<u>A BIAS CORRECTION OF NUMERICAL PREDICTION MODEL TEMPERATURE FORECAST</u>

PROJECT REPORT

Submitted by

PRITHU MISRA (ID : 12304507 , ROLL : RK22SAB64)
PRACHI (ID : 12212594, ROLL : RK22SAB41)
ADITI RAJ (ID : 12300419 , ROLL : RK22SAB62)
MANSI SINGH (ID : 12217190 , ROLL : RK22SAA15)

in partial fulfillment for the award of the degree

of

BACHELORS OF TECHNOLOGY

In

COMPUTER SCIENCE

LOVELY PROFESSIONAL UNIVERSITY

APRIL ,2024

CERTIFICATE

Certified that this project report "A MACHINE LEARNING MODEL FOR WEATHER FORECASTING" is the bonafide work of "Prachi, Mansi Singh, Aditi Raj and Prithu Misra**" who carried out the project workunder my supervision.**

Dr. Dhanpratap Singh
Guide

Head of the Department

COMPUTER SCIENCE & ENGINEERING

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my project guide Dr. Dhanpratap Singh Sir as well as our HOD who gave us the golden opportunity to do this wonderful project on the topic "**A BIAS CORRECTION OF NUMERICAL PREDICTION MODEL TEMPERATURE FORECAST**",

which also helped me in doing a lot of research and I came to know about so many new things I am really thankful to them.

Secondly I would also like to thank my parents and friends who helped me alot in finalizing this project within the limited time frame.

Date : 10-04-2024

Prachi

Mansi Singh

Aditi Raj

Prithu Misra

TABLE OF CONTENTS

ABSTRACT

Traditionally, climate assessment has been performed reliably by treating the environment as a liquid. However, this conventional approach, which views the atmosphere as a fluid system, encounters challenges in accurately predicting future states due to its susceptibility to oscillating effects and uncertainties. These limitations restrict climate forecasts to relatively short-term periods, typically no longer than 10 days. Recognizing the inadequacies of traditional methods, there is a growing interest in alternative approaches that leverage machine learning techniques. Machine learning offers a more robust and adaptable framework for weather prediction, exhibiting resilience to barometric destabilization effects and offering independence from the strict adherence to physical laws governing atmospheric processes.

Background

In the domain of weather forecasting, the India observatory relies on conventional methodologies encompassing four primary approaches: climatology, analog, persistence and trends, and numerical weather prediction. Climatology involves analyzing historical weather data spanning multiple years to establish average weather patterns. The analog method identifies past weather scenarios similar to the present forecast to inform predictions. Persistence and trends rely on historical trends without considering external factors, often resulting in less accurate forecasts. Numerical weather prediction employs complex mathematical models to simulate atmospheric conditions based on various atmospheric variables such as temperature, wind speed, pressure systems, and precipitation. While these traditional methods have provided valuable insights, they are not without limitations. They typically yield short-term forecasts and have yet to fully integrate machine learning algorithms, which offer potential enhancements in accuracy and predictive capability.

Objective (Brief)

This project aims to enhance weather prediction accuracy by integrating machine learning techniques into the forecasting process. Specifically, the objective is to develop models capable of predicting temperature variations over extended time periods, leveraging algorithms such as linear regression, random forest regression, and decision tree regression. The predictive models will incorporate a range of environmental variables including maximum and minimum temperatures, cloud cover, humidity, sun hours, precipitation, pressure, and wind speed to generate comprehensive forecasts.

1. INTRODUCTION

Weather prediction has traditionally relied on fluid dynamics principles, treating the atmosphere as a dynamic fluid system. This approach involves analyzing current environmental conditions and utilizing numerical methods to extrapolate future states of the atmosphere. However, despite advancements in modeling techniques, accurate weather forecasting beyond a 10-day horizon remains a challenge. Recognizing the potential for improvement through technological advancements, particularly in the field of machine learning, there is growing interest in exploring alternative methodologies to enhance predictive accuracy.

Machine Learning

Machine learning offers a paradigm shift in weather forecasting, providing a data-driven approach that is less reliant on traditional fluid dynamics models. Unlike traditional methods, which are susceptible to perturbations and require extensive physical variables for prediction, machine learning algorithms demonstrate robustness and adaptability in handling complex atmospheric phenomena. With the integration of machine learning, weather forecasting has witnessed notable improvements in accuracy and predictability, marking a significant evolution in the field.

USE OF ALGORITHMS:

There are different methods of foreseeing temperature utilizing Regression and a variety of Functional Regression, in which datasets are utilized to play out the counts and investigation. To Train, the calculations 80% size of information is utilized and 20% size of information is named as a Test set. For Example, if we need to anticipate the temperature of Kanpur, India utilizing these Machine Learning calculations, we will utilize 8 Years of information to prepare the calculations and 2 years of information as a Test dataset. The as opposed to Weather Forecasting utilizing Machine Learning Algorithms which depends essentially on reenactment dependent on Physics and Differential Equations, Artificial Intelligence is additionally utilized for foreseeing temperature: which incorporates models, for example, Linear regression, Decision tree regression, Random forest regression. To finish up, Machine Learning has enormously changed the worldview of Weather estimating with high precision and predictivity. What's more, in the following couple of years greater progression will be made utilizing these advances to precisely foresee the climate to avoid catastrophes like typhoons, Tornados, and Thunderstorms.

1. METHODOLOGY

The dataset utilized in this arrangement has been gathered from Kaggle which is “Historical Weather Data for Indian Cities” from which we have chosen the data for “WEST BENGAL”. The dataset was created by keeping in mind the necessity of such historical weather data in the community. The datasets for the top 8 Indian cities as per the population. The datasetwas used with the help of the worldweatheronline.com API and the wwo_hist package. The datasets contain hourly weather data from 08-04-2022 to 08-04-2024. The data of each cityis for more than 10 years. This data can be used to visualize the change in data due to globalwarming or can be used to predict the weather for upcoming days, weeks, months, seasons,etc.
Note: The data was extracted with the help of worldweatheronline.com API and we cannot guarantee the accuracy of the data.
The main target of this dataset can be used to predict the weather for the next day or week with huge amounts of data provided in the dataset. Furthermore, this data can also be used to make visualization which would help to understand the impact of global warming over the various aspects of the weather like precipitation, humidity, temperature, etc.

In this project, we are concentrating on the temperature prediction of Kanpur city with the help of various machine learning algorithms and various regressions. By applying various regressions on the historical weather dataset of Kanpur city we are predicting the temperature like first we are applying Multiple Linear regression, then Decision Tree regression, and after that, we are applying Random Forest Regression.

Table 2.1: Historical Weather Dataset of West Bengal

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	name	latitude	longitud	datetime	tempma	tempmir	temp	feelslike	feelslike	feelslike	dew	humidity	precip	precippr	precipco	preciptyp	snow	snowdep	windgus
2	West Be	22.57	88.371	4/8/2022	36.6	25	30.1	38.5	25	31.4	20.9	62.2	0	0	0		0	0	47.9
3	West Be	22.57	88.371	4/9/2022	39.3	25.6	30.9	40.1	25.6	32.6	21.4	62.2	0	0	0		0	0	45.4
4	West Be	22.57	88.371	######	36.6	26.3	30.4	40.5	26.3	33.1	22.6	66	0	0	0		0	0	55.8
5	West Be	22.57	88.371	4/11/2022	40.8	26.2	31.4	41.7	26.2	33.8	21.5	60.6	0	0	0		0	0	53.6
6	West Be	22.57	88.371	######	38.2	26.9	31.1	41.4	29.7	34.9	22.9	64.8	0	0	0		0	0	40
7	West Be	22.57	88.371	######	38.1	26.6	31.2	42.7	26.6	34.6	22.9	64.3	0	0	0		0	0	50.8
8	West Be	22.57	88.371	######	42.1	26.9	32.7	43.8	29.8	36.1	21.9	58	0.5	100	4.17	rain	0	0	46.4
9	West Be	22.57	88.371	######	43.5	27.8	33.4	43.5	31.1	36.2	20	54	0	0	0		0	0	49
10	West Be	22.57	88.371	######	40.9	27.3	32.4	42.6	30.8	35.9	22.1	59.7	0	0	0		0	0	37.8
11	West Be	22.57	88.371	######	39.1	27.5	32	42.1	30.9	35.6	22.2	59.8	0	0	0		0	0	51.8
12	West Be	22.57	88.371	######	39.7	26.6	32	41.8	26.6	35.2	22.2	61.4	0	0	0		0	0	39.2
13	West Be	22.57	88.371	######	36.9	26.3	31	40.3	26.3	34.3	22.9	65.3	0	0	0		0	0	49.3
14	West Be	22.57	88.371	######	36	27.1	31	40.9	30.5	35.3	23.4	66.3	0.3	100	4.17	rain	0	0	43.6
15	West Be	22.57	88.371	######	36.2	27.4	31	39.7	30.6	34.6	22.6	63	0.1	100	4.17	rain	0	0	40.3
16	West Be	22.57	88.371	######	39.2	27.5	31.7	42.6	30.7	35.8	22.8	62.4	0.1	100	4.17	rain	0	0	32.8
17	West Be	22.57	88.371	######	42.4	27.3	33.4	41.2	30.9	36	20.2	54.5	0	0	0		0	0	38.5
18	West Be	22.57	88.371	######	43.5	28	34.1	43.5	32	37	20	52.8	0	0	0		0	0	41.8
19	West Be	22.57	88.371	######	42.9	27.9	33.8	41.5	32.2	37	20.5	54.9	0	0	0		0	0	32.8
20	West Be	22.57	88.371	######	41.2	27.5	33	42.2	31.4	36.8	22.1	59.1	0	0	0		0	0	39.6
21	West Be	22.57	88.371	######	40.7	27.3	32.7	42.8	30.3	36.4	22.1	58.6	0	0	0		0	0	44.6
22	West Be	22.57	88.371	######	39.4	27.1	32.4	42.9	30.4	36.7	22.9	61.8	0	0	0		0	0	42.5
23	West Be	22.57	88.371	######	38.5	27	31.9	43.3	30.3	36.6	23.6	65.1	0	0	0		0	0	48.6
24	West Be	22.57	88.371	######	37.1	26.3	31.1	41.5	26.3	35	23.3	66.2	0	0	0		0	0	40.7
25	West Be	22.57	88.371	5/1/2022	37.4	25.6	31.1	41.3	25.6	34.8	22.9	64.6	2.7	100	20.83	rain	0	0	34.6
26	West Be	22.57	88.371	5/2/2022	38.3	27.7	31.7	42	30.4	35.8	22.8	62.1	2.8	100	20.83	rain	0	0	43.9
27	West Be	22.57	88.371	5/3/2022	36.6	24.5	30.7	41	24.5	34.5	22.9	65.4	23.3	100	25	rain	0	0	46.4
28	West Be	22.57	88.371	5/4/2022	37.2	26.9	31.3	40.6	29.7	34.8	22.5	62.5	0.8	100	12.5	rain	0	0	40
29	West Be	22.57	88.371	5/5/2022	38.3	24.8	30.7	39.9	24.8	33	21.2	59.9	16	100	25	rain	0	0	32
30	West Be	22.57	88.371	5/6/2022	37.9	26.4	31.5	39.2	26.4	33.8	20.6	55.1	0.2	100	4.17	rain	0	0	29.2

	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	windspe	windspe	windspe	windspe	winddir	sealevelp	cloudcov	visibility	solarradi	solarene	uvindex	severeris	sunrise	sunset	moonph	conditior	descripti	icon	source
2	28.8	28.8	20.7	11.9	189.3	1008.4	1.5	24.1	313.3	27.1	10	30	2022-04-	2022-04-	0.23	Clear	Clear cor	clear-day	obs
3	28.8	28.8	21.6	15.1	188.6	1006.7	1.2	24.1	315	27.2	10	30	2022-04-	2022-04-	0.25	Clear	Clear cor	clear-day	obs
4	32.8	32.8	26	17.6	187.5	1006	10.2	24.1	312.6	27.1	10	75	2022-04-	2022-04-	0.29	Clear	Clear cor	clear-day	obs
5	30.6	30.6	23.9	18	192.5	1004.1	35.7	24.1	318.8	27.6	10	75	2022-04-	2022-04-	0.32	Partially	Partly cl	partly-cl	obs
6	25.9	25.9	20.8	13	180	1005.1	25.7	24.1	311.9	27	10	75	2022-04-	2022-04-	0.35	Partially	Partly cl	partly-cl	obs
7	29.2	29.2	24.1	18	185.3	1004	11.4	24.1	313.4	27.1	10	75	2022-04-	2022-04-	0.39	Clear	Clear cor	clear-day	obs
8	34.9	34.9	22.3	7.6	187.7	1002.2	21.8	24.1	292.3	25.3	10	75	2022-04-	2022-04-	0.42	Rain, Pa	Partly cl	rain	obs
9	32.8	32.8	19.7	3.6	187.5	1002.3	37.8	24.1	284.6	24.5	10	75	2022-04-	2022-04-	0.45	Partially	Partly cl	partly-cl	obs
10	29.9	29.9	21.2	16.6	176	1002.3	41	24.1	290.2	25.1	10	60	2022-04-	2022-04-	0.48	Partially	Partly cl	partly-cl	obs
11	29.5	29.5	21.7	16.6	180	1003.2	51.9	24.1	314.6	27.2	10	60	2022-04-	2022-04-	0.5	Partially	Partly cl	partly-cl	obs
12	27	27	19.8	11.5	179.3	1005.4	25.6	24.1	321.3	27.6	10	60	2022-04-	2022-04-	0.55	Partially	Partly cl	partly-cl	obs
13	32.8	32.8	22.1	13.7	183.8	1006.7	7.6	24.1	318.8	27.5	10	60	2022-04-	2022-04-	0.59	Clear	Clear cor	clear-day	obs
14	29.9	29.9	23.4	12.6	181.7	1006.5	24.2	24.1	314.9	27.2	10	75	2022-04-	2022-04-	0.62	Rain, Pa	Clearing	rain	obs
15	28.8	28.8	19.9	11.2	188.4	1006.7	35.6	24.1	294	25.4	10	75	2022-04-	2022-04-	0.66	Rain, Pa	Becomir	rain	obs
16	28.4	28.4	20.3	14	187	1006.5	25.8	24.1	310.8	26.7	10	75	2022-04-	2022-04-	0.69	Rain, Pa	Partly cl	rain	obs
17	23	23	16.9	8.3	198	1005.8	9.8	24.1	298.6	25.8	10	75	2022-04-	2022-04-	0.75	Clear	Clear cor	clear-day	obs
18	27.7	27.7	18.7	9	201.9	1003.8	5.1	24.1	324.5	27.9	10	60	2022-04-	2022-04-	0.77	Clear	Clear cor	clear-day	obs
19	30.2	30.2	17.3	10.4	189.5	1003.7	2.5	24.1	326.8	28.2	10	60	2022-04-	2022-04-	0.8	Clear	Clear cor	clear-day	obs
20	30.6	30.6	21.7	15.8	188.7	1004.4	2.4	24.1	325.2	28	10	60	2022-04-	2022-04-	0.84	Clear	Clear cor	clear-day	obs
21	32	32	21.5	11.5	187.7	1004.2	3.8	24.1	324	28	10	60	2022-04-	2022-04-	0.87	Clear	Clear cor	clear-day	obs
22	24.8	24.8	19.6	7.2	183.1	1005.2	2.6	24.1	326.8	28.3	10	60	2022-04-	2022-04-	0.91	Clear	Clear cor	clear-day	obs
23	28.1	28.1	20.1	11.5	186	1005	1.7	24.1	325.7	27.9	10	60	2022-04-	2022-04-	0.94	Clear	Clear cor	clear-day	obs
24	29.5	29.5	21	11.9	176.7	1003.7	17.5	24.1	324	27.9	10	60	2022-04-	2022-04-	0.98	Clear	Clear cor	clear-day	obs
25	30.2	30.2	19.1	6.8	172.5	1004.1	37.1	23.8	322.7	27.8	10	60	2022-05-	2022-05-	0	Rain, Pa	Partly cl	rain	obs
26	31	31	20.3	14.8	181.8	1003.6	54.5	24	310.5	26.9	10	60	2022-05-	2022-05-	0.05	Rain, Pa	Partly cl	rain	obs
27	29.2	29.2	19.8	11.2	178.8	1005	38.8	22.3	285.8	24.5	10	75	2022-05-	2022-05-	0.08	Rain, Pa	Partly cl	rain	obs
28	24.8	24.8	14.8	2.5	197	1006.5	20.1	23.7	321.2	27.8	10	60	2022-05-	2022-05-	0.11	Rain, Pa	Partly cl	rain	obs
29	28.4	28.4	11.5	5.8	205.3	1005.8	36.8	22.6	293	25.3	10	30	2022-05-	2022-05-	0.14	Rain, Pa	Becomir	rain	obs
30	24.5	24.5	9.8	0.7	194.4	1005.1	29.1	24.1	301.9	26.1	10	30	2022-05-	2022-05-	0.18	Rain, Pa	Partly cl	rain	obs

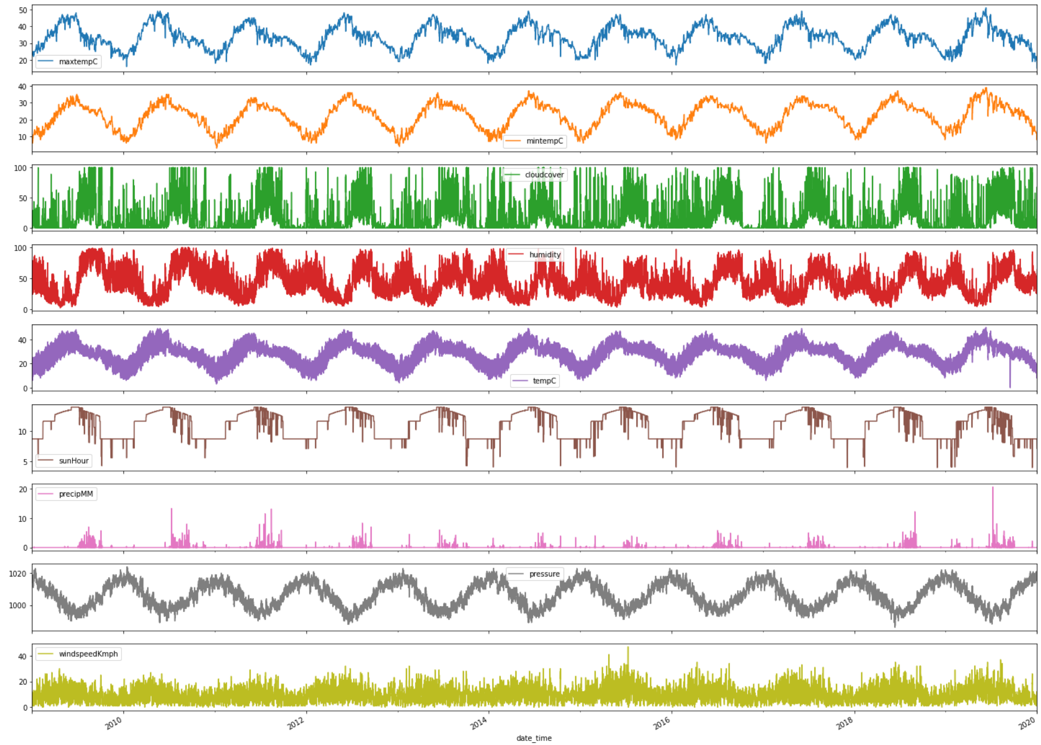


Figure 2.1: Plot for each factor for 3 years

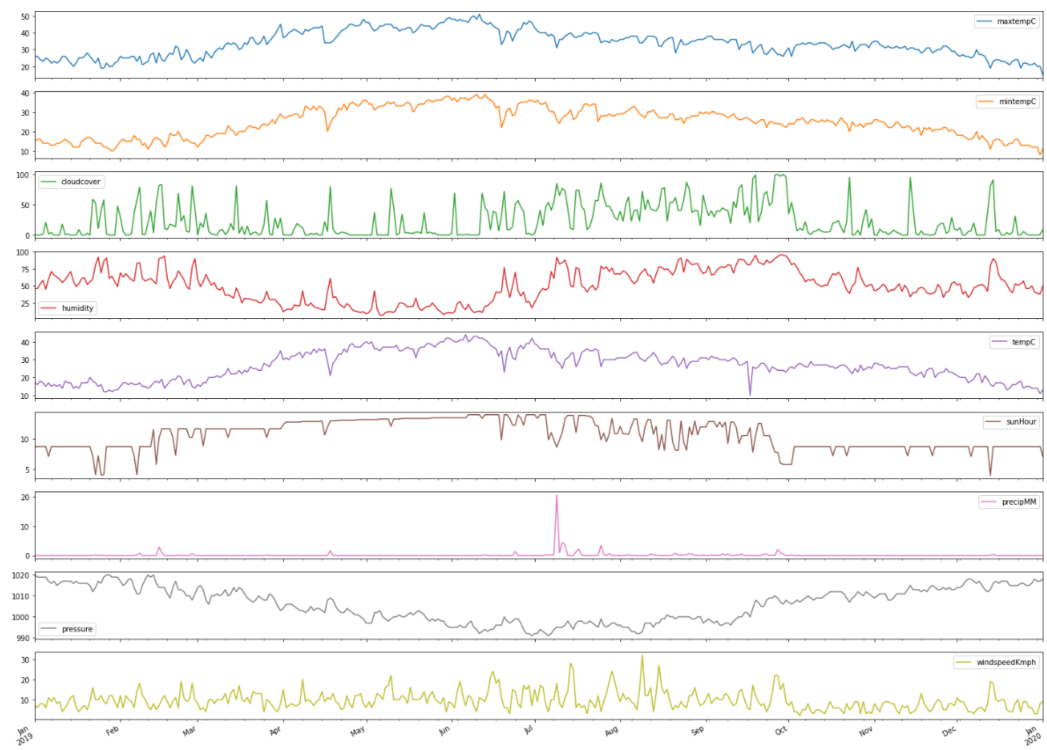
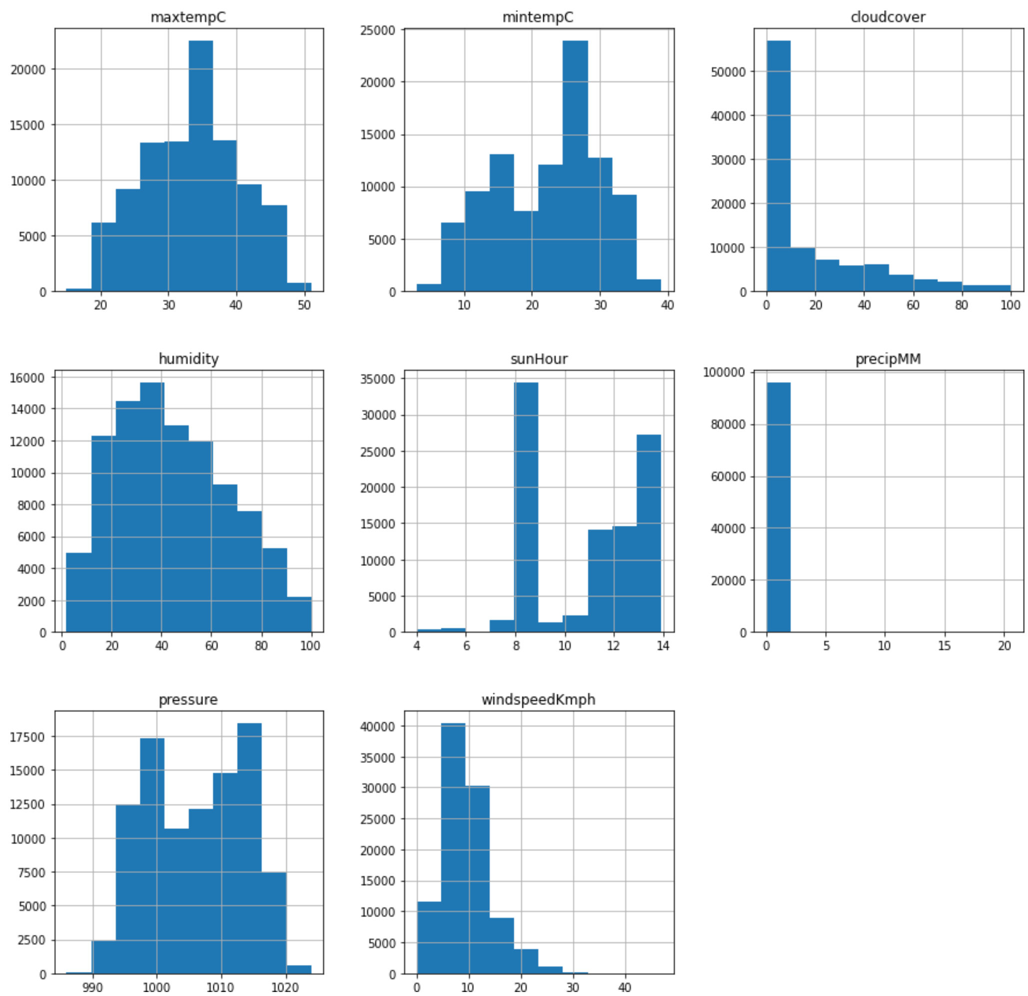


Figure 2.2: Plot for each factor for 1 year

2. EXPERIMENTATION

The record has just been separated into a train set and a test set. Each information has just been labeled. First, we take the trainset organizer. We will train our model with the help of histograms and plots. The feature so extracted is stored in a histogram. This process is done for every data in the train set. Now we will build the model of our classifiers. The classifiers which we will take into account are Linear Regression, Decision Tree Regression, and Random Forest Regression. With the help of our histogram, we will train our model. The most important thing in this process is to tune these parameters accordingly, such that we get the most accurate results. Once the training is complete, we will take the test set. Now for each data variable of the test set, we will extract the features using feature extraction techniques and then compare its values with the values present in the histogram formed by the train set. The output is then predicted for each test day. Now in order to calculate accuracy, we will compare the predicted value with the labeled value. The different metrics that we will use confusion matrix, R2 score, etc.



3. 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.

datetime	Actual	Prediction	diff
2023-12-01	26.5	26.58	-0.08
2023-05-20	33.0	32.81	0.19
2023-02-12	26.3	26.22	0.08
2023-01-15	23.2	23.42	-0.22
2023-08-17	30.9	30.92	-0.02
...	...	...	...
2023-05-27	33.2	32.88	0.32
2023-09-02	32.4	32.55	-0.15
2023-08-21	30.1	29.97	0.13
2023-11-01	28.3	28.07	0.23
2023-06-09	31.7	31.31	0.39

73 rows × 3 columns

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.

	Actual	Prediction	diff
datetime			
2023-12-01	26.5	26.3	0.2
2023-05-20	33.0	34.3	-1.3
2023-02-12	26.3	26.5	-0.2
2023-01-15	23.2	23.2	0.0
2023-08-17	30.9	30.3	0.6
...	...	...	...
2023-05-27	33.2	34.5	-1.3
2023-09-02	32.4	33.0	-0.6
2023-08-21	30.1	30.0	0.1
2023-11-01	28.3	28.8	-0.5
2023-06-09	31.7	30.9	0.8

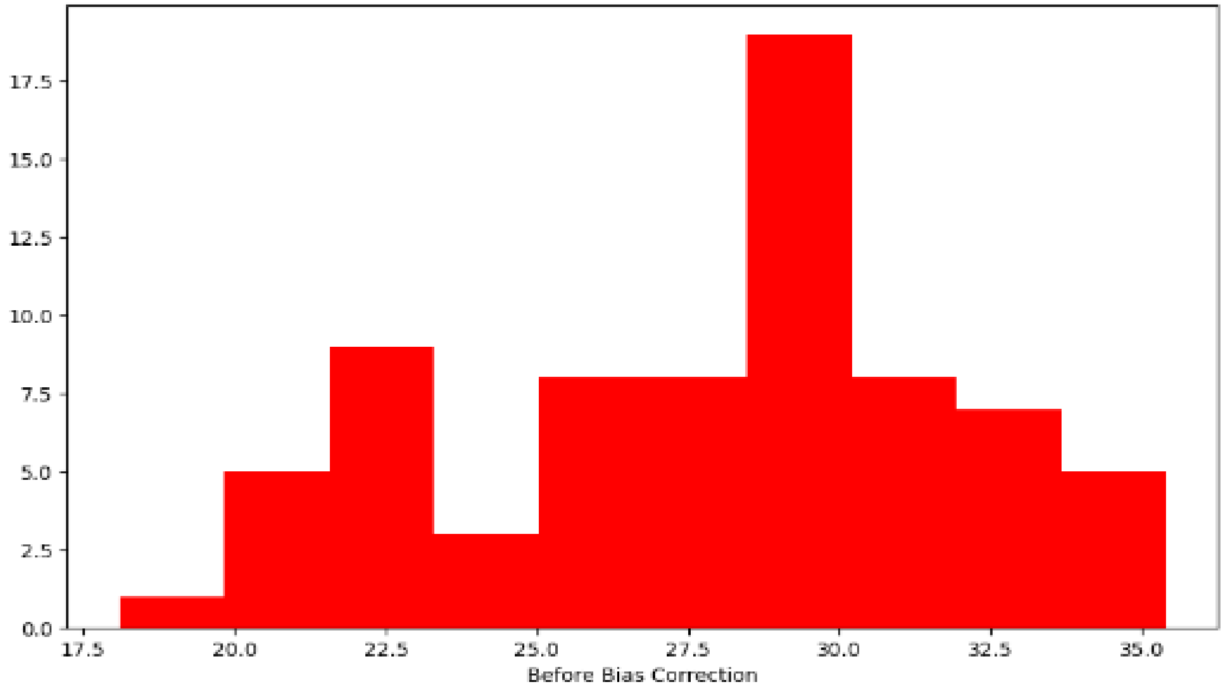
73 rows × 3 columns

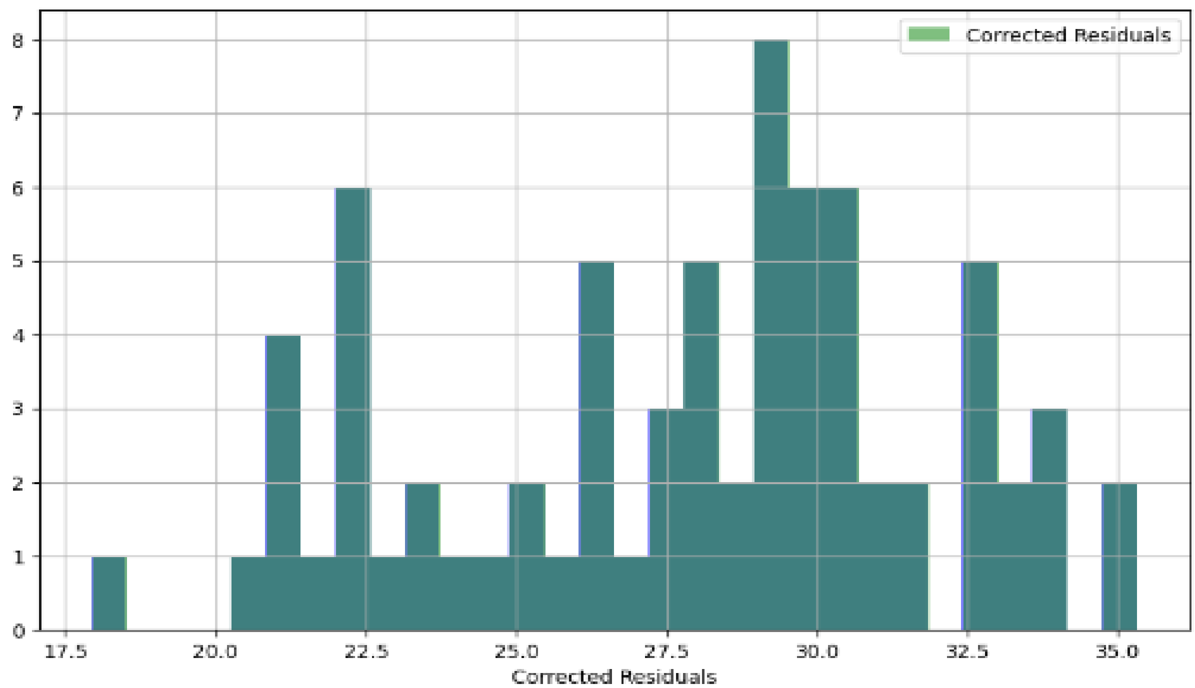
Random Forest Regression:

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.

	Actual	Prediction	diff
datetime			
2023-12-01	26.5	26.52	-0.02
2023-05-20	33.0	33.68	-0.68
2023-02-12	26.3	25.98	0.32
2023-01-15	23.2	23.27	-0.07
2023-08-17	30.9	30.98	-0.08
...	...	...	...
2023-05-27	33.2	33.47	-0.27
2023-09-02	32.4	32.40	0.00
2023-08-21	30.1	30.20	-0.10
2023-11-01	28.3	28.27	0.03
2023-06-09	31.7	31.15	0.55

73 rows × 3 columns


17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
17.5
20.0
22.5
25.0
27.5
30.0
32.5
35.0
Before Bias Correction


Corrected Residuals
8
7
6
5
4
3
2
1
0
17.5
20.0
22.5
25.0
27.5
30.0
32.5
35.0
Corrected Residuals

4. 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.

Linear regression demonstrated to be a low predisposition, high fluctuation model though polynomial regression demonstrated to be a high predisposition, low difference model. Linear regression is naturally a high difference model as it is unsteady to outliers, so one approach to improve the linear regression model is by gathering more information. Practical regression, however, was high predisposition, demonstrating that the decision of the model was poor and that its predictions can't be improved by the further accumulation of information. This predisposition could be expected to the structure decision to estimate temperature dependent on the climate of the previous two days, which might be too short to even think about capturing slants in a climate that practical regression requires. On the off chance that the figure was rather founded on the climate of the past four or five days, the predisposition of the practical regression model could probably be decreased. In any case, this would require significantly more calculation time alongside retraining of the weight vector w, so this will be conceded to future work.

Talking about Random Forest Regression, it proves to be the most accurate regression model. Likely so, it is the most popular regression model used, since it is highly accurate and versatile. Below is a snapshot of the implementation of Random Forest in the project.

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.