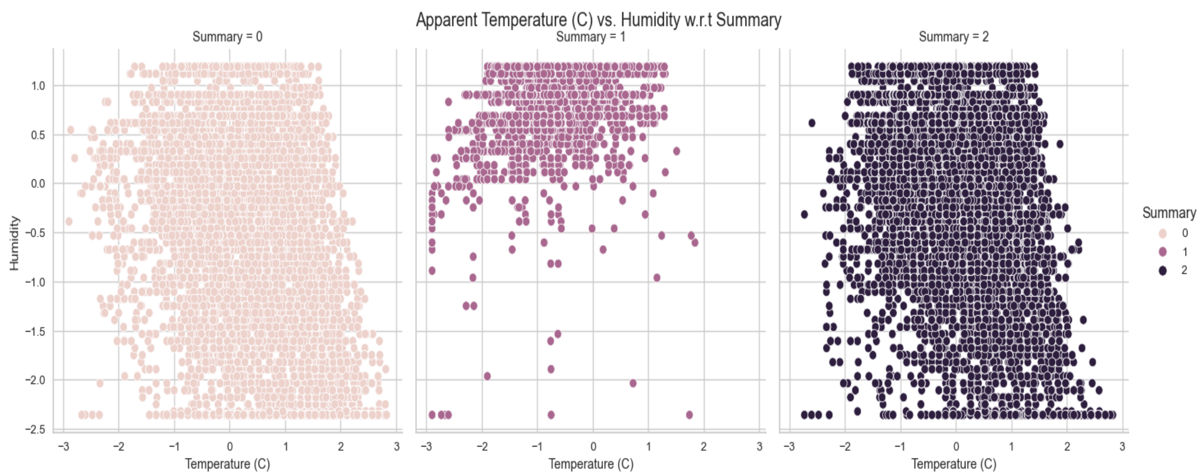

Figure 13. Apperant Temp vs Humidity

CHAPTER 5
CONCLUSION AND VALIDATION

5.1. Conclusion

The study on weather analysis using different models has provided valuable insights into the strengths and limitations of various weather modeling approaches. Through a comprehensive analysis of numerical weather prediction (NWP), statistical models, and machine learning-based methods, we have gained a deeper understanding of their performance in capturing atmospheric phenomena and forecasting weather patterns.

The research revealed that NWP models excel in capturing large-scale weather patterns and dynamic processes, making them valuable tools for medium to long-range weather forecasts. However, their performance may be impacted by uncertainties in initial conditions and model parameterizations, highlighting the importance of data assimilation and model calibration.

Machine learning-based methods showcased their ability to capture intricate nonlinear relationships within weather data, resulting in improved short-term weather forecasting. These models show promise in handling large datasets and adapting to complex weather patterns. However, their application may require careful data preprocessing and selection of appropriate algorithms.

Ensemble approaches, combining multiple models, emerged as a promising strategy to address uncertainties and enhance overall forecasting accuracy. By leveraging the complementary strengths of different models, ensemble techniques offer a more reliable range of predictions, especially for extreme events.

It is evident that weather modeling remains a dynamic and evolving field, driven by advancements in technology, data availability, and climate research. The challenges posed by climate change and the need for accurate weather forecasts call for continuous research and improvements in modeling techniques.

In conclusion, the findings from this study underscore the importance of a diverse set of weather analysis models and the potential for integrating different methodologies. By leveraging the strengths of each approach and addressing their limitations, we can advance weather modeling capabilities and make significant strides in improving weather analysis and forecasting accuracy. This, in turn, will support informed decision-making in various sectors, ensuring better preparedness for weather-related challenges and contributing to overall climate resilience.

5.2. Future Work

Building upon the findings and insights from this study on weather analysis using different models, several areas of future work emerge to further advance the field and address the challenges in weather modeling:

1. Model Development and Calibration: Future research could focus on refining existing weather analysis models and developing new ones to enhance accuracy and robustness. Improving model calibration techniques and data assimilation methods could help reduce uncertainties and improve forecasts.

2. Integration of Advanced AI Techniques: With the rapid advancements in AI and machine learning, integrating cutting-edge deep learning algorithms into weather analysis models holds great potential. Investigating the application of neural networks, recurrent networks, and transformers can further improve short-term and long-term forecasting accuracy.

3. Big Data and High-Resolution Modeling: As computational capabilities continue to improve, exploring the integration of big data and high-resolution modeling can provide more detailed and precise predictions. Leveraging remote sensing and Earth observation data can further enhance model inputs and outputs.

4. Climate Change Impact Studies: Given the growing concern over climate change and its impact on weather patterns, future research should focus on conducting in-depth studies to understand the linkages between climate change and weather modeling. Investigating how climate change influences extreme weather events and long-term climate trends is critical for informed decision-making and climate adaptation strategies.

5. Ensemble Techniques: Developing more sophisticated ensemble techniques and exploring the optimal combination of different models could lead to further improvements in forecasting accuracy and uncertainty quantification.

6. Real-Time and High-Frequency Forecasting: Enhancing real-time weather forecasting capabilities, especially for short-term and high-frequency predictions, can support various sectors, such as transportation, energy, and emergency response, where timely and accurate information is crucial.

By addressing these future research areas, the field of weather modeling can progress and better serve society by providing more accurate and reliable weather forecasts, supporting climate studies, and helping communities

5.3. Future Scope

The future scope for weather analysis using different models is vast, with numerous opportunities for advancements and applications. Some of the key future scopes in this field include:

7. Improved Forecasting Accuracy: Advancements in modeling techniques, data assimilation, and high-performance computing will continue to enhance the accuracy and reliability of weather forecasts. Fine-tuning models to better capture complex atmospheric processes and leveraging AI algorithms can lead to more precise predictions.

8. Extreme Event Prediction: Addressing the challenges posed by extreme weather events, such as hurricanes, heatwaves, and floods, will be a significant future focus. Developing specialized models to better predict and prepare for extreme events can help minimize their impacts on communities and infrastructure.

9. Climate Adaptation and Mitigation: Weather analysis models will play a crucial role in supporting climate adaptation and mitigation strategies. By accurately projecting long-term climate trends, decision-makers can implement measures to mitigate climate change effects and foster resilient communities.

10. Real-Time Decision Support Systems: The development of real-time weather analysis systems will continue to grow, providing timely and accurate information to support emergency response, disaster management, and other critical decision-making processes.

11. AI-Driven Innovations: The integration of AI and machine learning in weather modeling will evolve, allowing for more sophisticated pattern recognition, data processing, and model optimization. AI-driven innovations will lead to novel approaches in understanding weather dynamics.

12. Earth System Modeling: Expanding weather analysis models to encompass Earth system modeling will enable a holistic view of interactions between the atmosphere, oceans, land, and cryosphere. This integrated approach will contribute to a comprehensive understanding of climate dynamics.

REFERENCES

1. Agrawal, R., Jain, R.C., Jha, M.P. and Singh, D. (1980): Forecasting of rice yield using climatic variables. Indian Journal of Agricultural Science, Vol. 50, No. 9, pp. 680-684.
2. Lee, S., Cho, S.& Wong, P.M., (1999) : Rainfall prediction using artificial neural network.― J. Geog. Inf. Decision Anal. 2, 233–242 1998. [10] Wong, K. W., Wong, P. M., Gedeon, T. D. & Fung, C. C. ―Rainfall Prediction Using Neural Fuzzy Technique.
3. C. Hamzacebi, "Improving artificial neural networks' performance in seasonal time Series Forecasting", Information Sciences 178 (2008), pages: 4550-4559.
4. Lin, Long-Ji. (1993): Scaling up reinforcement learning for robot control. Proceedings of the tenth International Conference on Machine Learning.
5. G.E.P. Box, G. Jenkins (1970), "Time Series Analysis, Forecasting and Control", Holden-Day, SanFrancisco,CA
6. Chatfield, C. (1994): The Analysis of Time Series-An Introduction. Chapman and Hall.
7. Sivakumar, B. (2001): Rainfall dynamics in different temporal scale: A Chaotic percepective. Hydrology and Earth System Science, Vol.5, pp. 645-651.
8. Guhathakurta, P. (2006). Long range of monsoon rainfall prediction of 2005 for the districts and sub-division in kerala with artificial neural network. Current science, Vol. 90, pp. 773-779.
9. Saima, H., Jaafar, J., Belhaouari, S. and Jillani, T.A. (2011): ARIMA based Interval Type-2 Fuzzy Model for Forecasting. International Journal of Computer Applications, Vol. 28, No. 3, pp. 17-21.
10. M.Tektas, "Weather Forecasting Using ANFIS and ARIMA (2010): A Case study of Istanbul," Enviornment Research , Engineering and Management , vol. 1(51), pp.5-10.
11. Mahmudur Rahman, A.H.M. Saiful Islam , Sahah Yaser Maqnoon Nadvi , Rashedur M Rahman (2013) : Comparative Study of ANFIS and ARIMA Model for weather forecasting in Dhaka" IEEE.
12. Sarah N. kohail, Alaa M. EI-Halees(2011): "Implementation of Data Minning Technique for Metrological Data Analysis",IJICT Journal Volume 1 No.3, july .

13. S.A. Shamsnia, N. Shahidi, A. Liaghat, A.Sarraf and S.F. Vahdat. 2011. Modeling Of Weather Parameters Using Stochastic Methods. Internat. Conference on Environment and Industrial Innovation, IPCBEE, Singapore, 282-285.

14. R Development Core Team (2008). R: A language and environment for analytic computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.Rproject. org.

15. Montgomery, D.C., and Lynwood A.J. (1996): Forecasting and Time Series Analysis, Mc. Graw-Hill.