WEATHER FORECASTING USING MACHINE LEARNING

Chimma Dharma
Student
Department of Computer Science and Engineering
Narasaraopet, India
Dharma.chimma123@gmail.com

Sk. Allabakshu
Student
Department of Computer Science and Engineering
Narasaraopet, India
Allabaashu37470@gmail.com

P. Siva Sankar
Student
Department of Computer Science and Engineering
Narasaraopet, India
Sivasankarpusunuri2@gmail.com

U. Sai Anusha
Asst.Professor
Department of Computer Science and Engineering
Narasaraopet, India
uppalasaianusha@gmail.com

Abstract: - Weather forecasting is the process of predicting the upcoming state of the atmosphere by analysing current weather conditions and using mathematical models to simulate future weather patterns. Weather forecasting using machine learning involves training models that can automatically learn and analyze large amounts of weather data to predict future weather patterns. This approach involves using a variety of machine learning models, including neural networks, decision tree, and support vector machines, to analyze historical weather data and generate forecasts. The process involves collecting vast amounts of data from various sources, including weather stations, satellites, and other sensors, to create a comprehensive data set for analysis. Machine learning models can improve the accuracy of weather forecasting by detecting patterns in data that may not be apparent to humans. The use of machine learning in weather forecasting is still in its initial stages, but it has the possible to improve weather forecasts significantly in the future.

Keywords:-Weather Forecasting, Prediction, Weather Data, Data Analysis, Accuracy, Decision tree, Multiple Linear Regression, flask.

I. INTRODUCTION

Weather forecasting using machine learning is an emerging field that involves the use of advanced algorithms to analyze past weather data and make predictions about future weather patterns. Machine learning techniques, such as artificial neural networks, decision trees. can be used to detect complex patterns in large data sets, allowing for more accurate and reliable weather forecasts.

Primary benefits of using Machine learning in weather forecasting is that it can benefit improve forecast accuracy, particularly for short-term predictions. Machine learning models can analyze large volumes of data from various sources, including weather stations, satellites, and other sensors, to identify patterns that may not be apparent to humans. This enables them to generate more accurate forecasts and alerts for severe weather events.

Another benefit of using machine learning in weather forecasting is that it can help automate the process of generating weather forecasts, reducing the time and resources required to produce them. This can help make weather forecasts more accessible and affordable, particularly in developing countries where resources may be limited.

Despite these benefits, weather forecasting using machine learning is still in its initial stages, and there are some tasks that need to be overcome. For example, ensuring the quality and accuracy data inputs is essential, and machine learning models need to be continually updated and refined to keep up with changes in the atmosphere.

Overall, weather forecasting using machine learning holds excessive promise for improving weather prediction accuracy and making weather forecasts more accessible to people worldwide. Applications of Machine learning in agriculture is the development of crop recommendation systems.

Weather forecasting is the procedure of forecasting the future atmosphere by analyzing current weather conditions and using mathematical models to simulate future weather patterns.

Weather forecasting is a crucial aspect of modern life, as it helps individuals, businesses, and governments plan and prepare for weather-related events. Weather forecasting involves forecasting the future state of the atmosphere, including changes in temperature, humidity, wind speed, precipitation, and other atmospheric conditions.

Weather forecasts are used to warn people of impending storms, scorchers, or other severe weather events, allowing them to take necessary precautions to protect themselves and their property. They are also used by farmers to plan their planting and harvesting, and by airlines to make decisions about flight routes and schedules.

Weather forecasting has improved significantly over the years with advances in technology, particularly the use of satellites, radar systems, and computer models. In recent years the use of Machine learning and artificial intelligence techniques has also been applied to weather forecasting, leading to further improvements in accuracy and reliability.

II. LITERATURE SURVEY

A literature survey aims to identify the latest trends, advancements, and research gaps in a particular field of study. It helps researchers to get an overview of the current state of research in the area and to identify the most important and relevant research papers and studies. Here is a brief literature review on weather forecasting:

1 “A Review of Machine Learning Approaches in Weather Forecasting" by O. S. Ayodeji and A. O. Awodele. This article provides a inclusive review of Machine learning techniques used in weather forecasting, including artificial neural networks, decision trees.

2 "Advances in Weather Forecasting: A Review of Machine Learning Techniques" by T. Balasubramanian et al. This paper explores the latest advances in weather forecasting using machine learning techniques, including deep learning models and ensemble learning algorithms.

3 "A Review of Statistical Weather Forecasting Methods" by S. H. Lee et al. This article provides an overview of statistical forecasting methods used in weather prediction, including autoregressive combined touching average (ARIMA) models, linear regression, multiple linear regression.

4 "Advances in Numerical Weather Prediction" by E. Kalyan. This paper discusses the history and development of numerical weather prediction models, which use mathematical equations to simulate weather patterns and make predictions about future conditions.

5 "Challenges and Opportunities in Advancing Mesoscale Weather Forecasting" by M. Xue et al. This article explores the challenges and opportunities for improving mesoscale weather forecasting, which focuses on predicting weather patterns at a smaller scale than global weather models.

Overall, these studies demonstrate the significant advancements made in weather forecasting and the potential for further improvements using Machine learning techniques and other approaches.

III. EXISTING SYSTEM

The existing system of weather forecasting typically involves a combination of observational data collection, computer modelling, and meteorological analysis. Here are the main steps involved:

Data collection: Weather data is collected from various sources, such as satellites, weather stations, radars, and buoys. These sources provide information on temperature, humidity, pressure, wind speed, and other meteorological parameters.

Data analysis: The collected data is processed and analysed using mathematical models and statistical methods to create a comprehensive picture of the current weather conditions.

Numerical weather prediction: Numerical weather models are used to simulate the behavior of the atmosphere based on the collected data. These models use complex mathematical equations to calculate the future weather conditions.

Weather forecasting: Based on the data analysis and numerical modeling, meteorologists make weather forecasts for different timeframes, from a few hours to several days or even weeks in advance.

Dissemination of information: The weather forecasts are communicated to the public through various media channels, such as TV, radio, newspapers, and internet websites, as well as through weather apps and social media.

The existing system of weather forecasting is constantly evolving, with new technologies and methods being developed to improve the accuracy and timeliness of weather predictions.

IV. METHODOLOGY

We have used different Machine Learning algorithms to predict the temperature by using mintemp, maxtemp, humidity, sun hours, precipitation, cloud cover features Different evaluation metrics like error rate, accuracy are used to compare the four models. The methodology for weather forecasting using machine learning typically involves the following steps:

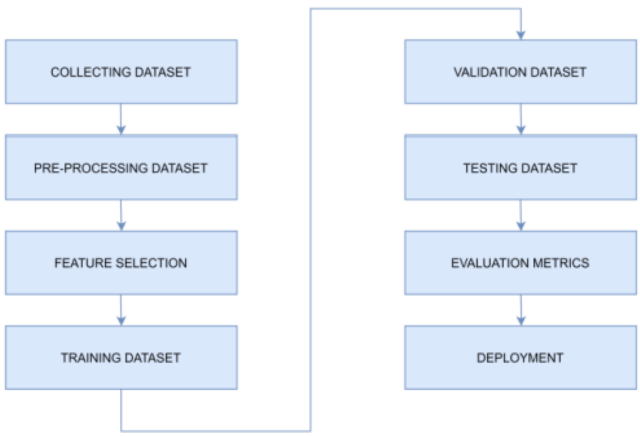


Fig. 1: Proposed system workflow

- Data Collection:

The first step is to gather relevant weather data from various sources, such as weather stations, satellites, and sensors. This data may include historical weather data, current weather conditions, and weather forecasts. Here we collect data from Kaggle Historic weather data of Indian cites from that we choose one of the city weather details. here almost 6K records are present and 19 features from that we only choose 9 features to predict the temperature.

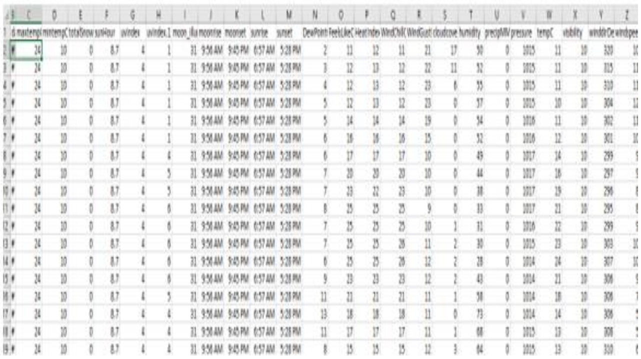
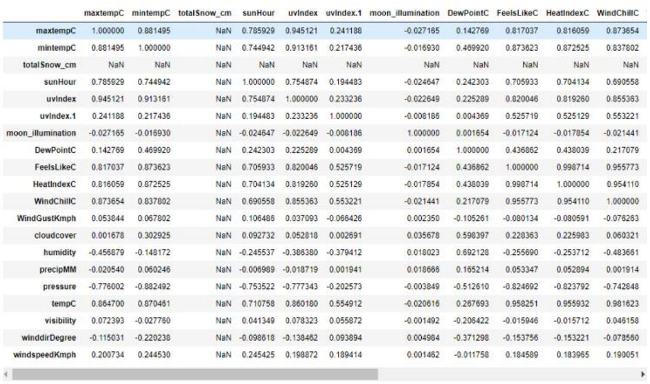


Fig. 2: Dataset diagram

- Data Cleaning and Feature Extraction

Data Preprocessing: Once the data is collected, it needs to be preprocessed to remove any noise or outliers and to normalize the data. This step is essential for ensuring the accuracy of the machine learning model. Fig 3 shows the correlation between the features of the dataset.



	maxtempC	mintempC	totalSnow_cm	sunHour	uvIndex	uvIndex.1	moon_illumination	DewPointC	FeelsLikeC	HeatIndexC	WindChillC
maxtempC	1.000000	0.881495	NaN	0.785929	0.945121	0.241188	-0.027165	0.142769	0.817037	0.816059	0.873654
mintempC	0.881495	1.000000	NaN	0.744942	0.913161	0.217436	-0.016930	0.469920	0.873623	0.872525	0.837802
totalSnow_cm	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
sunHour	0.785929	0.744942	NaN	1.000000	0.754874	0.194483	-0.024647	0.242303	0.705933	0.704134	0.690558
uvIndex	0.945121	0.913161	NaN	0.754874	1.000000	0.233236	-0.022649	0.225289	0.820046	0.819260	0.855363
uvIndex.1	0.241188	0.217436	NaN	0.194483	0.233236	1.000000	-0.008186	0.004369	0.525719	0.525129	0.553221
moon_illumination	-0.027165	-0.016930	NaN	-0.024647	-0.022649	-0.008186	1.000000	0.001654	-0.017124	-0.017854	-0.021441
DewPointC	0.142769	0.469920	NaN	0.242303	0.225289	0.004369	0.001654	1.000000	0.436862	0.438039	0.217079
FeelsLikeC	0.817037	0.873623	NaN	0.705933	0.820046	0.525719	-0.017124	0.436862	1.000000	0.998714	0.955773
HeatIndexC	0.816059	0.872525	NaN	0.704134	0.819260	0.525129	-0.017854	0.438039	0.998714	1.000000	0.954110
WindChillC	0.873654	0.837802	NaN	0.690558	0.855363	0.553221	-0.021441	0.217079	0.955773	0.954110	1.000000
WindGustKmph	0.053844	0.067802	NaN	0.106486	0.037093	-0.066426	0.002350	-0.105261	-0.080134	-0.080591	-0.076263
cloudcover	0.001678	0.302925	NaN	0.092732	0.052818	0.002691	0.035678	0.598397	0.228363	0.225983	0.060321
humidity	-0.456879	-0.148172	NaN	-0.245537	-0.386380	-0.379412	0.018023	0.692128	-0.255690	-0.253712	-0.483661
precipMM	-0.020540	0.060246	NaN	-0.006989	-0.018719	0.001941	0.018666	0.165214	0.053347	0.052894	0.001914
pressure	-0.776002	-0.882492	NaN	-0.753522	-0.777343	-0.202573	-0.003849	-0.512610	-0.824692	-0.823792	-0.742848
tempC	0.864700	0.870461	NaN	0.710758	0.860180	0.554912	-0.020616	0.267693	0.958251	0.955932	0.981623
visibility	0.072393	-0.027760	NaN	0.041349	0.078323	0.055872	-0.001492	-0.206422	-0.015946	-0.015712	0.046158
winddirDegree	-0.115031	-0.220238	NaN	-0.098618	-0.138462	0.093894	0.004984	-0.371298	-0.153756	-0.153221	-0.078560
windspeedKmph	0.200734	0.244530	NaN	0.245425	0.198872	0.189414	0.001462	-0.011758	0.184589	0.183965	0.190051

Fig.3 Correlation

- Feature Extraction:

In this step, the relevant features from the data are extracted, such as temperature, humidity, wind speed, and pressure. These features will be taken as participations for the Machine learning algorithm.

- Model Training:

The next stage is to train the machine learning model using the preprocessed data. Different algorithms can be used, such as artificial neural networks, decision trees. The model is trained to identify patterns in the data and make predictions about future weather conditions.

- Model Evaluation:

Once the model is trained, it needs to be evaluated to assess its accuracy and performance. This is typically done using a separate test dataset that was not used during training.

- Model Deployment:

Lastly, the model is deployed in a making situation, where it can be used to generate weather forecasts in real-time. The model can be continually updated and refined based on new data and feedback.

Overall, the methodology for weather forecasting using machine learning involves a combination of data collection, feature extraction, preprocessing, model training, evaluation, and deployment. These steps are essential for developing accurate and reliable weather prediction systems.

1. Random Forest: Random forest is a popular machine learning algorithm that is widely used for classification and regression tasks. It is an ensemble learning method that combines multiple decision trees to produce more accurate and robust predictions.

Random forest is a versatile and effective algorithm that can be applied to a wide range of machine learning tasks, such as image classification, text classification, and regression analysis. It is particularly useful when dealing with high-dimensional and noisy data, and when the relationships between the input variables and the output variable are complex and nonlinear.

2. Decision Tree: Decision trees are highly interpretable, easy to understand, and can handle both categorical and numerical data. They are also computationally effective and can holder high-dimensional data. But decision trees can be disposed to overfitting, particularly if the tree is too deep or if the dataset is noisy. Techniques such as pruning, regularization, and ensemble methods such as random forests can help to overcome these limitations.

3. Linear Regression: Linear regression can also be used to model the relationship between multiple predictor variables and a target variable, such as temperature or precipitation. In this case, we would collect a dataset of weather observations that includes information on the predictor variables (such as location, time, atmospheric pressure, humidity, wind speed, and precipitation) and the target variable (such as temperature or precipitation). We would then use linear regression to learn the coefficients of the linear equation that best fits the data and predicts the target variable based on the predictor variables.

Linear regression has limitations in weather forecasting, as it assumes a linear relationship between the predictor variables and the target variable, which may not always hold in complex weather systems. However, it can still be a useful tool for simple weather forecasting tasks, especially when combined with other machine learning algorithms and meteorological models.

V. RESULT AND ANALYSIS

After created all the models by using the training data , we test the models by using the test data and calculates the accuracies of the models. The following table shows the accuracies of different models:

Algorithms	Accuracy
Linear Regression	94.15
Decision tree	96.52
Random forest	93.21

Fig.4 Accuracy Table

From the above table we selected Decision tree model as our final model because it got more accuracy compared to the other models.

VI. DEPLOYMENT

The final model is deployed into an application by using Flask module of python by creating the user interface. The application simply takes the input and provides output.

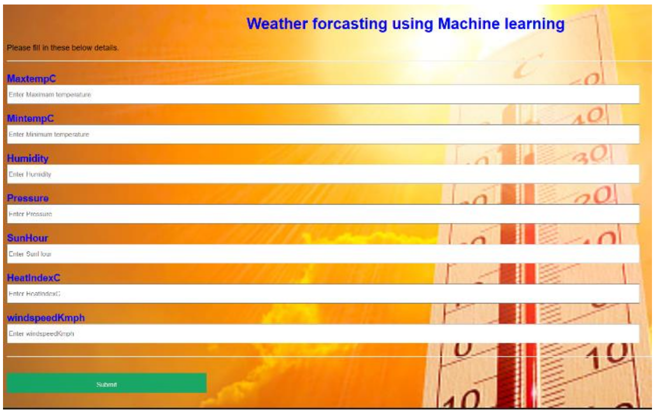


Fig.5 User interface

VII. CONCLUSION

Machine learning algorithms can identify complex patterns in weather data and generate forecasts that are more precise and localized. This technology is being used by various organizations, including government agencies, private companies, and research institutions, to develop more advanced weather prediction systems.

However, there are still challenges associated with Weather forecasting using machine learning, such as data quality, model complexity, and interpretability. It is crucial to ensure that the data used to train machine learning models is of high quality and representative of the weather patterns being predicted. Additionally, the models used must be interpretable, so that meteorologists can understand how they are making predictions and adjust them accordingly.

VIII. REFERENCES

[1] M. Ahmed, “A data fusion framework for real-time risk assessment on freeways,” 2012.

[2] AHMJakaria, “Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee,” 2018.

[3] O. Y. Al-Jarrah, “Efficient Machine Learning for Big Data: A Review,” 2015.

[4] M. S. S. S. A. S. A. J. Amit Kumar Agarwal, “Forecasting using Machine Learning,” 2019.

[5] N. D. Hoai, “Downscaling Global Weather Forecast Outputs Using ANN for Flood Prediction,” 2011.

[6] C. Choi, “Development of Heavy Rain Damage Prediction Model Using Machine Learning Based on Big Data [Changhyun Choi] [2018],” 2018.

[7] S. Madan, “Analysis of Weather Prediction using Machine Learning & Big Data,” p. 2018.

[8] C. Feng, “Hourly-Similarity Based Solar Forecasting Using Multi-Model Machine Learning Blending,” 2018.

[9] GylianVerstraete, “A data-driven sframework for predicting weather impact on high-volume low-margin retail products,” 2018.

.