ADDIS ABABA UNIVERSITY
ADDIS ABABA INSTITUTE OF TECHNOLOGY

Introduction to AI

Final Project Report – Weather Forecast for Addis Ababa

Group Members

1. Yohannes Solomon UGR/3323/13	Section 1
2. Dawit Belay UGR/8622/14	Section 1
3. Sifan Fita UGR/8856/14	Section 2
4. Ahmed Muhammed UGR/8920/14	Section 1

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.

Machine-Learning-Model-for-Weather-Forecasting

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.

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 Addis Ababa, Ethiopia 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.

METHODOLOGY

The dataset was used with the help of the worldweatheronline.com API and the wwo_hist package. The datasets contain hourly weather data from 01-01-2014 to 01-01-2024. The data of each city is for more than 10 years. This data can be used to visualize the change in data due to global warming 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 Addis Ababa 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.

Historical Weather Dataset of Addis Ababa

| maxtempC	mintempC	totalSnow_cm	sunHour	uvIndex	moon_illumination	moonrise	moonset	sunrise	sunset	...
date-time										
2014-01-01 0	20	9	0.0	10.0	4	0	06:20 AM	06:25 PM	06:40 AM	06:17
2014-01-01 300	20	9	0.0	10.0	4	0	06:20 AM	06:25 PM	06:40 AM	06:17
2014-01-01 600	20	9	0.0	10.0	4	0	06:20 AM	06:25 PM	06:40 AM	06:17
2014-01-01 900	20	9	0.0	10.0	4	0	06:20 AM	06:25 PM	06:40 AM	06:17
2014-01-01 1200	20	9	0.0	10.0	4	0	06:20 AM	06:25 PM	06:40 AM	06:17

Couldn't contain the whole table so it's listed in the notebook.

Plots for each factor for 10 years
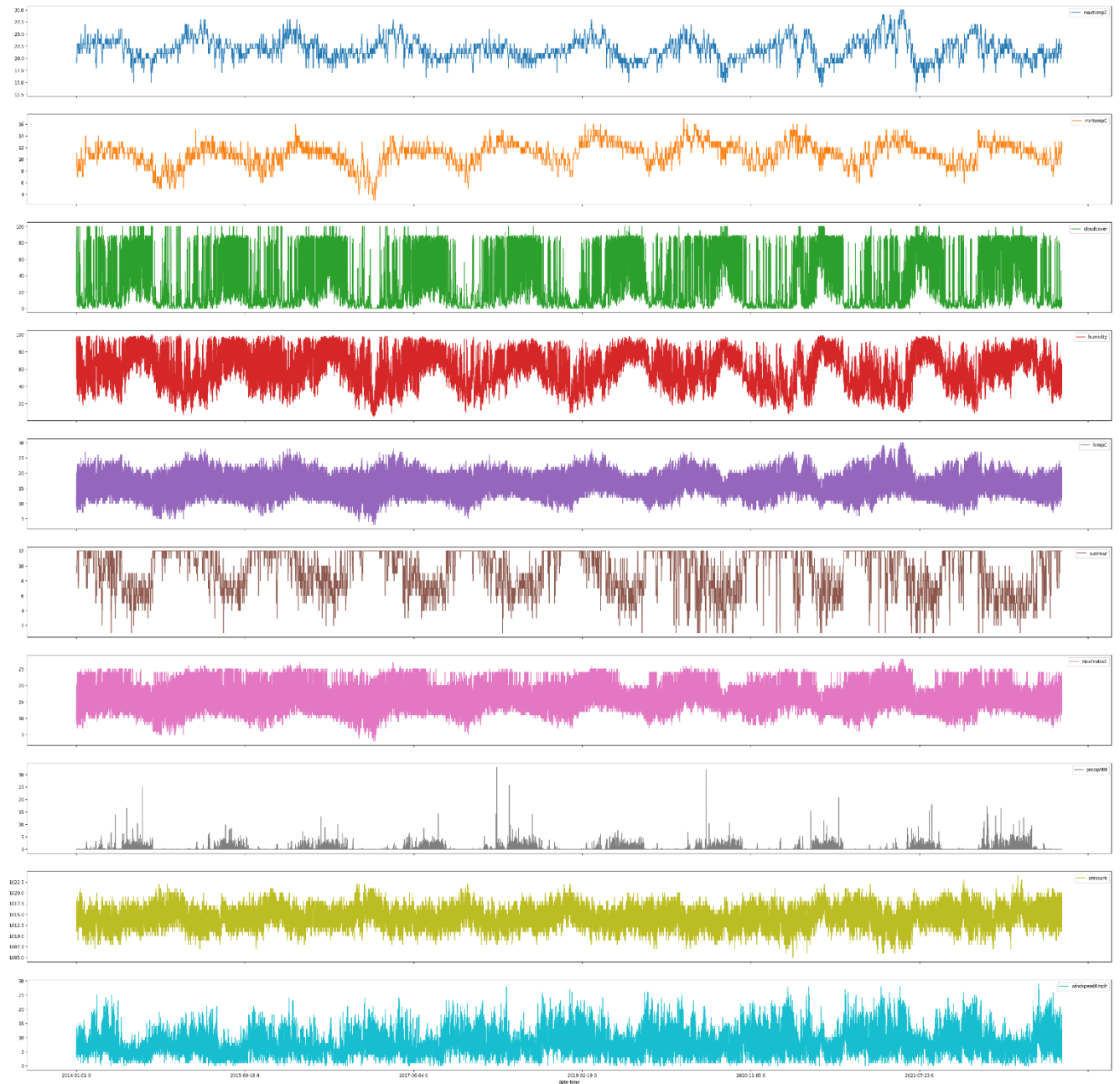
Plot for each factor for one year
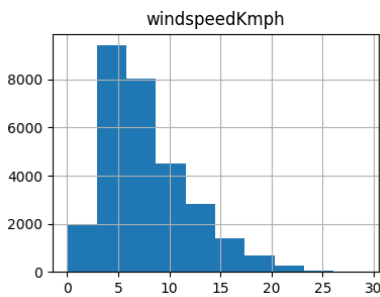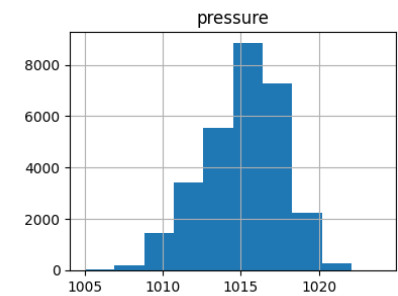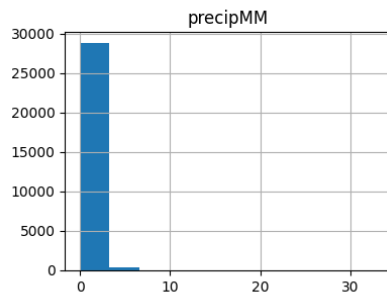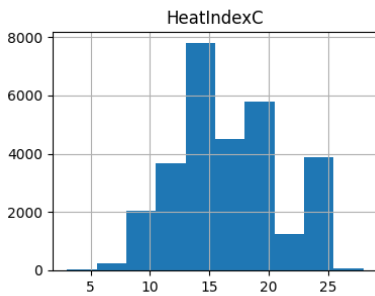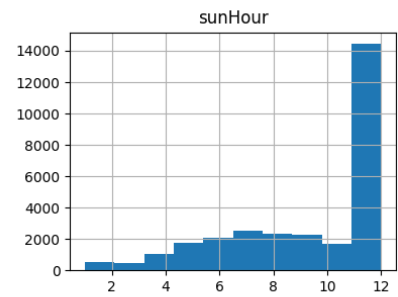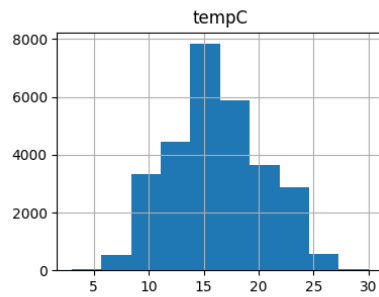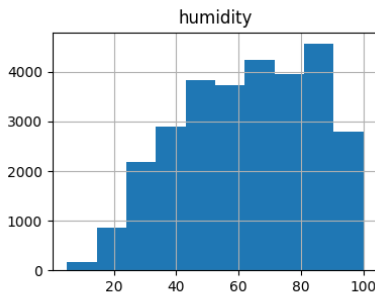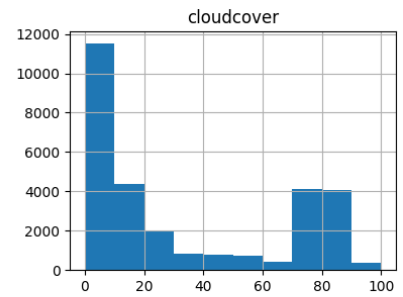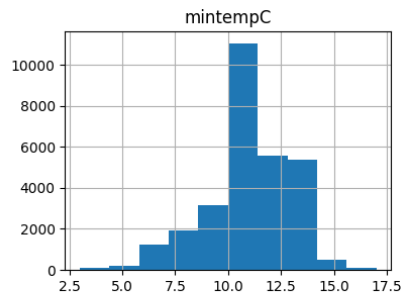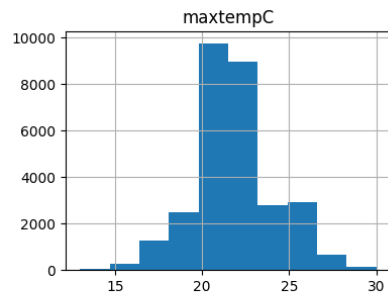
Multiple Linear Regression

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
2014-05-31 1200	24	24.11	-0.11
2021-03-25 300	14	14.33	-0.33
2015-12-11 2100	15	14.71	0.29
2017-04-22 0	15	14.91	0.09
2022-01-27 900	13	13.10	-0.10
...	...	...	...
2014-11-30 1800	17	16.63	0.37
2019-02-26 300	13	13.13	-0.13
2016-04-12 0	15	14.86	0.14
2015-12-01 300	10	10.08	-0.08
2020-12-24 1500	23	23.23	-0.23

Decision Tree Regression

This regression model has medium equal to the next regression in most tests 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
2014-05-31 1200	24	23.0	1.0
2021-03-25 300	14	14.0	0.0
2015-12-11 2100	15	15.0	0.0
2017-04-22 0	15	15.0	0.0
2022-01-27 900	13	13.0	0.0
...	...	...	...
2014-11-30 1800	17	17.0	0.0
2019-02-26 300	13	13.0	0.0
2016-04-12 0	15	15.0	0.0
2015-12-01 300	10	10.0	0.0
2020-12-24 1500	23	23.0	0.0

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
2014-05-31 1200	24	23.73	0.27
2021-03-25 300	14	14.00	0.00
2015-12-11 2100	15	15.00	0.00
2017-04-22 0	15	15.00	0.00
2022-01-27 900	13	13.00	0.00
...	...	...	...
2014-11-30 1800	17	17.00	0.00
2019-02-26 300	13	13.00	0.00
2016-04-12 0	15	15.00	0.00
2015-12-01 300	10	10.00	0.00
2020-12-24 1500	23	22.95	0.05