Manjot Singh, Ranbeer Singh, Karanvir Singh
Presenter

Machine Learning in Weather

Utilising Machine Learning for Accurate Weather Predictions

Certificate of Project Completion

01

Certification Title: Certificate

This certificate acknowledges the completion of the project titled 'A Machine Learning Model for Weather Forecasting'.

02

Project Authors: Ranbeer Singh Boparai, Karanveer Singh, Manjot Singh

The project was carried out by three dedicated members who contributed significantly to its success.

03

Supervision: Dr. Ashima Khosla

The project was supervised by Dr. Ashima Khosla, who provided guidance and support throughout the process.

04

Project Focus: Weather Forecasting

This project aimed to develop a machine learning model specifically for predicting weather patterns and conditions.

Acknowledgement of Support

Gratitude to Dr. Ashima Khosla

We extend our heartfelt appreciation to our project guide, **Dr. Ashima Khosla**, for her invaluable support during our project on weather forecasting.

Invaluable Project Opportunity

We are thankful for the opportunity to work on the project titled *“A Machine Learning Model for Weather Forecasting.”*

Deep Exploration in Subject

Dr. Khosla's guidance allowed us to explore the subject matter deeply, enhancing our understanding of machine learning applications.

Appreciation for Family Support

Our sincere gratitude goes to our parents and friends whose encouragement was crucial for completing this project on time.

Team Collaboration

The successful completion of this project was made possible through effective teamwork and collaboration among the authors.

Completion Date

This project was completed on **12-07-2021**, marking an important milestone for our team.

Weather Forecasting Contents

01 **TITLE**
The main focus of the presentation is identified.

02 **ABSTRACT**
Summarises the key aspects of the

03 **BACKGROUND**
Provides context and relevance of the

04 **OBJECTIVE**
States the main goals of the research.

05 **1.0 INTRODUCTION**
Introduces the topic and its significance.

06 **2.0 METHODOLOGY**
Details the methods used for the research.

07 **3.0 EXPERIMENTATION**
Describes the experimental

08 **4.0 RESULT AND DISCUSSION**
Presents findings and insights from

09 **5.0 CONCLUSION**
Summarises the findings and implications.

ABSTRACT

Traditionally, climate assessment has been performed reliably by treating the environment as a liquid. The current wind condition is being observed. The future state of the environment is recorded by understanding thermodynamics and the numerical position of the liquid elements. Nevertheless, this traditional arrangement of differential conditions as observed by physical models is at times unstable under oscillating effects and uncertainties when estimating the underlying states of air. This indicates an insufficient understanding of environmental variations, so it limits climate forecasts to 10-day periods because climate projections are essentially unreliable. But machine learning is moderately hearty for most barometric destabilizing effects compared to traditional techniques. Another favorable position of machine learning is that it does not depend on the physical laws of environmental processes.

BACKGROUND

For the current situation, India observatory conducts traditional weather forecasting. There are four common methods to predict the weather. The first method is the climatology method that is reviewing weather statistics gathered over multiple years and calculating the averages. The second method is an analog method that is to find a day in the past with weather similar to the current forecast. The third method is the persistence and trends method that has no skill to predict the weather because it relies on past trends. The fourth method is numerical weather prediction the is making weather predictions based on multiple conditions in the atmosphere such as temperatures, wind speed, high-and low- pressure systems, rainfall, snowfall, and other conditions. So, there are many limitations of these traditional methods. Not only it forecasts the temperature in the current month at most, but also it predicts without using machine learning algorithms. Therefore, our project is to increase the accuracy and predict the weather in the future for at least one month by applying machine learning techniques

OBJECTIVE

Purpose of this project is to predict the temperature using different algorithms like linear regression, random forest regression, and Decision tree regression. The output value should be numerically based on multiple extra factors like maximum temperature, minimum temperature, cloud cover, humidity, and sun hours in a day, precipitation, pressure and wind speed.

INTRODUCTION

Weather prediction is the task of predicting the atmosphere at a future time and a given area. This has been done through physical equations in the early days in which the atmosphere is considered fluid. The current state of the environment is inspected, and the future state is predicted by solving those equations numerically, but we cannot determine very accurate weather for more than 10 days and this can be improved with the help of science and technology.

Machine learning can be used to process immediate comparisons between historical weather forecasts and observations. With the use of machine learning, weather models can better account for prediction inaccuracies, such as overestimated rainfall, and produce more accurate predictions. Temperature prediction is of major importance in a large number of applications, including climate-related studies, energy, agricultural, medical, or etc.

There are numerous kinds of machine learning calculations, which are Linear Regression, Polynomial Regression, Random Forest Regression, Artificial Neural Network, and Recurrent Neural Network. These models are prepared dependent on the authentic information gave of any area. Contribution to these models is given, for example, if anticipating temperature, least temperature, mean air weight, greatest temperature, mean dampness, and order for 2 days. In light of this Minimum Temperature and Maximum Temperature of 7 days will be accomplished.

Machine Learning in Weather

01

Initially based on physical equations considering the atmosphere as fluid, weather prediction has advanced with technology.

Weather prediction methods have evolved

Machine learning improves accuracy

02

Machine learning processes historical weather data to enhance the accuracy of weather forecasts significantly.

03

Accurate temperature forecasts are crucial for various sectors including agriculture, energy, and healthcare.

Temperature prediction is vital

Various machine learning models exist

04

Models such as Linear Regression, Random Forest, and Neural Networks are employed based on historical weather data parameters.

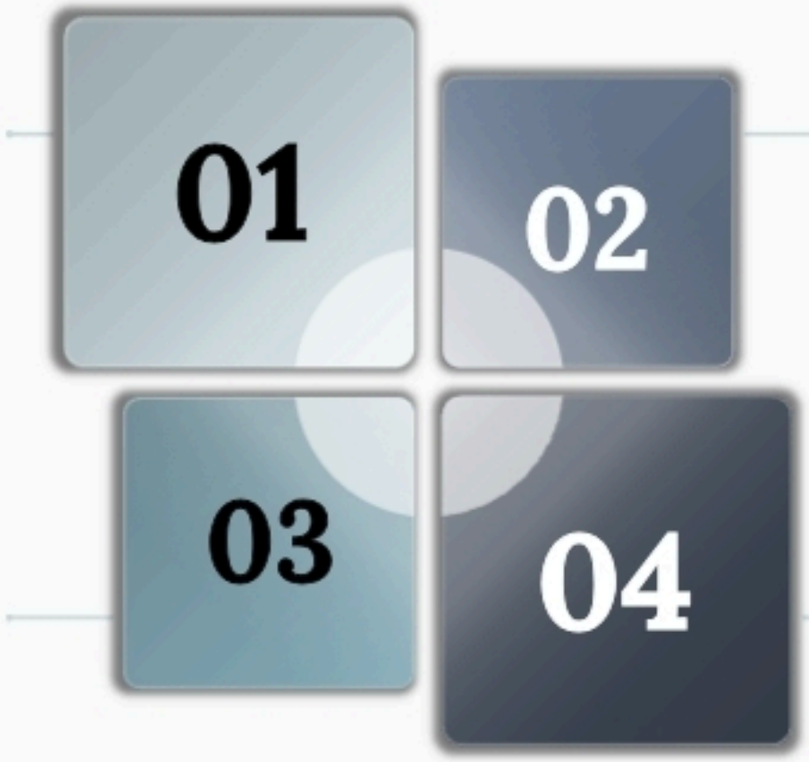
Algorithms for Temperature Forecasting

Data Division for Training

In temperature forecasting, **80%** of the dataset is used for training while **20%** is reserved for testing, ensuring robust model evaluation.

Machine Learning Models

Various **machine learning** models, such as **Linear Regression**, **Decision Tree Regression**, and **Random Forest Regression**, significantly enhance forecasting accuracy compared to traditional methods.



Historical Data Usage

For accurate predictions, **8 years** of historical temperature data is leveraged to train models, while **2 years** are used to validate performance.

Comparative Precision

Machine learning approaches offer a paradigm shift in weather forecasting, providing higher **precision** and **predictivity** than classical physics-based methods.

Methodology for Weather Prediction

Dataset Source

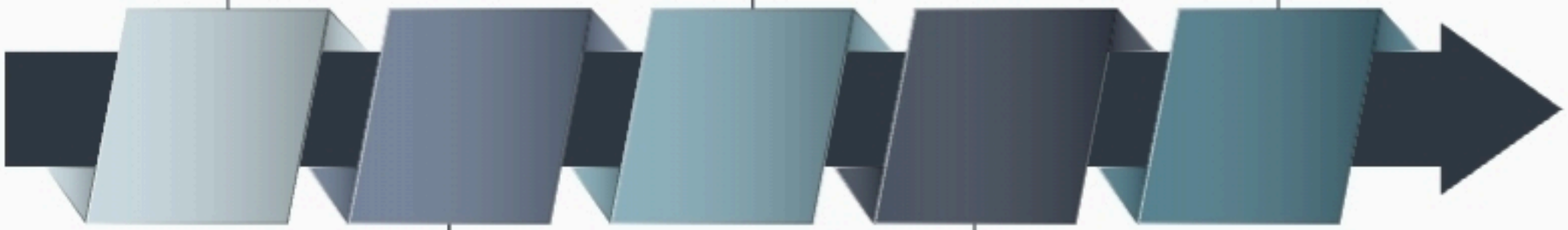
The dataset is sourced from Kaggle, specifically for 'Kanpur City'.

Primary Usage

The dataset aids in predicting future weather and visualising global warming effects.

Data Accuracy Note

Data accuracy cannot be guaranteed as it was extracted using an external API.



Historical Data Range

Data spans from 01-01-2009 to 01-01-2020, providing a comprehensive view.

Machine Learning Focus

Concentration on temperature prediction using algorithms like Linear Regression and Random Forest.

Table 2.1: Historical Weather Dataset of Kanpur City

date_time	maxtempC	mintempC	totalSnow_cm	sunHour	uvIndex	uvIndex.1	moon_illumination	moonrise	moonset	sunrise	...	WindChillC	WindGustKmph	cloud
2009-01-01 00:00:00	24	10	0.0	8.7	4	1	31	09:56 AM	09:45 PM	06:57 AM	...	11	21	
2009-01-01 01:00:00	24	10	0.0	8.7	4	1	31	09:56 AM	09:45 PM	06:57 AM	...	12	22	
2009-01-01 02:00:00	24	10	0.0	8.7	4	1	31	09:56 AM	09:45 PM	06:57 AM	...	12	23	
2009-01-01 03:00:00	24	10	0.0	8.7	4	1	31	09:56 AM	09:45 PM	06:57 AM	...	12	23	
2009-01-01 04:00:00	24	10	0.0	8.7	4	1	31	09:56 AM	09:45 PM	06:57 AM	...	14	19	

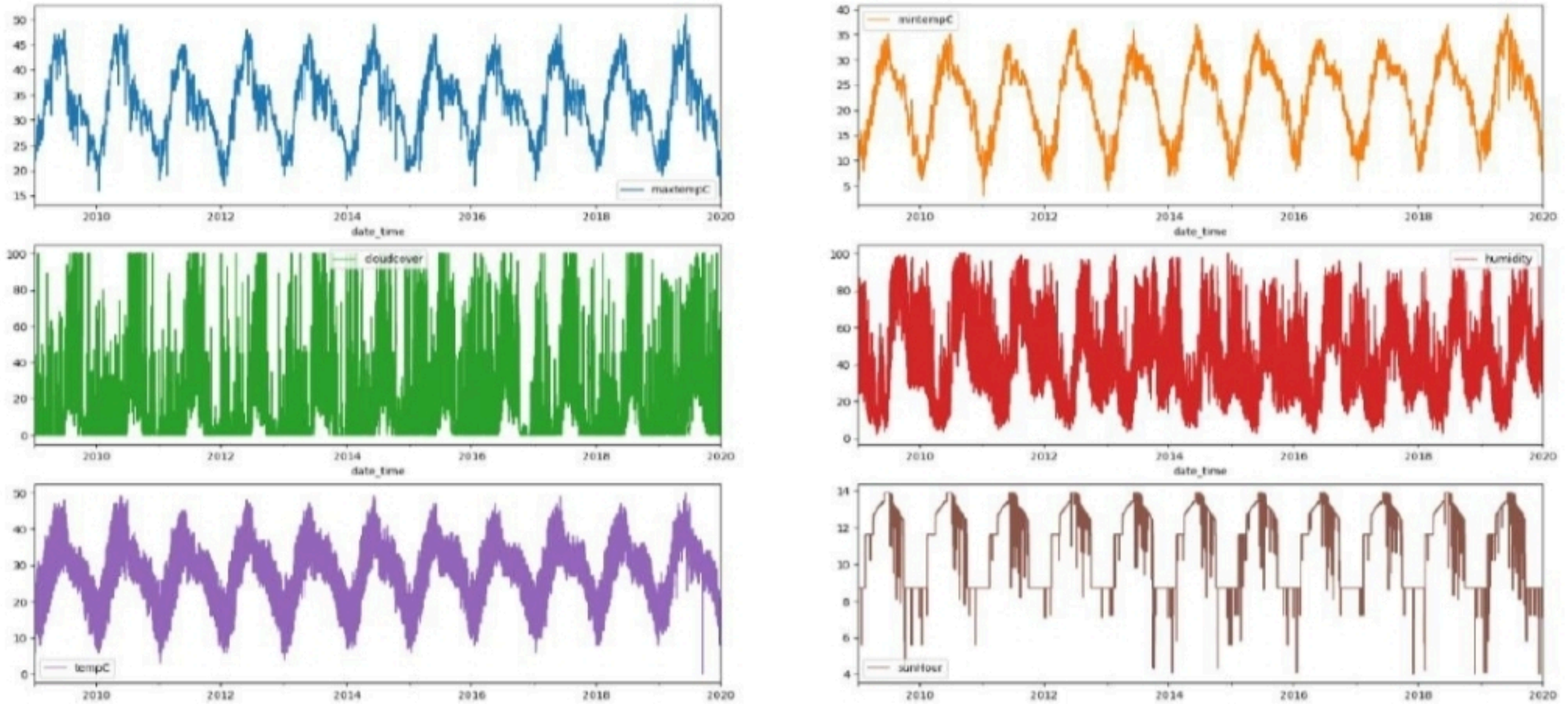


Figure 2.1: Plot for each factor for 10 years

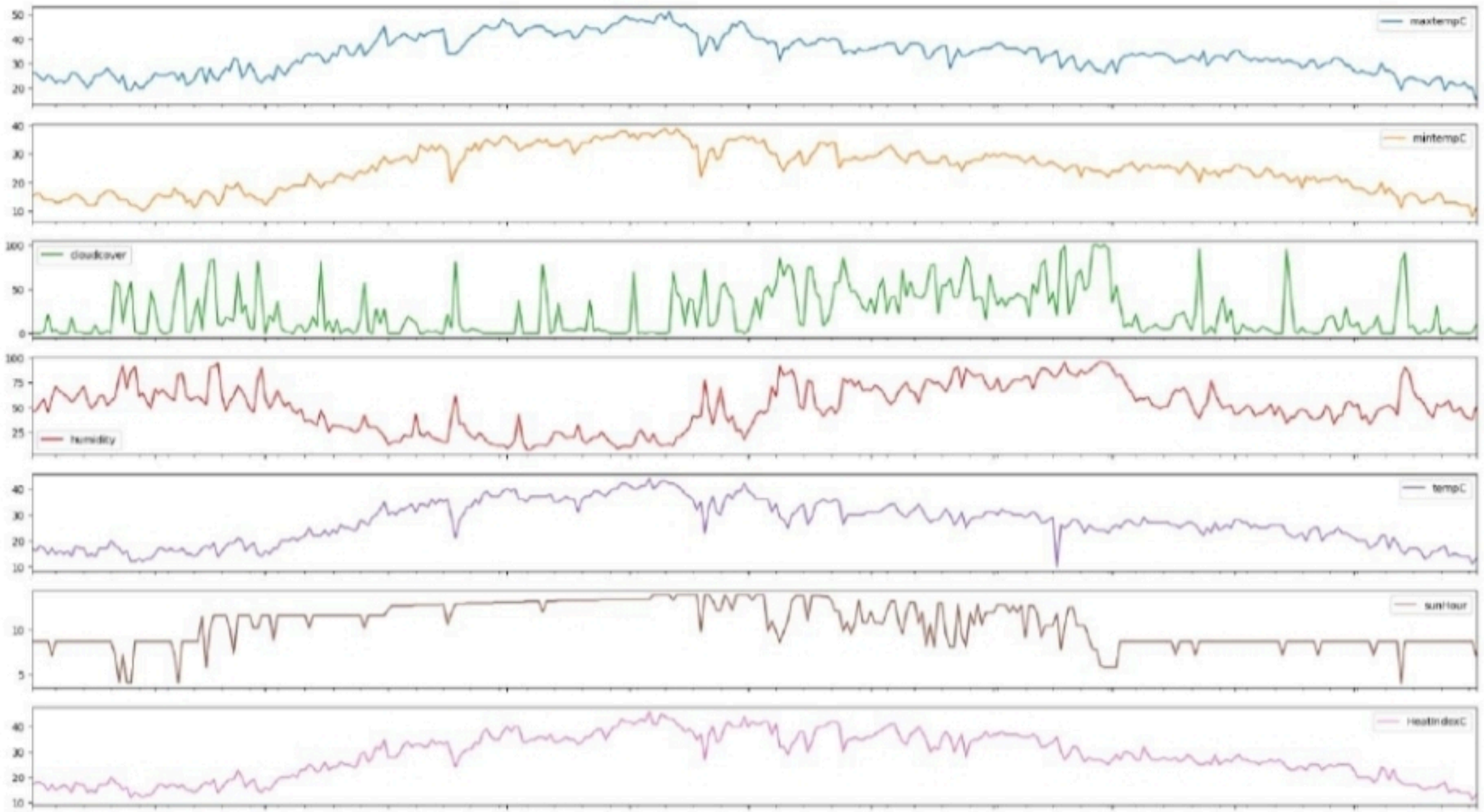


Figure 2.2: Plot for each factor for 1 year

Weather Forecasting Experimentation

Prediction and Evaluation

After training, we use the test set for **predictions** and evaluate them using metrics like **confusion matrix** and **R2 score**.

Model Training

We build classifiers such as **Linear Regression**, **Decision Tree Regression**, and **Random Forest Regression** using the train set data.

Data Separation

The dataset is divided into a **train set** and a **test set** for effective model training and evaluation.

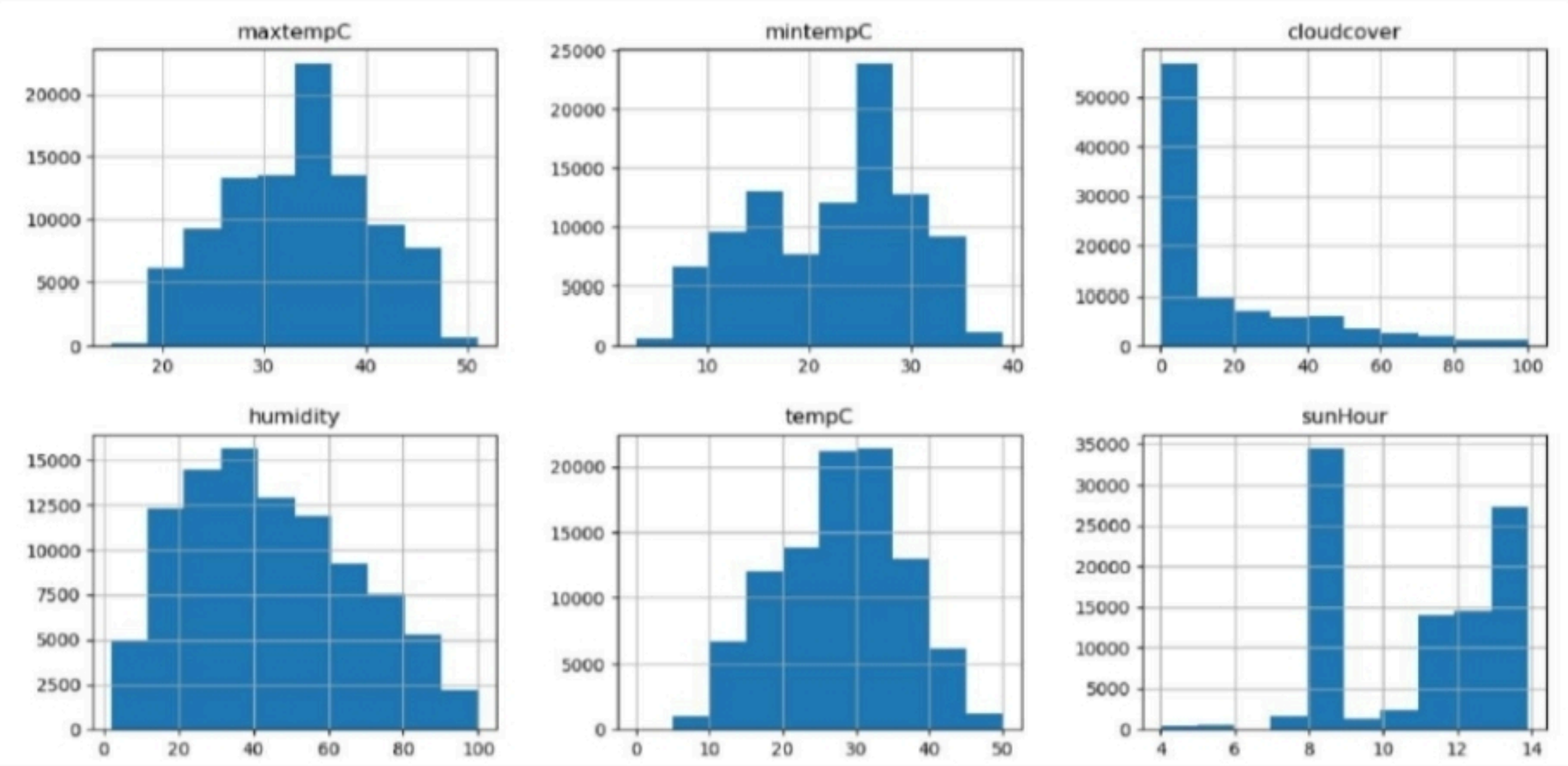
Parameter Tuning

Tuning the classifier **parameters** is crucial for achieving optimal accuracy in predictions.

Feature Extraction

Utilising **histograms** and **plots**, we extract features from the train set to facilitate model training.

EXPERIMENTATION



4. RESULT AND DISCUSSION

The results of the implementation of the project are demonstrated below.

Multiple Linear Regression:

This regression model has high mean absolute error, hence turned out to be the least accurate model. Given below is a snapshot of the actual result from the project implementation of multiple linear regression.

date_time	Actual	Prediction	diff
2013-07-10 08:00:00	34	34.89	-0.89
2015-11-04 20:00:00	25	24.57	0.43
2015-09-21 09:00:00	34	35.08	-1.08
2017-02-16 11:00:00	28	25.22	2.78
2012-07-21 01:00:00	28	28.04	-0.04
...	...	...	...
2019-03-30 09:00:00	37	33.55	3.45
2015-11-12 12:00:00	32	30.36	1.64
2019-12-31 05:00:00	8	9.13	-1.13
2019-08-02 17:00:00	35	35.92	-0.92
2019-10-22 08:00:00	26	25.77	0.23

Decision Tree Regression:

This regression model has medium mean absolute error, hence turned out to be the little accurate model. Given below is a snapshot of the actual result from the project implementation of multiple linear regression.

date_time	Actual	Prediction	diff
2013-07-10 08:00:00	34	34.0	0.0
2015-11-04 20:00:00	25	24.0	1.0
2015-09-21 09:00:00	34	34.0	0.0
2017-02-16 11:00:00	28	27.0	1.0
2012-07-21 01:00:00	28	28.0	0.0
...	...	...	...
2019-03-30 09:00:00	37	32.0	5.0
2015-11-12 12:00:00	32	32.0	0.0
2019-12-31 05:00:00	8	9.0	-1.0
2019-08-02 17:00:00	35	35.0	0.0
2019-10-22 08:00:00	26	26.0	0.0

19287 rows × 3 columns

Random Forest Regression:

This regression model has low mean absolute error, hence turned out to be the more accurate model. Given below is a snapshot of the actual result from the project implementation of multiple linear regression.

:

date_time	Actual	Prediction	diff
2013-07-10 08:00:00	34	33.92	0.08
2015-11-04 20:00:00	25	24.84	0.16
2015-09-21 09:00:00	34	34.25	-0.25
2017-02-16 11:00:00	28	27.00	1.00
2012-07-21 01:00:00	28	27.99	0.01
...	...	...	...
2019-03-30 09:00:00	37	32.79	4.21
2015-11-12 12:00:00	32	31.91	0.09
2019-12-31 05:00:00	8	8.81	-0.81
2019-08-02 17:00:00	35	34.98	0.02
2019-10-22 08:00:00	26	26.32	-0.32

19287 rows × 3 columns

CONCLUSION

All the machine learning models: linear regression, various linear regression, decision tree regression, random forest regression were beaten by expert climate determining apparatuses, even though the error in their execution reduced significantly for later days, demonstrating that over longer timeframes, our models may beat genius professional ones

Polynomial regression was shown to be a high predilection, low difference model, while linear regression was shown to be a low predisposition, high fluctuation model. Obtaining additional data is one way to enhance the linear regression model, which is inherently a high difference model due to its instability with outliers. However, practical regression showed a high propensity, indicating that the model's choice was bad and that additional information collection would not improve its predictions. This tendency may be due to the structure's choice to base temperature estimation on the climate over the last two days, which may be too brief to consider capturing slants in a climate as required by practical regression.

The practical regression model's bias could likely be reduced if the figure was instead based on the weather during the previous four or five days. Regardless, this will be left for later work because it would necessitate a lot more computation effort in addition to retraining the weight vector w.

When it comes to regression models, Random Forest Regression is the most accurate. It is probably the most widely used regression model because of its great accuracy and adaptability. An overview of the project's Random Forest implementation can be found below.

Weather Forecasting has a major test of foreseeing the precise outcomes which are utilized in numerous ongoing frameworks like power offices, air terminals, the travel industry focuses, and so forth. The trouble of this determining is the mind-boggling nature of parameters. Every parameter has an alternate arrangement of scopes of qualities.