Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024

WEATHER FORECASTING

Satya Ranjan Pattanaik, Computer Science And Engineering Gandhi Institute For Technology, INDIA satyaranjan.pattanaik2020@gift.edu.in
Sudhira Dash, Computer Science And Engineering Gandhi Institute For Technology, INDIA sudhira.dash2020@gift.edu.in
Prof. Rajeswari Chaini Computer Science And Engineering Gandhi Institute For Technology, INDIA

ABSTRACT— In Data science Weather forecasting is an application of time series, forecasting where we use time series data and algorithms to make a forecast for a particular time and with respect to location.
By using the suitable algorithms we can predict the weather condition for the next few days. After finding the suitable data set we have to analyse by using the respective modules like pandas, non pie, math plot lib, So that we can predict and analyse weather and forecasting.
Weather forecasting is the process of predicting atmospheric conditions at a specific location over a given period, typically ranging from hours to days in advance.
It involves the analysis of various meteorological data, such as temperature, humidity, wind speed, and atmospheric pressure, collected from ground-based stations, satellites, radar systems, and numerical models.
weather forecasting remains a crucial tool for mitigating risks associated with weather-related events and enhancing societal resilience to natural hazards.

Keywords*:*
“Temperature”, “Pressure”, “Humidity”, “Wind”, “Precipitation”, “Cloud cover”, “Visibility”, “Dew point”, “Fronts”, “Forecast”, “Severe weather”, “Climate”, “Model”.

1. INTRODUCTION

Information about final paper submission is available from the conference website. In the world wide the major concern is the climate change, Which has many factors like forest deforestation, Air pollution, Water pollution, Soil pollution etc. The result of climate change comes as a unpredicting weather.
Weather forecasting is the task of concern department with respect to location and time. In case of our state cyclone come very often and we also found the plot very regularly.
weather forecasting has significant implications for public safety and disaster preparedness. Timely and accurate forecasts enable governments, emergency responders, and communities to anticipate and mitigate the impacts of natural disasters such as hurricanes, tornadoes, floods, and wildfires, thereby saving lives and reducing property damage.
weather forecasting represents a remarkable fusion of scientific knowledge, technological innovation, and human expertise. It serves as a vital tool for navigating the dynamic and often unpredictable nature of the Earth's atmosphere, providing valuable insights that help society adapt to and thrive in a world shaped by weather and climate.
As we are in the line of equator we found heat wave, because of the climate change. Weather forecasting is necessary to avoid the human livestock dearth.

1.1 OVERVIEW

Weather forecasting is the science and art of predicting atmospheric conditions at a specific location and time in the future. It involves analyzing current weather data, understanding atmospheric processes, and using mathematical models to forecast future weather patterns. Weather forecasts provide valuable information about temperature, precipitation, wind, humidity, and other

meteorological parameters, helping individuals, businesses, and governments make informed decisions and take appropriate actions to mitigate risks and optimize operations.

1.2 PROBLEM STATEMENT

Heavy rainfall can lead to numerous hazards, for instance: flooding, including danger to human life, harm to structures and framework, and loss of products and domesticated animals. avalanches, which can compromise human life, upset transport and interchanges, and cause harm to structures and foundation.

Where overwhelming precipitation happens with high breezes, hazard to ranger service crops is high.. In the case of initial treatment of patients, the probability of survival has increased significantly with early diagnosis of breast cancer. With proper tumor classification, unnecessary treatment can be avoided. Each volume should be treated differently.

Therefore, if there is no proper diagnosis then there is a high risk of death for the patient. Correct diagnosis of breast cancer and classification of tumors in benign and malignant tumors is an area of investigation. For example if we consider an area affected by tropical cyclone the fundamental impacts of tropical cyclone incorporate heavy rain, strong wind, huge tempest floods close landfall, and tornadoes.

The devastation from a tropical cyclone, for example, a sea tempest or hurricane, depends for the most part on its power, its size, and its area.

Tropical tornados act to evacuate woods shade and additionally change the scene close beach front zones, by moving and reshaping sand ridges and causing broad disintegration along the drift. Indeed, even well inland, overwhelming precipitation can prompt mudslides and avalanches in rugged regions.

Their belongings can be detected after some time by concentrate the convergence of the Oxygen-18 isotope inside caverns inside the region of typhoons' ways. So we are providing a better way to get accurate predictions. As mentioned above, the benefits of identifying important features of mechanical learning, complex data sets, play an important role in forecasting of weather. Since the best results can be achieved with engineering learning algorithms, we should use these techniques to aware people from natural disasters. This is because learning engineering algorithms can provide more accurate results. Apart from this, the results are achieved at a short time and people get enough time to do preparations or to escape from that place

1.3 OBJECTIVES

Our project aims to predict the Weather and Atmosphere conditions using the previous dataset of the weather forecasting with a focus on improving the accuracy of prediction. This will increase the accuracy of the weather prediction and we will get accurate results than the traditional methods. Our dataset consists of max and min. temperature of everyday from the specific location.

Classifications: When gathering datasets to give to the models there are sure parameters which are called as ordered information which incorporates: snow, rainstorm, rain, mist, cloudy, for the most part overcast, halfway shady, scattered mists, and clear. These can be additionally ordered into four classes.

1. Rain, tempest, and snow into precipitation
2. For the most part shady, foggy, and cloudy into exceptionally shady
3. Scattered mists and somewhat shady into modestly shady
4. Clear as clear Thus our aim is to provide accurate result in order to provide correct prediction of weather for future so in critical conditions people can be aware of upcoming natural calamities.

2. LITERATURE SURVEY

Weather forecasting has come a long way from simply observing red sunsets and predicting fair weather. Today, it's a complex field that integrates various methodologies to provide increasingly accurate predictions. Here's a look at what the literature reveals:

2.1 Traditional Techniques

- **Statistical Methods:** These methods, like ARIMA (Autoregressive Integrated Moving Average) models, analyse historical data to identify patterns and trends for future predictions.
- **Numerical Weather Prediction (NWP):** This approach relies on powerful computer models that solve complex equations governing atmospheric behaviour, providing detailed forecasts.

2.2 Modern Advancements

- **Machine Learning (ML):** ML algorithms like Artificial Neural Networks (ANNs) are trained on vast weather datasets to learn complex relationships and improve forecast accuracy.
- **Data Assimilation:** This technique incorporates real-time observations from satellites, weather stations, and balloons into NWP models for more up-to-date predictions.

A literature survey of weather forecasting involves reviewing relevant research articles, academic papers, books, and other scholarly sources to understand the current state of knowledge, recent advancements, and ongoing research trends in the field. Here's an overview of key topics and areas typically covered in a literature survey of weather forecasting.

- **Historical Development:** Explore the historical evolution of weather forecasting, from early empirical methods to modern numerical weather prediction techniques. Identify seminal works and milestones that have shaped the field over time.
- **Observational Techniques:** Review the principles and technologies used for weather observations, including ground-based weather stations, satellites, radar systems, weather balloons, and unmanned aerial vehicles (UAVs). Examine recent advancements in observational capabilities and their impact on forecast accuracy.
- **Model Evaluation And Modification:** Discuss methods for evaluating the performance of weather forecast models, including verification scores, skill metrics, and case studies. Analyse recent studies assessing the strengths and limitations of NWP models under different weather scenarios and geographical regions.
- **Predictability and Uncertainty:** Explore the factors influencing weather predictability, such as atmospheric dynamics, boundary conditions, model resolution, and initial condition uncertainties. Review research on quantifying and reducing forecast uncertainty through ensemble forecasting, probabilistic methods, and model post-processing techniques.
- **Extreme Weather Events:** Investigate the prediction and impacts of extreme weather phenomena, including hurricanes, tornadoes, heatwaves, droughts, floods, and severe storms. Examine studies on the predictability, intensity, and frequency of extreme events under changing climate conditions.
- **Climate Change and Weather Forecasting:** Review literature on the interactions between weather forecasting and climate change, including the influence of climate variability on weather patterns, long-term trends in extreme weather, and the role of weather forecasts in climate adaptation and mitigation strategies.
- **Machine Learning and Artificial Intelligence:** Explore the applications of machine learning, neural networks, and other artificial intelligence techniques in weather forecasting. Survey recent research on data-driven approaches for improving forecast skill, enhancing model performance, and addressing computational challenges in NWP.
- **Communication and Decision Support:** Examine studies on effective communication strategies for conveying weather forecast information to end users, including the general public, emergency managers, policymakers, and industry stakeholders. Discuss the role of decision support systems, visualization tools, and user-centric design principles in enhancing weather-related decision-making and risk management.
- **Future Directions and Challenges:** Identify emerging research areas, unresolved questions, and future challenges in weather forecasting, such as improving forecast lead times, enhancing regional and local-scale predictions, integrating socio-economic factors into forecast models, and addressing ethical and equity considerations in weather-related decision support.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024

3. 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 "Kanpur City".

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 dataset was used with the help of the worldweatheronline.com API and the wwo_hist package.

The datasets contain hourly weather data from 01-01-2009 to 01-01-2020. 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 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 3.1: Historical Weather Dataset of Kanpur City

date_time	maxtempC	mintempC	cloudcover	humidity	tempC	sunHour	precipMM	pressure	windspeedKmph
2009-01-01 00:00:00	24	10	17	50	11	8.7	0.0	1015	10
2009-01-01 01:00:00	24	10	11	52	11	8.7	0.0	1015	11
2009-01-01 02:00:00	24	10	6	55	11	8.7	0.0	1015	11
2009-01-01 03:00:00	24	10	0	57	10	8.7	0.0	1015	12
2009-01-01 04:00:00	24	10	0	54	11	8.7	0.0	1016	11

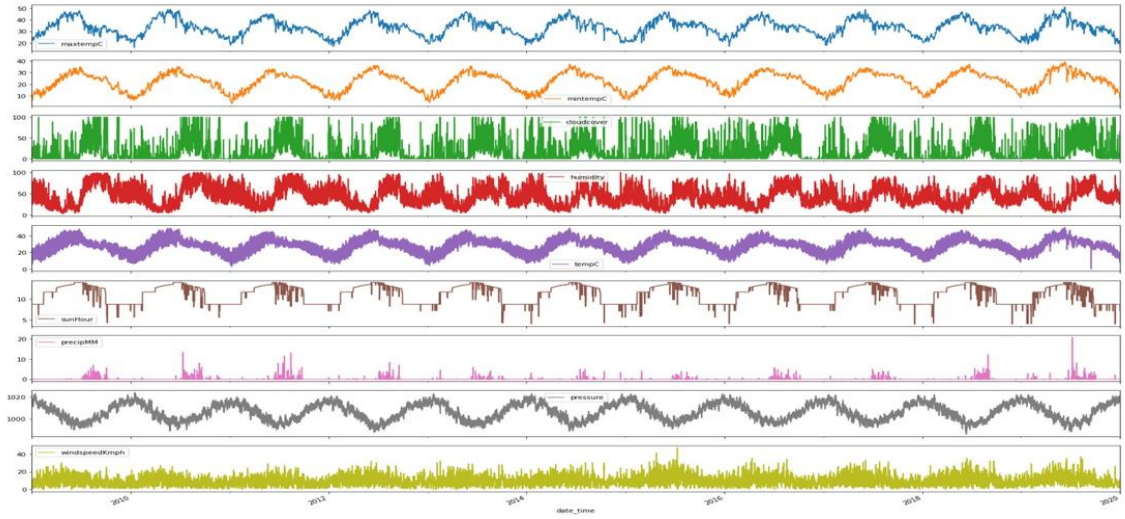

Figure 3.1: Plot for each factor for 10 years

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024

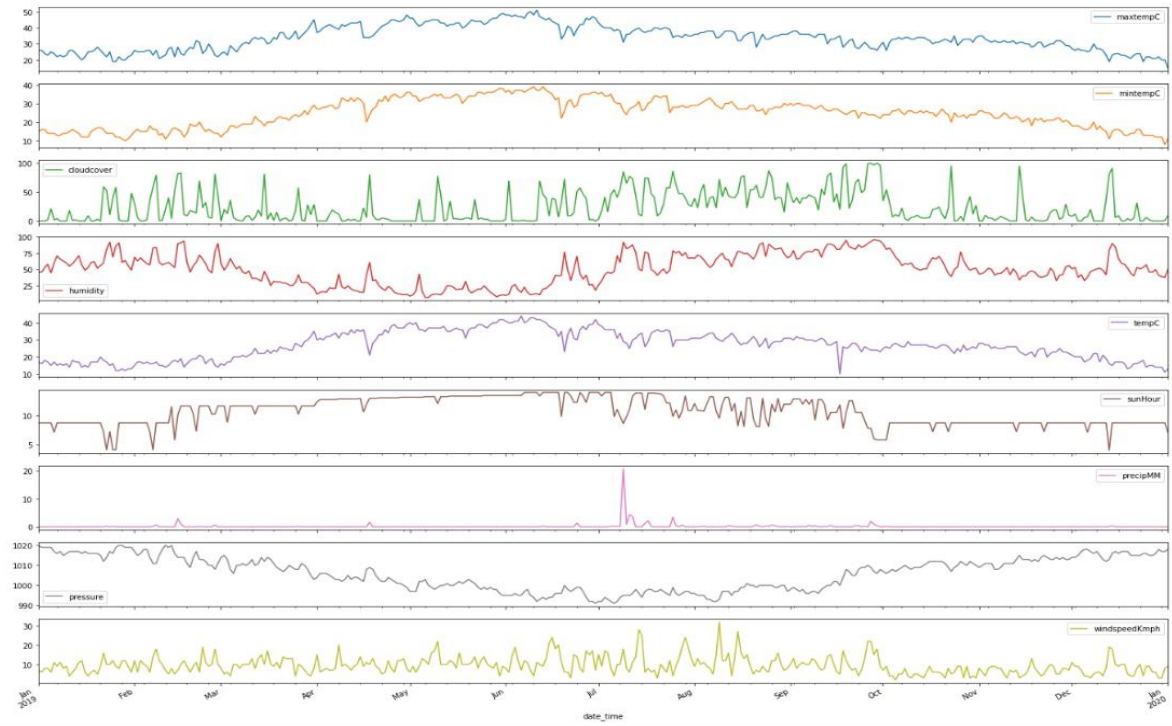

Figure 3.2: Plot for each factor for 1 year

4. RESULT & DISCUSSION

4.1 Result of Weather Forecasting

Weather forecasting has become a vital tool for societies around the world. Here are some key results:

- **Improved Accuracy:** Modern weather forecasts are significantly more accurate than ever before, thanks to advancements in data collection, computer modeling, and meteorological science.
- **Informed Decisions:** Accurate forecasts empower individuals and businesses to make informed decisions about daily activities, travel plans, agricultural practices, and many other areas.
- **Economic Benefits:** Weather forecasts contribute significantly to economic well-being by reducing risks and optimizing operations in various sectors.
- **Safety and Preparedness:** Weather forecasts play a crucial role in warnings and advisories for severe weather events, helping communities prepare and mitigate potential damage.

4.2 Discussion on Weather Forecasting

There are also ongoing discussions and areas for improvement in weather forecasting:

- **Accuracy Limitations:** While forecasts have improved, there are still limitations, especially for long-term predictions and highly localized events.
- **Climate Change Impact:** The complexity of climate change introduces new challenges for forecasting models, requiring ongoing research and adaptation.
- **Communication and Public Understanding:** Effectively communicating forecast uncertainties and probabilities to the public remains an important area of focus.

Overall, weather forecasting is a success story in scientific advancement with significant societal benefits. However, researchers and forecasters continue to strive for better accuracy, especially in the face of a changing climate.

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

4.3 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.

Table 4.1:

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024

date_time	Actual	Prediction	diff
2013-07-10 08:00:00	34	33.209030	0.790970
2015-11-04 20:00:00	25	25.275755	-0.275755
2015-09-21 09:00:00	34	31.975338	2.024662
2017-02-16 11:00:00	28	20.496727	7.503273
2012-07-21 01:00:00	28	28.401085	-0.401085
...	...	...	...
2019-03-30 09:00:00	37	33.187428	3.812572
2015-11-12 12:00:00	32	28.483724	3.516276
2019-12-31 05:00:00	8	15.177361	-7.177361
2019-08-02 17:00:00	35	35.363251	-0.363251
2019-10-22 08:00:00	26	27.890691	-1.890691

19287 rows × 3 columns

4.4 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.

Table 4.2:

date_time	Actual	Prediction	diff
2013-07-10 08:00:00	34	34.0	0.0
2015-11-04 20:00:00	25	25.0	0.0
2015-09-21 09:00:00	34	34.0	0.0
2017-02-16 11:00:00	28	28.0	0.0
2012-07-21 01:00:00	28	28.0	0.0
...	...	...	...
2019-03-30 09:00:00	37	39.0	-2.0
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	36.0	-1.0
2019-10-22 08:00:00	26	27.0	-1.0

19287 rows × 3 columns

4.5 Random Forest Regression:

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024

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.

Table 4.3:

date_time	Actual	Prediction	diff
2013-07-10 08:00:00	34	33.94	0.06
2015-11-04 20:00:00	25	24.43	0.57
2015-09-21 09:00:00	34	34.36	-0.36
2017-02-16 11:00:00	28	26.35	1.65
2012-07-21 01:00:00	28	28.17	-0.17
...	...	...	...
2019-03-30 09:00:00	37	32.99	4.01
2015-11-12 12:00:00	32	31.74	0.26
2019-12-31 05:00:00	8	10.62	-2.62
2019-08-02 17:00:00	35	35.72	-0.72
2019-10-22 08:00:00	26	26.85	-0.85

19287 rows × 3 columns

5. SYSTEM DEVELOPMENT

5.1 Analysis

The learning procedure starts with the perception of information, so examples can be discovered in information and prevalent choices can be taken later on which depend on the precedents gave. The principle point is to enable PCs to learn without human help or collaboration and modify their activities as needs be

The amount and size of malignant growth databases are expanding quickly, yet most are not dissected to discover covered up and profitable learning. Machine learning procedures can be utilized to find shrouded connections and examples. Models created utilizing machine learning systems enable specialists to settle on exact choices.

Accordingly, we utilize programmed learning strategies, for example, random forests, linear regression, polynomial regression, and so on to prepare our machine. The gadget adjusts to the predefined information record and gains from the predetermined parameters. From that point forward, machine learning strategies have turned out to be precise in a few fields before. Along these lines, the utilization of machine learning is helpful for the conclusion of malignant growth.

Collapsing neural systems work surprisingly better than linear regression, polynomial regression and Random Forests. This is on the grounds that, at every one of the dimensions, the weights proceed to return and attempt to diminish the mistake.

The most critical piece of our task is the examination of data when programmed learning methods are utilized. To dissect the pictures, we utilize a few descriptors, for example, nearby double examples, ORB, edge nearness measurements without parameters (PFTAS), GLCM. These element extractors help remove the usefulness of each picture. Subsequent to seeing these element vectors, we can at long

last train our machine in like manner. At last, this will enable us to get an exact determination of the forecast.

The most extreme exactness was accomplished when Parameter Free Threshold

Adjacency Statistics was utilized as the component extractor and SVM was utilized as the Machine Learning Algorithm. The best outcome was accomplished when parameters were tuned in like manner. With the end goal to tune parameters and gets quick outcomes, Grid Search technique was utilized. In network seek strategy, a scope of parameters is given to the classifier and the calculation at long last takes up the best blend of all the given parameters. The best arrangement of parameters is taken with the end goal that it gives the most extreme exactness.

5.2 System Design

The record has just been separated into train set and test set. Each information has just been labeled. First we take the trainset organizer.

We will train our model with the help of histograms. 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, Polynomial Regression, Random Forest and Neural Networks. With the help of our histogram, we will train our model. The most important thing to in this process is to tune thee parameters the 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 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 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 are confusion matrix, accuracy score, f1 score etc.

5.3 Model Development

Our strategy for model improvement is exploratory. The objective of our undertaking is to ensure the conclusion of malignancy with greatest accuracy. This must be accomplished by exploring different avenues regarding distinctive systems from a specific field. We have considered the programmed learning descriptors and algorithms.

The Machine Learning Algorithms that we are using are:

- Linear Regression
- Polynomial Regression
- Random Forest
- Neural Networks

Subsequently our point is to locate the best mix which will furnish us with greatest precision. Along these lines this task is absolutely test based. In addition parameter tuning is a noteworthy piece of any Machine Learning Algorithm. Regardless of whether the calculation works exceptionally solid in specific conditions, at that point too because of terrible determination of parameters, the precision could be low. In this manner we likewise need to center around the right arrangement of parameters. Hence parameter tuning must be done in whichever show we pick.

Parameter tuning should either be possible physically or by utilizing the lattice seek technique.

Network looking is the procedure in which information is checked with the end goal to discover ideal parameters for some random model. Contingent upon the kind of model that we are utilizing, tuning of specific parameters is vital. Framework seeking applies to a solitary model sort as well as number of models.

Network looking can be connected in machine learning with the end goal to ascertain the best parameters for its utilization in some random model. It very well may be computationally greatly costly and may set aside a long opportunity to keep running on the machine.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, June : 2024

Matrix Search constructs a model on every conceivable parameter mix. At that point it repeats through every parameter blend lastly stores a model for each mix.

5.4 PANDAS

Pandas is an open source, BSD-authorized library giving superior, simple to-utilize information structures and information investigation apparatuses for the Python programming language. Pandas has been heavily utilized in the development of this project. Given below are a few snapshots from the code.

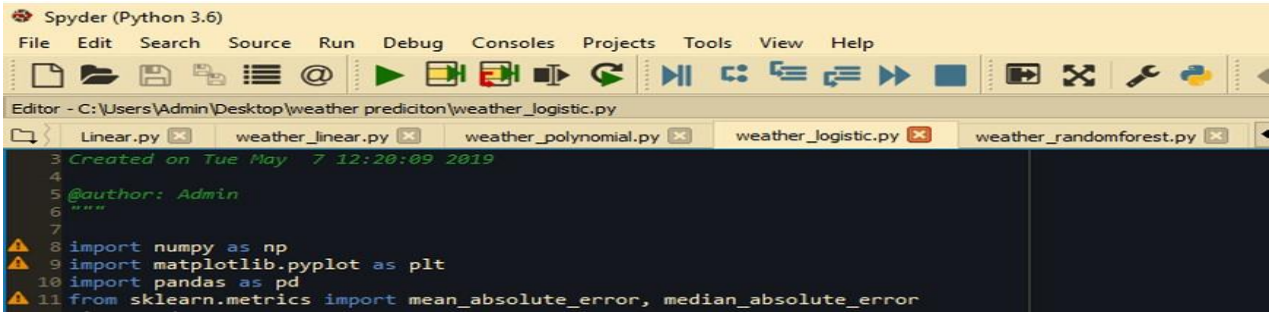

Figure 5.1: Pandas being imported to be used in the code

```
#import dataset
dataset = pd.read_csv('dataset.csv')
X = dataset.iloc[:,5:7].values
y = dataset.iloc[:,8].values
```

Figure 5.2: Using the Pandas object to import the dataset

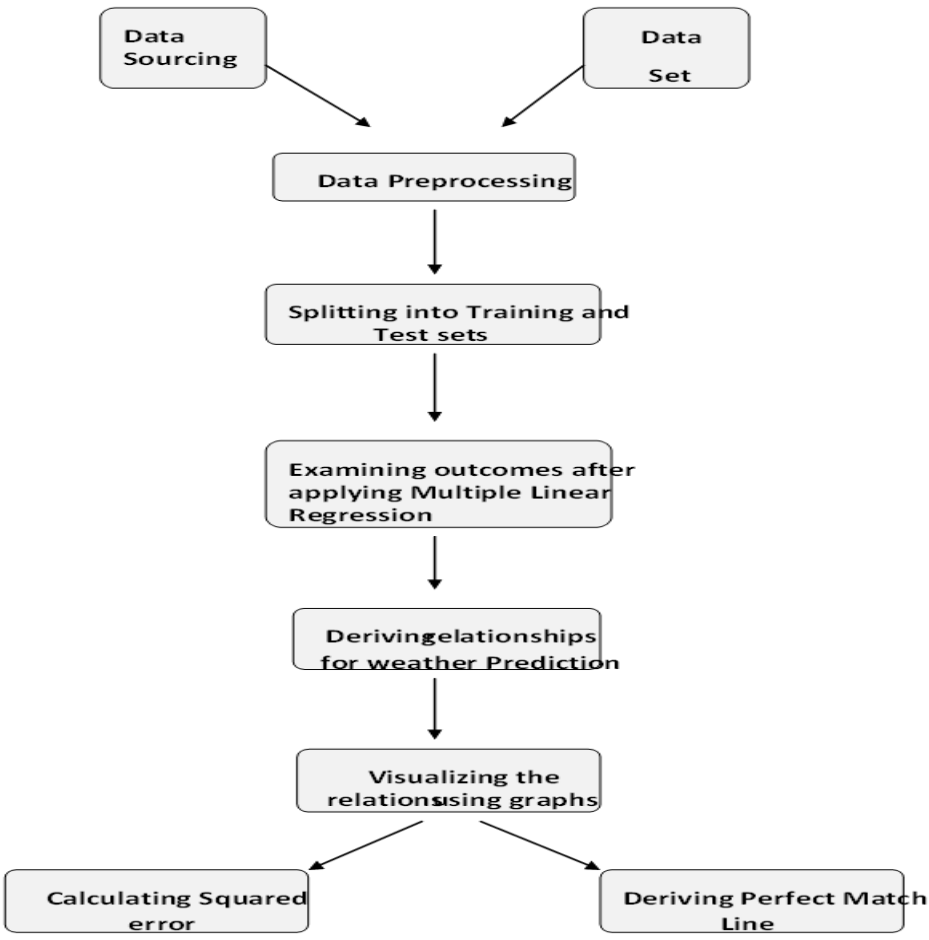

Fig. 5.1 Flow Chart

6. 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.

7. REFERENCES

[1] Mohammad Wahiduzzaman, Eric C. J. Oliver, Simon J Wotherspoon, Neil J. Holbrook, “A climatological model of North Indian Ocean tropical cyclone genesis, tracks and landfall”.

[2] Jinglin Du, Yayun Liu , Yanan Yu and Weilan Yan, “A Prediction of Precipitation Data Based on Support Vector Machine and Particle Swarm Optimization (PSO-SVM) Algorithms” [3] Prashant Kumar, Atul K. Varma, “ Atmospheric and Oceanic Sciences Group, EPSA, Space Applications Centre (ISRO), Ahmedabad, IndiaAssimilation of INSAT-3D hydro-estimator method retrieved rainfall for short-range weather prediction”

[4] Prashant Kumar, C. M. Kishtawal, P. K. Pal, “Impact of ECMWF, NCEP, and NCMRWF global model analysis on the WRF model forecast over Indian Region”

[5] H. Vathsala, Shashidhar G. KoolagudiPrediction, “Model for peninsular Indian summer monsoon rainfall using data mining and statistical approaches”

[6] Mark Holmstrom, Dylan Liu, Christopher Vo, “Machine Learning applied to weather forecasting”, Stanford University, 2016.

[7] Gyanesh Shrivastava, Sanjeev Karmakar, Manoj Kumar Kowar, “ Application of Artificial Neural Networks in Weather Forecasting: A Comprehensive Literature Review”, International Journal of Computer Application, 2012.

[8] Meera Narvekar, Priyanca Fargose, “Daily weather forcasting using Artificial Neural Network”, International Journal of Computer Application, 2015.