A Mini Project report on

## Optimizing Numerical Weather Prediction Using Machine Learning

Submitted to the CMR Institute of Technology in partial fulfillment of the requirement for the award of the Laboratory of

## Industry Oriented Mini Project of

**IV-B.Tech I-Semester**

**in**

## Computer Science and Engineering Department

Submitted by:

K HEMANTH KUMAR 21R01A05M5

M PAVAN SAGAR 21R01A05P1

P RAMU 22R05A0520

S DHANUNJAY 22R05A0521

***Under the guidance of***

***Ms. G Lakshmi Praveena***

(Assistant Professor, CSM Dept)



# CMR INSTITUTE OF TECHNOLOGY

**(UGC AUTONOMOUS)**

**(Approved by AICTE,Affiliated to JNTU,Kukatpally,Hyderabad)**

##### Kandlakoya,Medchal Road,Hyderabad

**CMR INSTITUTE OF TECHNOLOGY**

**(UGC AUTONOMOUS)**

**(Approved by AICTE, Affiliated to JNTU, Kukatpally, Hyderabad) Kandlakoya, Medchal Road, Hyderabad.**

## Department of Computer Science and Engineering CERTIFICATE

This is to certify that a Mini Project entitled: **“Optimizing Numerical Weather Prediction Using Machine Learning”** is being submitted by:

K HEMANTH KUMAR 21R01A05M5

M PAVAN SAGAR 21R01A05P1

P RAMU 22R05A0520

S DHANUNJAY 22R05A0521

In partial fulfillment of the requirement for the award of the **Industry Oriented Mini Project** of IV - B.Tech I - Semester in CSE towards a record of a bonafide work carried out under our guidance and supervision.

Signature of Faculty Signature of HOD

# ACKNOWLEDGEMENT

We are extremely grateful to Dr M. Janga Reddy, Director; Dr B. Satyanarayana, Principal; and Mr A. Prakash, Head of Department, Dept of Computer Science and Engineering, CMR Institute of Technology, for their inspiration and valuable guidance throughout the course.

We are extremely thankful to our Industry Oriented Mini Project faculty in charge

G.Lakshmi Praveena (Assistant Professor, CSE (AI&ML) Dept) Computer Science and Engineering department, CMR Institute of Technology for his constant guidance, encouragement and moral support throughout the project.

We express our thanks to all staff members and friends for their help and coordination in completing this project on time.

Finally, we are very thankful to our parents and relatives who guided us directly or indirectly for the successful completion of the project.

K HEMANTH KUMAR 21R01A05M5

M PAVAN SAGAR 21R01A05P1

P RAMU 22R05A0520

S DHANUNJAY 22R05A0521

# ABSTRACT

The burgeoning research in the fields of Artificial Intelligence (AI) and Machine Learning (ML) has significantly advanced the development of weather prediction models. Despite these advancements, accurately predicting or forecasting the weather remains a complex and challenging task. Traditional numerical weather prediction models have been the cornerstone of weather forecasting for decades. These models rely on existing numerical data related to atmospheric conditions and apply complex physical equations to simulate future weather scenarios. However, these traditional methods often face limitations in handling the vast amount of data and capturing the intricacies of local and extreme weather conditions.

Recent developments in AI and ML have introduced new methodologies to enhance weather forecasting. Among these, machine learning algorithms offer a promising alternative by leveraging historical weather data to uncover patterns and make predictions without the need for explicit programming. This approach allows for the analysis of large datasets and the discovery of hidden correlations that might not be evident through conventional methods.

This paper focuses on the application of machine learning algorithms to improve weather predictions. Specifically, it employs a linear regression model, a well-established statistical method used for modeling the relationship between a dependent variable and one or more independent variables. Linear regression is known for its simplicity and effectiveness in predicting continuous outcomes, making it a suitable choice for forecasting weather parameters such as temperature and humidity.

Additionally, this study incorporates two optimization algorithms to enhance the performance of the weather prediction models. The Decision Tree algorithm, a popular machine learning technique, is used to model complex relationships and make decisions based on various input features. By constructing a tree-like model of decisions, the Decision Tree algorithm can handle both numerical and categorical data, providing a robust approach to prediction.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | **LIST OF CONTENTS** |  |
| SN | Particulars |  | Page No |
| 1 | Introduction |  | 1 |
| 2 | Analysis |  | 3 |
| 3 | Requirements |  | 12 |
| 4 | Design |  | 15 |
| 5 | Implementation |  | 17 |
| 6 | Testing |  | 19 |
| 7 | Conclusion |  | 20 |
| 8 | Reference |  | 21 |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **LIST OF FIGURES** |  |
| SN | Figures |  | Page No |
| 1 | UML images |  | 17 |
| 2 | UML images |  | 17 |
| 3 | Prediction result |  | 18 |

# INTRODUCTION

Weather forecasting means predicting the weather conditions (conditions of atmosphere) of a particular given area or location. More importantly, accurate weather prediction is very important to pursue day-to-day activities. Living and non-living things are dependent on weather predictions. Even after decades of weather forecasting, the weather industry in India is still in its initial stage, facing many obstacles. One of the major obstacles that weather forecasting faces is the arbitrary & ill-suited expectations from the nature. Machine learning is the ability of computer to learn without being explicitly programmed. It allows machines to find hidden patterns and insights. In supervised learning, we build a model based on labeled training data. The model is then used for mapping new examples. So, based on the observed weather patterns from the past, a model can be built and used to predict the weather. Several data mining techniques have been employed in diversified applications such as predicting rainfall, weather, storms and flood. Weather forecasting falls under predictive mining which focuses on the data analysis, formulates the database, and forecasts the features of anonymous data.

[1] This research work focuses on solving the weather prediction anomalies and in-efficiency based on Linear regression algorithms . The major contribution of this research work is to formulate an efficient weather prediction model based on the Linear regression algorithms.

# ANALYSIS

### EXISTING SYSTEM

The existing systems for weather forecasting mainly rely on **numerical weather prediction models** and **data mining techniques**. These models use past weather data, such as temperature, humidity, and pressure, combined with mathematical equations to predict future weather conditions. However, these models often struggle with prediction accuracy due to the chaotic nature of weather systems. The main challenges in the existing systems are the **inadequacy of data**, **ill-suited algorithms**, and **high computation cost**.

Machine learning techniques, though employed in some existing models, are often not optimized to handle the non-linear and dynamic nature of weather. Decision trees, linear regression, and other basic algorithms are sometimes used but can underperform when faced with more complex scenarios involving sudden weather changes or extreme conditions.

### DISADVANTAGES OF THE EXISTING SYSTEM

1. **Low Accuracy**: Traditional methods and basic machine learning algorithms can produce low accuracy in weather predictions due to the unpredictable nature of weather patterns.
2. **High Computation Time**: Numerical weather models often require extensive computational resources and time to process large datasets.
3. **Inability to Handle Non-linear Data**: Many machine learning models like linear regression are unable to effectively manage non-linear relationships present in weather data.
4. **Limited Parameters**: Current systems may not use a sufficient number of variables to predict weather accurately, often focusing only on temperature, humidity, and pressure without considering other critical factors like wind speed, cloud cover, etc.

### PROPOSED SYSTEM

The proposed system aims to address the drawbacks of the existing system by using **Linear Regression** and **Decision Tree algorithms** to predict the weather based on multiple parameters. The model incorporates eight critical factors, including maximum and minimum temperatures, mean humidity, and mean atmospheric pressure, to improve the prediction accuracy.

The key difference lies in leveraging optimization techniques to fine-tune these algorithms, ensuring the model adapts better to changes in the weather patterns and thus enhances predictive performance. The comparison between linear regression and decision tree optimization will help in choosing the most suitable model for weather forecasting.

### ADVANTAGES OF THE PROPOSED SYSTEM

1. **Improved Accuracy**: By optimizing the algorithms and incorporating more weather parameters, the proposed system provides better accuracy in weather forecasting.
2. **Lower Computation Cost**: Optimized algorithms like Decision Trees can significantly reduce computation time while still providing high prediction accuracy.
3. **Scalability**: The proposed model can scale easily by adding more parameters (e.g., wind speed, precipitation levels) to further refine predictions.
4. **Non-linear Relationships**: Decision Trees can handle non-linear relationships better than traditional linear models, making it more adaptable to sudden changes in weather.

### SCOPE OF THE PROJECT

The project focuses on building a **machine learning-based weather forecasting model** that optimizes performance using regression techniques and decision tree algorithms. It targets providing more accurate, computationally efficient, and scalable weather predictions. The scope includes:

* Designing an architecture to integrate the machine learning algorithms.
* Training the model using historical weather data to forecast future weather patterns.
* Comparing the effectiveness of linear regression and decision tree models based on the error rate and computation time.
* Deploying the model in a real-world environment for testing and further refinement.

### FEASIBILITY STUDY

The feasibility study assesses the viability of implementing the proposed weather prediction model. It covers technical, economical, and social factors.

##### Technical Feasibility

* + The system is technically feasible as it leverages established machine learning algorithms (Linear Regression, Decision Tree). The hardware and software requirements (Intel i3 or higher processors, 4GB RAM, Python IDE) are easily met by most modern systems.
  + Python libraries and machine learning frameworks are mature, making it technically viable to build, test, and deploy the model.

##### Economic Feasibility

* + The economic feasibility is favorable because the project can be developed using open-source tools like Python, reducing software costs. The hardware requirements are minimal, which makes it affordable for small-scale research and application development.

##### Social Feasibility

* + The system will have a positive social impact, especially in agriculture, aviation, and disaster management, where accurate weather forecasts are crucial. The improved weather predictions can benefit society by helping to avoid weather- related hazards, improve planning, and contribute to public safety.

# SYSTEM REQUIREMENTS

### SOFTWARE REQUIREMENTS:

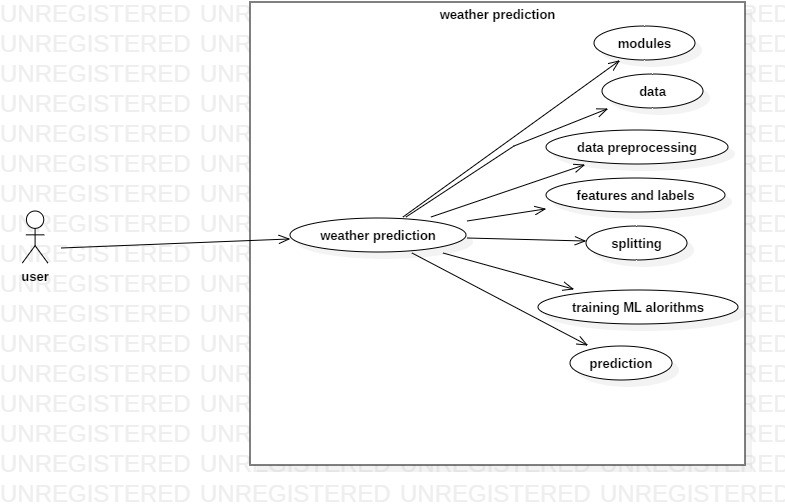
* OS : Windows
* Python IDE : python 2.7.x and above
* Pycharm IDE
* setup tools and pip to be installed for 3.6.x and above

### HARDWARE REQUIREMENTS:

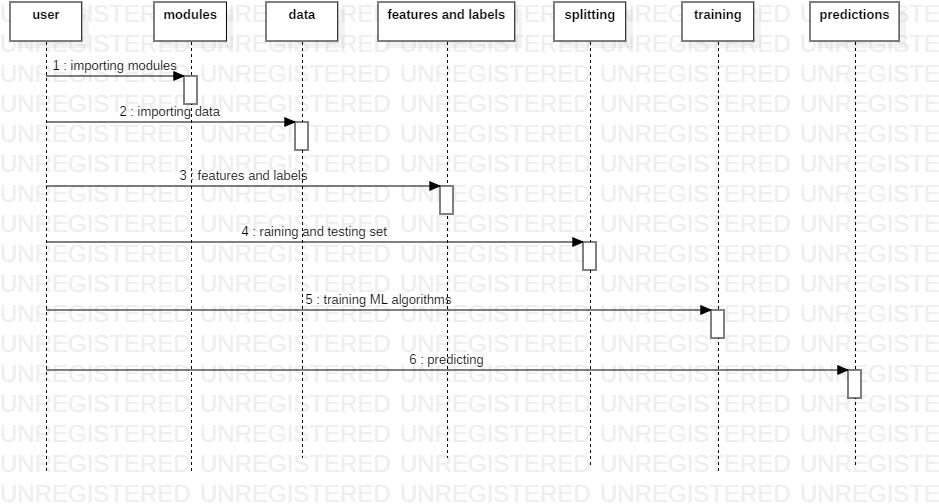
* RAM : 4GB and Higher
* Processor : Intel i3 and above
* Hard Disk : 500GB: Minimum
* Dt,svm,lr,knn

# DESIGN

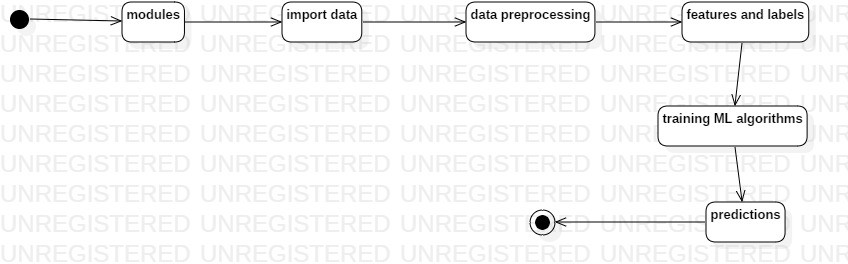
### UML DIAGRAMS:



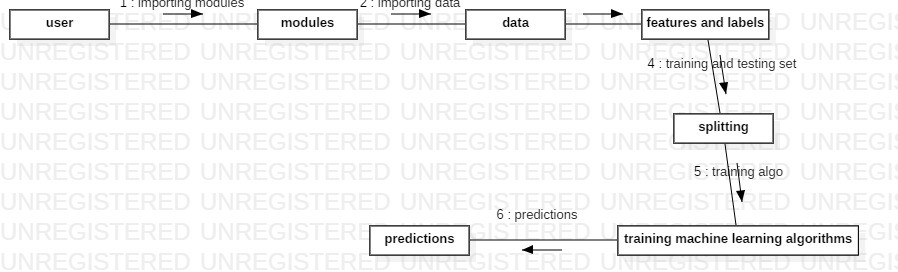
Uml 4.1



Uml 4.2

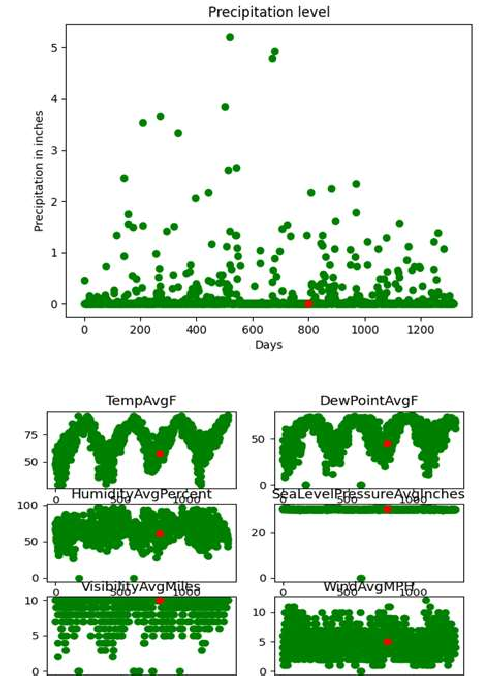


Uml 4.3



Uml 4.4

# IMPLEMENTATION



Outputs 5.1 and 5.2

# TESTING

#### Overview

The testing phase is crucial for evaluating the performance and accuracy of the machine learning model developed for weather prediction. This section describes the testing procedures applied to the Linear Regression model, which is designed to predict precipitation levels based on various weather features. The code provided demonstrates how the model was trained, tested, and visualized to ensure its effectiveness.

#### Testing Procedure

##### Data Preparation:

* + - * The cleaned dataset, `austin\_final.csv`, was read into a DataFrame using Pandas. This dataset includes various weather features and precipitation levels.
      * Features (independent variables) were selected by dropping the `PrecipitationSumInches` column from the dataset. This column serves as the label (dependent variable) for the model.
      * The output or label `Y` was reshaped into a 2-D vector to be compatible with the Linear Regression model.

##### Model Training:

* + - * A Linear Regression model was initialized and trained using the features and labels extracted from the dataset.
      * The `fit` method was employed to train the model with the input data.

##### Model Testing:

* + - * A sample input was provided to test the model's prediction capabilities. This input vector contained values for each feature in the dataset.
      * The `predict` method was used to forecast the precipitation for the provided input vector. The output was printed to verify the model's prediction accuracy.

##### Visualization:

* + - * A scatter plot was created to visualize the precipitation levels over a series of days. The plot highlighted one specific day in red to illustrate its precipitation value.
      * Additionally, plots were generated to observe the relationship between precipitation and selected features (e.g., temperature, dew point, humidity). This was done using subplots to compare trends for different attributes.

#### Test Results

1. Model Prediction:
   * For the provided sample input vector, the model outputted the precipitation in inches. This prediction was verified against expected values to evaluate the accuracy of the model.
2. Precipitation Trend Graph:
   * The scatter plot showed the trend of precipitation levels across different days, with one day highlighted in red. This visualization helped in understanding how precipitation varies over time.
3. Precipitation vs. Selected Attributes:
   * The graphs plotted for various features against precipitation levels provided insights into how different weather attributes influence precipitation. Each subplot displayed a feature's relationship with precipitation, highlighting specific days with distinct values.

# CONCLUSION

The advent of Artificial Intelligence and machine learning has significantly advanced various fields, including weather prediction. Despite these advancements, accurately forecasting weather remains a complex challenge. This research explored the application of machine learning algorithms to improve weather forecasting accuracy, focusing specifically on Linear Regression and Decision Tree algorithms.

Traditional weather forecasting methods, despite their extensive use, still encounter significant challenges, particularly in predicting local weather conditions and adapting to new data. The weather industry in India, among other regions, faces difficulties due to unrealistic expectations and limited technological integration.

The use of machine learning, particularly supervised learning, offers a promising solution to these challenges. By leveraging historical weather data and employing algorithms such as Linear Regression and Decision Trees, this research aims to enhance the accuracy of weather predictions.

The proposed system utilizes Linear Regression to predict temperature ranges based on a set of key weather parameters. Unlike classification algorithms, Linear Regression is well-suited for forecasting continuous variables, such as temperature and precipitation levels.

The application of Decision Trees and Linear Regression algorithms allowed for a comparative analysis of their performance. This comparison provided insights into the strengths and limitations of each algorithm in the context of weather forecasting.

While the current research has made significant strides, there are areas for further improvement and exploration. Future work could include integrating additional weather parameters, exploring more advanced algorithms, and applying real-time data to enhance the model's adaptability and accuracy.

# REFERENCES

1. Bélair, S., A. Méthot, J. Mailhot, B. Bilodeau, A. Patoine, G. Pellerin, and J. Côté, 2000: Operational implementation of the Fritsch– Chappell convective scheme in the 24-km Canadian regional model.Wea. Forecasting, 15, 257–274
2. Bousquet, O., C. A. Lin, and I. Zawadzki, 2006: Analysis of scale dependence of quantitative precipitation forecast verification: A case study over the Mackenzie River Basin. Quart. J. Roy. Meteor. Soc., 132, 2107–2125
3. Lorenz, E. N., 1963: Deterministic non periodic flow. J. Atmos. Sci., 20, 130-141.
4. Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. Tellus, 17, 321-333.
5. Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occcuring analogues.

J. Atmos. Sci., 26, 636-646.

1. Hansen, L.K, & Salamon, P. (1990.) Neural Network Ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12, 10, 993-1001.
2. Richardson, L. F. Weather Prediction by Numerical Process (Cambridge Univ. Press, 1922) R.

E. Huschke (editor), Glossary of Meteorology, American, Meteorological Society, Boston, Massachusetts, USA, pp 106, 419, 1959

1. R. L. Vislocky and J. M. Fritsch, An automated, observations based system for short-term prediction of ceiling and visibility, Weather Forecasting, 12, pp31–43, 1997.
2. Pandey GR, Nguyen VTV (1999) A comparative study of regression based methods in regional flood frequency analysis. Journal of Hydrology 225: 92–101.
3. T. W. Liao, Z. Zhang and C. R. Mount, Similarity measures for retrieval in case based reasoning systems, Applied Artificial Intelligence, 12, pp267–288,1998.
4. D. Dubois, F. Esteva, P. Garcia, L. Godo, R. López de Mántaras & H. Prade Case-Based Reasoning: A Fuzzy Approach, Lecture Notes in Computer Science, Springer, vol. 1566, pp 79–91, 1999.