**Comparative Study of Weather Prediction Model using ML**

**ABSTRACT**

In the past, climate assessment has relied on a conventional approach of treating the environment like a liquid, which involves observing current wind conditions and using thermodynamics and numerical models to record the future state of the environment. However, this traditional method has its limitations due to the unstable nature of oscillating effects and uncertainties when estimating air conditions. As a result, climate forecasts have been limited to short-term periods of up to 10 days due to an insufficient understanding of environmental variations. Machine learning, on the other hand, is more resilient to most barometric destabilizing effects compared to traditional techniques. Additionally, machine learning does not rely on physical laws governing environmental processes, which is another advantage.

**Background**

In response to the current situation, the Indian Astronomical Observatory is preparing traditional weather forecasts. There are four general methods of weather forecasting. The first method is the climatological method, which looks at meteorological statistics collected over several years and calculates an average. The second method is the method of analogy, which involves finding that the weather forecast for a certain day in the past is similar to the current weather forecast. The third method is the persistence and trend method, which lacks the ability to predict the weather because it relies on past trends.

The fourth method is numerical weather forecasting, which predicts the weather based on various conditions such as atmospheric temperature, wind speed, high and low pressure systems, precipitation and snowfall. Therefore, these traditional methods have many limitations. Not only does it predict the temperature for the current month in full, but it does so without using machine learning algorithms. So my project is to improve the accuracy and forecast the weather at least one month in the future using machine learning techniques

**Objective**

The aim of this project is to predict temperature using different algorithms such as linear regression, random forest regression and Decision tree regression. The output value should be based on several additional factors such as maximum temperature, minimum temperature, cloud cover, humidity and hours of the day, rainfall, pressure and wind speed.

1. **Introduction**

Weather forecasting involves forecasting the future weather and atmosphere of a given region. This was done using physics equations in the early days when the atmosphere was considered a liquid. When we look at the current state of the environment and numerically solve these equations to predict the future state, we cannot determine a very precise time beyond 10 days that can be improved by science and technology.

Machine learning can be used to process instant comparisons of historical weather forecasts and observational data. Using machine learning, weather models can better account for forecast inaccuracies, such as overestimating precipitation, and produce more accurate forecasts. Temperature prediction is important in many applications, including climate-related research, energy, agriculture, medicine, and more.

There are many types of machine learning calculations, including linear regression, polynomial regression, random forest regression, artificial neural networks, and recurrent neural networks. These models are prepared based on real information from any region. Contributions to these models are given, such as whether 2-day temperature is expected, minimum temperature, average air mass, maximum temperature, average humidity, and order. With that in mind, there will be 7 days of low and high temperatures.

**Machine Learning**

Machine learning is relatively fault-tolerant and does not require any additional physical variables for prediction. Therefore, machine learning is a better opportunity in development weather forecast. Weather forecasting used to be a complex problem until technology improved. Meteorologists rely on satellites and the atmospheric conditions in them the data model is less accurate. Over the past 40 years, weather forecasting and analysis have improved dramatically in accuracy and predictability using the Internet of Things. Thanks to advances in data science and artificial intelligence, scientists can now make extremely accurate and predictable weather predictions.

**Use of Algorithms:**

There are different ways to provide temperature using regression and different functions regression where a data set is used to calculate a score and surveys. For training, the calculation used 80% size and other 20% size as a test set. For example, if we are to use these machine learning calculations to make predictions temperature in Kanpur, India, we will use 8 years information to prepare the calculation and 2 years of information as a test data set. Unlike weather forecast that uses machine learning algorithms, relies heavily on reconstruction physics and differential equations, artificial intelligence is also used to ensure temperature: including models such as, stochastic decay, stochastic decay, stochastic decay. The forest is coming back. Achieve machine learning has radically changed high accuracy weather forecast and a world view of the forecast. Plus over the next few years will use these advances to further advance the accurate prediction of the climate to be prevented disasters such as typhoons, tornadoes and thunderstorms.

1. **Methodology**

The dataset used in this layout is collected from Kaggle and is “Historical Weather Data for Indian Cities' from which we have selected the data for 'Kanpur City'. The datasets are created keeping in mind the need for such historical weather data in the community. Dataset of top 8 Indian cities by population. Dataset was used using the worldweatheronline.com API and the wwo\_hist package. The datasets contain hourly weather data from 01.01.2009 to 01.01.2020. Data for in each city it is over 10 years old. This data can be used to visualize data changes due to global warming, or can be used to predict the weather for the next few days, weeks, months, seasons, etc.

Note: Data extracted using worldweatheronline. com API, we can't guarantee the correctness of the data.

The main purpose of this dataset is to forecast the weather for the next day or next week. The dataset provides a large amount of data. In addition, this data can be used for visualizations to help understand the impact of global warming on various aspects weather such as rainfall, humidity, temperature and more.

In this project we focus on forecasting temperature in Kanpur city using different machine learning algorithms and various regressions. Using multiple regression, we forecast the temperature on the historical weather data set for Kanpur city because we will first use multiple linear regression, decision tree regression, then we use random forest regression.



Table 2.1: Historical Weather Dataset of Kanpur City

A screenshot of a computer

Description automatically generated with low confidence

Figure 2.1: Plot for each factor for 10 years

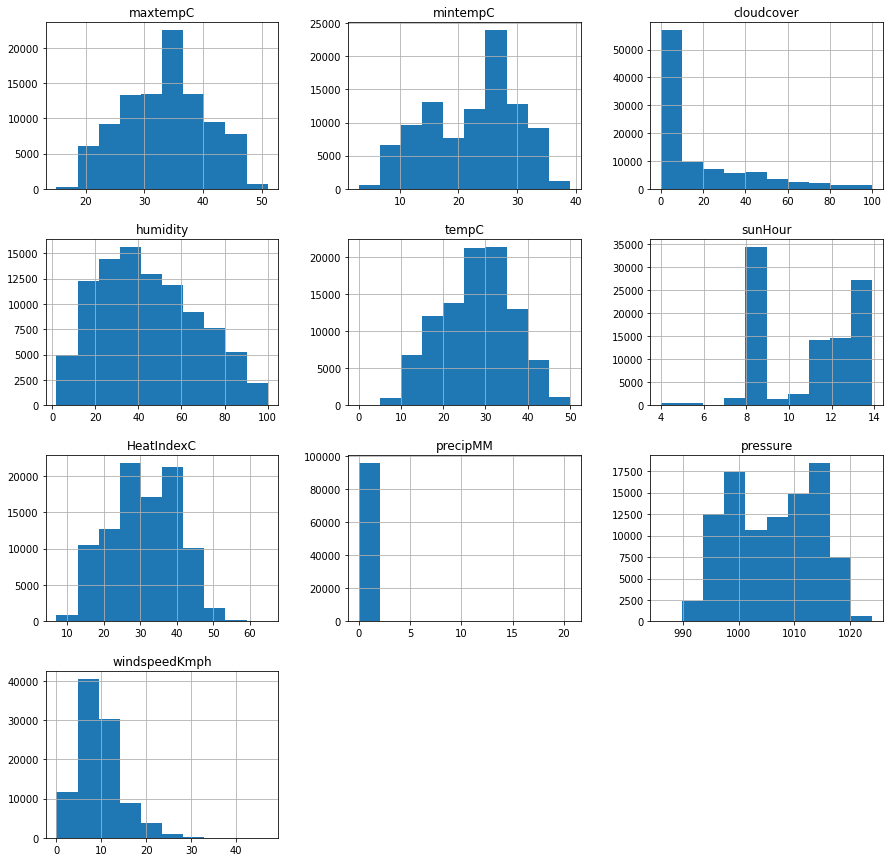
Background pattern

Description automatically generated

Figure 2.2: Plot for each factor for 1 year

1. **Experimentation**

The recordings have just been split into practice and test sets. Each information has just been labelled. First we take the trainset organizer. We will use histograms and graphs for training our model. The characteristics extracted in this way are stored in a histogram. The process is being executed for each data item in the training set. Now we will create a classifier model. Classifiers that we will consider are linear regression, decision tree regression and Random Forest regression. We use our histogram to train our model. Most important aspect in this process is to fine tune these parameters accordingly so that we can get the most accurate results. After training is complete, we will take the test set. Now for each data variable in the test set, we will extract features using feature extraction techniques and then compare their values ​​with the values ​​present in the created histogram from learning from the test set. The output is then predicted for each test day. Now for the calculation accuracy, we compare predicted and labelled values. Various metrics for our project confusion matrix, R2 scores, etc.



**References**

1. Kaggle – Dataset for Kanpur City (<https://www.kaggle.com/datasets/hiteshsoneji/historical-weather-data-for-indian-cities?select=kanpur.csv>)
2. Numpy Library (<https://numpy.org/>)
3. Pandas Library (<https://pandas.pydata.org/>)
4. Advances in Weather Prediction. (<https://www.science.org/doi/abs/10.1126/science.aav7274>)