A

Mini-Project Report

on

WEATHER REPORTING SYSTEM USING MACHINE LEARNING

Submitted as partial fulfilment for the award of

BACHELOR OF TECHNOLOGY DEGREE

Session 2021-22

in

Computer Science & Engineering

Submitted By:

ABHAY SRIVASTAVA – 2000971540001

SUYASH RAJPUT - 2000971540056

SANSKRITI SHARMA - 2000971540048

SAKSHI TAYAL - 2000970310145

Under the guidance of:

Mr. Ravindra Kumar

Assistant professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GALGOTIAS COLLEGE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

AFFILIATED TO
DR. A.P.J ABDUL KALAM TECHNICAL UNIVERSITY, LUCKNOW, UP
(Formerly UPTU)

CERTIFICATE

This is to certify that the mini-project report entitled "**WEATHER REPORTING SYSTEM USING MACHINE LEARNING**" submitted by Mr. Abhay Srivastava **2000971540001** Mr. Suyash Rajput **2000971540056,** Ms. Sanskriti Sharma **2000971540048 ,** Ms. Sakshi Tayal **2000970310145** to the Galgotias College of Engineering & Technology, Greater Noida, Utter Pradesh, affiliated to Dr. A.P.J. Abdul Kalam Technical University Lucknow, Uttar Pradesh in partial fulfillment for the award of Degree of Bachelor of Technology in Computer science & Engineering is a bonafide record of the project work carried out by them under my supervision during the year 2021-2022.

Mr. Ravindra Kumar
(Project Coordinator)
Assistant professor
Dept. of CSE

Dr. Vishnu Sharma
Professor and Head
Dept. of CSE

1. ACKNOWLEDGEMENT

We have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. We would like to extend my sincere thanks to all of them.

We are highly indebted to **Mr. Ravindra Kumar** for his guidance and constant supervision. Also, we are highly thankful to them for providing necessary information regarding the project & also for their support in completing the project.

We are extremely indebted to Dr. Vishnu Sharma, HOD, Department of Computer Science and Engineering, GCET and **Mr. Ravindra Kumar** Project Coordinator, Department of Computer Science and Engineering, GCET for their valuable suggestions and constant support throughout my project tenure. We would also like to express our sincere thanks to all faculty and staff members of Department of Computer Science and Engineering, GCET for their support in completing this project on time.

We also express gratitude towards our parents for their kind co-operation and encouragement which helped me in completion of this project. Our thanks and appreciations also go to our friends in developing the project and all the people who have willingly helped me out with their abilities.

ABHAY SRIVASTAVA
SUYASH RAJPUT
SANSKRITI SHARMA
SAKSHI TAYAL

2. <u>ABSTRACT</u>

The weather is disordered in nature and hence has always been difficult for meteorologists in forecasting the weather accurately. Different methods and new models are being updated to keep pace with the everchanging nature of weather. Weather is measured from days to months where it includes the following components such as: Precipitation, Temperature, Sunshine, Wind, Lightning, Direction of Wind, Humidity, Cloud Cover, Speed of Wind, Snow, Pressure etc. The nature of the Weather tends to change every second on earth and has a great influence even if there are some small changes that occurs at any point of time on the surface. Meteorologists have done researches in weather prediction using various mathematical models. This project aims to obtain the highest accuracy to predict weather using Random Forest algorithm.

TABLE OF CONTENTS

3. <u>INTRODUCTION</u>

Weather forecasting has been a standout amongst the most experimentally and technologically troublesome issues over the world in the most recent century . Environmental change has been looking for a great deal of consideration since a long time because of the sudden changes that happen. There are several limitations in better execution of weather forecasting thus it ends up hard predicting weather here and now with effectiveness. Weather forecasting assumes a significant role in meteorology .To makes an exact prediction is one of the significant troubles standing up to meteorologist wherever all through the world. Weather warnings are vital in light of the fact that they are utilized to ensure life and property. Forecasts dependent on Temperature, Outlook, Humidity and Wind are important to farming, and along these lines to traders inside product markets. Temperature forecasts are utilized by utility companies to assess request over coming days. Since outdoor activities are seriously reduced by substantial rain, snow and wind chill, estimates can be utilized to design activity around these occasions, and to prepare and survive them. Without precise weather forecasts individuals may end up in hazardous circumstances as they were unprepared for and end up harmed or worse.

The difficulties of weather forecasting, among others, are learning weather representation utilizing an enormous volume of weather dataset. For this purpose, analysis of different data mining procedure is performed. Data mining techniques enables users to analyse data from a wide range of dimensions or angles, classify it, and condense the connections recognized. Some fundamental terms related to Data Mining are: Classification, Learning and Prediction. Classification is a data mining (machine learning) method used to predict aggregate participation for information cases. For instance, classification can be utilized to predict whether the weather on a specific day will be "sunny", "rainy" or "cloudy" .Learning refers to training and mapping contribution to yield information. It tends to be performed in two different ways: Supervised and Unsupervised learning. A supervised learning algorithm analyses the training data and produces a derived capacity utilizing Classifier . In machine learning, unsupervised learning alludes to the issue of trying to hidden structure in unlabelled information . Since the precedents given to the learner are unlabelled, there is no mistake or reward signal to assess a potential solution. This recognizes unsupervised learning from supervised learning. Prediction identifies with modelling and the logical relationship of the model sooner or later. Finding patterns and data may prompt sensible predictions .

We have used Random Forest Algorithm for weather forecasting. In this system, state of weather is classified in some attribute like as Humidity, Pressure, Wind speed and dew point. Using those attribute the system will predict the temperature

4. <u>LITERATURE REVIEW</u>

There are many research papers that have been published related to predicting the weather. A paper was published on 'The Weather Forecast Using Data Mining Research Based on Cloud Computing' This paper proposes a modern method to develop a service oriented architecture for the weather information systems which forecast weather using these data mining techniques. This can be carried out by using Artificial Neural Network and Decision tree Algorithms and meteorological data collected in Specific time. Algorithm has presented the best results to generate classification rules for the mean weather variables. The results showed that these data mining techniques can be enough for weather forecasting. Another paper was published on 'Analysis on The Weather Forecasting and Techniques' where they decided that artificial neural network and concept of fuzzy logic provides a best solution and prediction comparatively. They decided to take temperature, humidity, pressure, wind and various other attributes into consideration .

Another research paper titled 'Issues with weather prediction' discussed the major problems with weather prediction. Even the simplest weather prediction is not perfect. The one-day forecast typically falls within two degrees of the actual temperature. Although this accuracy isn't bad, as predictions are made for further in time. For example, in a place like New England where temperatures have a great variance the temperature prediction are more inaccurate than a place like the tropics. Another research paper titled 'Current weather prediction' used numerical methods to stimulate what is most likely going to happen based on known state of the atmosphere .

5.1   PROBLEM FORMULATION

Weather warnings are important forecasts **because they are used to protect life and property**. Since outdoor activities are severely curtailed by heavy rain, snow and wind chill, forecasts can be used to plan activities around these events, and to plan ahead and survive them.

Weather Forecasting is important since **it helps determine future climate expectations**. Through the use of latitude, one can determine the likelihood of snow and hail reaching the surface. You can also be able to identify the thermal energy from the sun that is accessible to a region.

System is trained with the historic weather data of cities. When city is entered by the user the system will display the current weather report along with the temperature predictions of the next 48 hours on the hourly basis. User can easily find out Weather condition by using this system. The primary advantage of forecasting is that it provides the business with valuable information that the business can use to make decisions about the future of the organization Disadvantages are Weather forecast by the system is not very accurate. Previous data is required by the system to forecast weather.

The goal of weather prediction is to provide information people and organizations can **use to reduce weather-related losses and enhance societal benefits**, including protection of life and property, public health and safety, and support of economic prosperity and quality of life. Farmers can predict when to plant or harvest their crop. People can choose where and when to take their holidays to take advantages of good weather. Surfers known when large waves are expected. Regions can be evacuated if hurricanes or floods are expected. Aircraft and shipping rely heavily on accurate weather forecasting.

5.2   REQUIREMENTS

Software Requirement

The software used in our projects are:

- Python 3.7: Python is an interpreted, high level, general programming language. Its formatting is visually uncluttered, and it often uses English keywords where other languages use punctuation. It provides a vast library for data mining and predictions.
- Jupiter Notebook/ Spyder/ VsCODE: It is an open source cross-platform integrated development environment (IDE) for scientific programming in the Python language. Spyder integrates with a number of prominent packages as well as another open source software.
- Numpy: Numpy was used for building the front-end part of the system.
- Pandas: Pandas was used for the data preprocessing and statistical analysis of data.
- Matplotlib: Matplotlib was used for the graphical representation of our prediction.

Functional Requirements

- The system must provide the predicted weather.
- The system must have an easy to use interface for using the system for all the users.
- The Admin must be able to update/modify the Dataset.
- The Dataset of the weather must be available for the system.

5.3 PROPOSED WORK AND SYSTEM DESIGN

In the past, the human forecasters were responsible for generating the entire weather forecast based upon available observation. Today, human input is generally confined to choosing a model based on various parameters, such as model biases and performance. Using a consensus of forecast models, as well as ensemble members of the various models, can help reduce forecast error.

The chosen weather data is divided into two groups, the training group, corresponding to 80% of the data, and the test group corresponding to 20% of data. Evaluation is done through test data and model score().

The use of ML will lead to significant progress over the whole range of "classical" modelling, data assimilation and post-processing algorithms in the coming decade. It can be applied at each step of the weather forecasting process. Advantages include the following:

 • ML can improve the utilization of observing systems, allowing forward planning of maintenance activities, improved quality control and intelligent filling of data gaps

 • ML can improve the ability to input information from complex observing systems into models' data assimilation by providing better compression rates, which will significantly increase the precision and speed of forward operators

 • ML can help model and estimate observation errors in data assimilation.
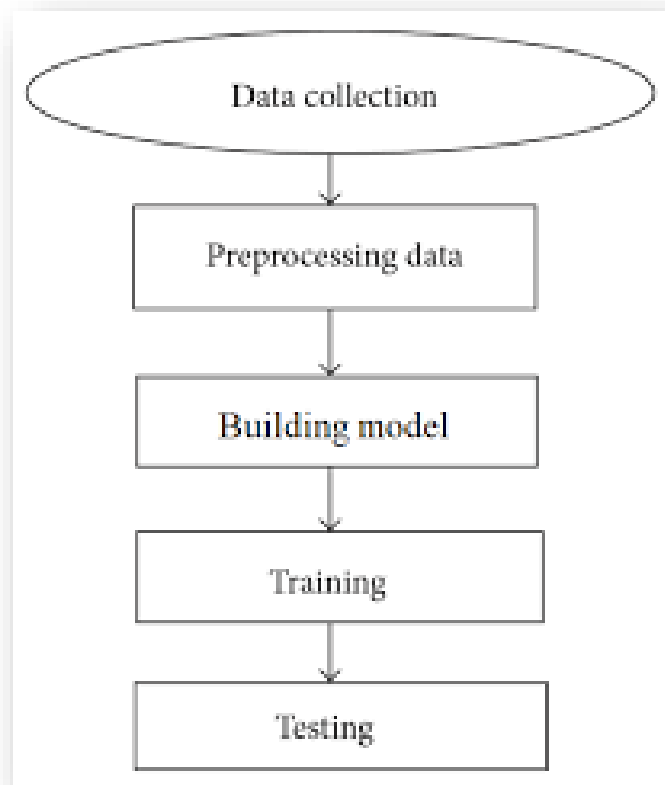
SYSTEM DESIGN

This weather reporting system is designed in a very user-friendly interface where user is given the following three options:-

1. **Current weather report**
2. **Forecast next 48 hours weather**
3. **Complete weather report**
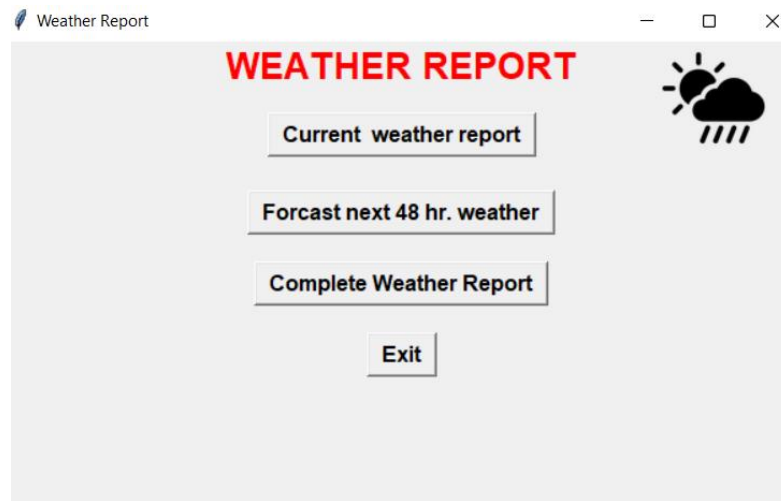
After selecting any one the above three options user is asked to choose the city from the combo box and press the submit button. Once the submit button is pressed corresponding output will be displayed on the interface.

Data flow diagram
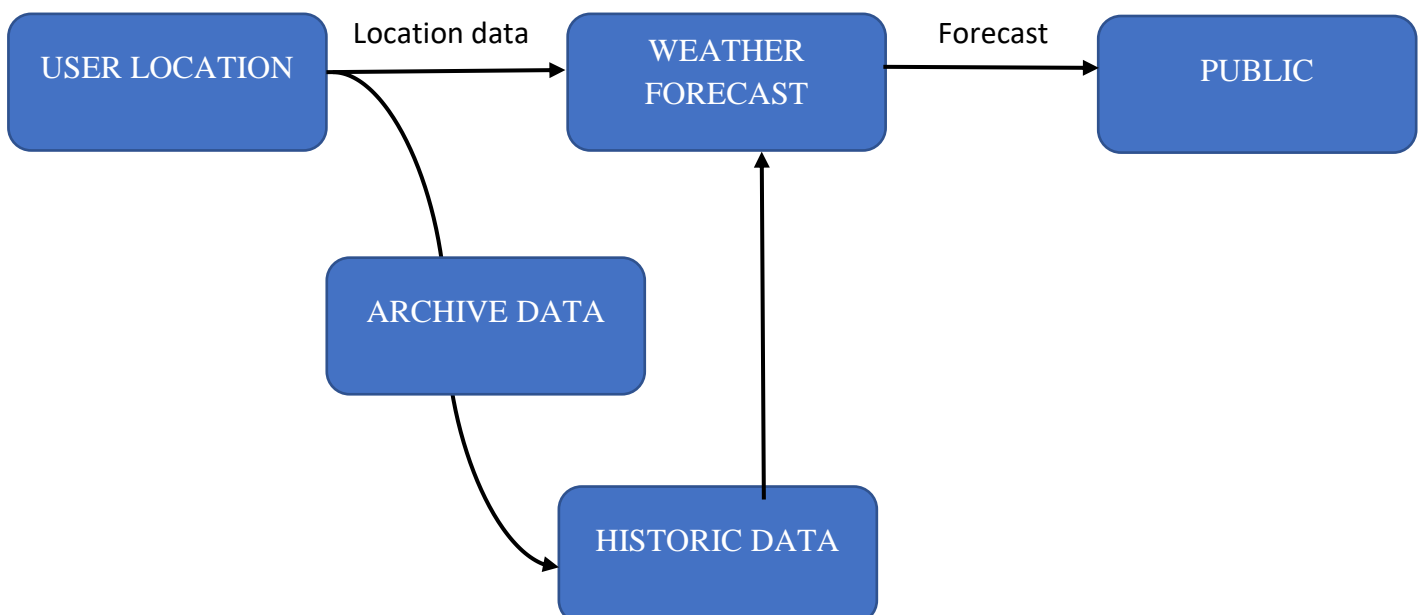
5.4 <u>IMPLEMENTATION</u>

1. DATA COLLECTION

For current weather report data is extracted with the help of **Open weather map** API. The data of weather forecast was obtained from **Kaggle**. We took historic dataset of 8 Indian cities namely Delhi, Kanpur, Nagpur, Pune, Jaipur, Hyderabad, Bombay and Bengaluru. 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. Parameters are :-

- Temperature
- Pressure
- Humidity
- Dewpoint
- Precipitation

2. DATA PRE-PROCESSING

The steps involved in pre-processing are: –

- **Features selection**

The data we have collected has many unwanted attributes which will not be needed in our project. Hence, we use the attributes which we need only.

- **Identifying and handling the missing values**

In data pre-processing, it is pivotal to identify and correctly handle the missing values, failing to do this, we might draw inaccurate and faulty conclusions and inferences from the data. So, after the dataset is filtered from unwanted attributes we check for our dataset if it contains any null or missing values, if occurred the missing value is replaced with the mean of remaining values in the column.

- **Splitting the dataset**

In the next step we split our model into two separate sets – training set and test set. Training set corresponds to 80% of dataset and test set corresponds to 20% of dataset.
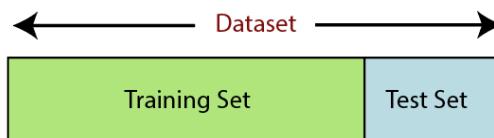
3. CHOOSING A MODEL

Model selection is the process of choosing one of the models as the final model that addresses the problem.

In this project the best fit algorithm for our model is found to be random forest

Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model.*

As the name suggests, **"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."** Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

The below diagram explains the working of the Random Forest algorithm:

.
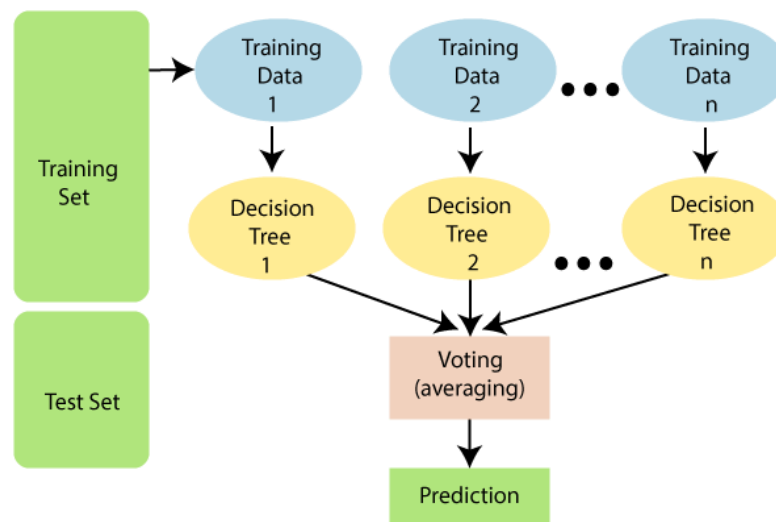
4. TRAINING A MODEL

After selecting the best fit algorithm we consistently trained our model so that it can significantly improve the prediction rate of the model. The weights of the model must be initialized randomly. This way the algorithm will learn to adjust the weights accordingly.

5. MODEL ASSESSMENT

Once the model is fit with the algorithm the next step is to evaluate it whether it performs well or not. So we have evaluated our model against the test data. This helps us to assess the accuracy of the model.

6. PARAMETER TUNNING

This step is for further improvement of the model if possible. To increase the accuracy of model we have made three changes in the default parameters of algorithm. They are :-

1. max_depth = 10,
2. n_estimators = 100, and
3. random_state = 0

7. GRAPHICAL USER INTERFACE

After the successful completion of the model we have created a very user friendly interface with the help of tkinter module of python through which user can easily interact with the model. We have given 3 options to the user, they are :-

1. **Current weather report :-** In this only the current weather condition will be displayed to the user. It includes current temperature, humidity, feels like, description, pressure, dew point, visibility and clouds.
2. **Forecast next 48 hours weather :-** Along with the current weather report, the user will get the next 48 hours temp predictions on the hourly basis.
3. **Complete weather report :-** In this the user will get the complete weather report i.e. current as well as the next hourly 48hrs temp predictions along with the graph (temp vs time)

After selecting any one of the three options listed above, user is required to select the city from the combo box and press the submit button. The output will be displayed on the interface.

5.5 RESULT AND DEMONSTRATION

RESULT

The prediction system works properly. The values of all the attributes were properly pre-processed. After all the pre-processing was completed, model was implemented and it was trained using train data. The GUI for the system was made with tkinter. The coding was done in VsCODE and Jupyter notebook. After completing all the process, we connected the front-end with the back-end.

Our accuracy was found to be around 95.48%.

DEMOSTRATION

CURRENT WEATHER REPORT

<--
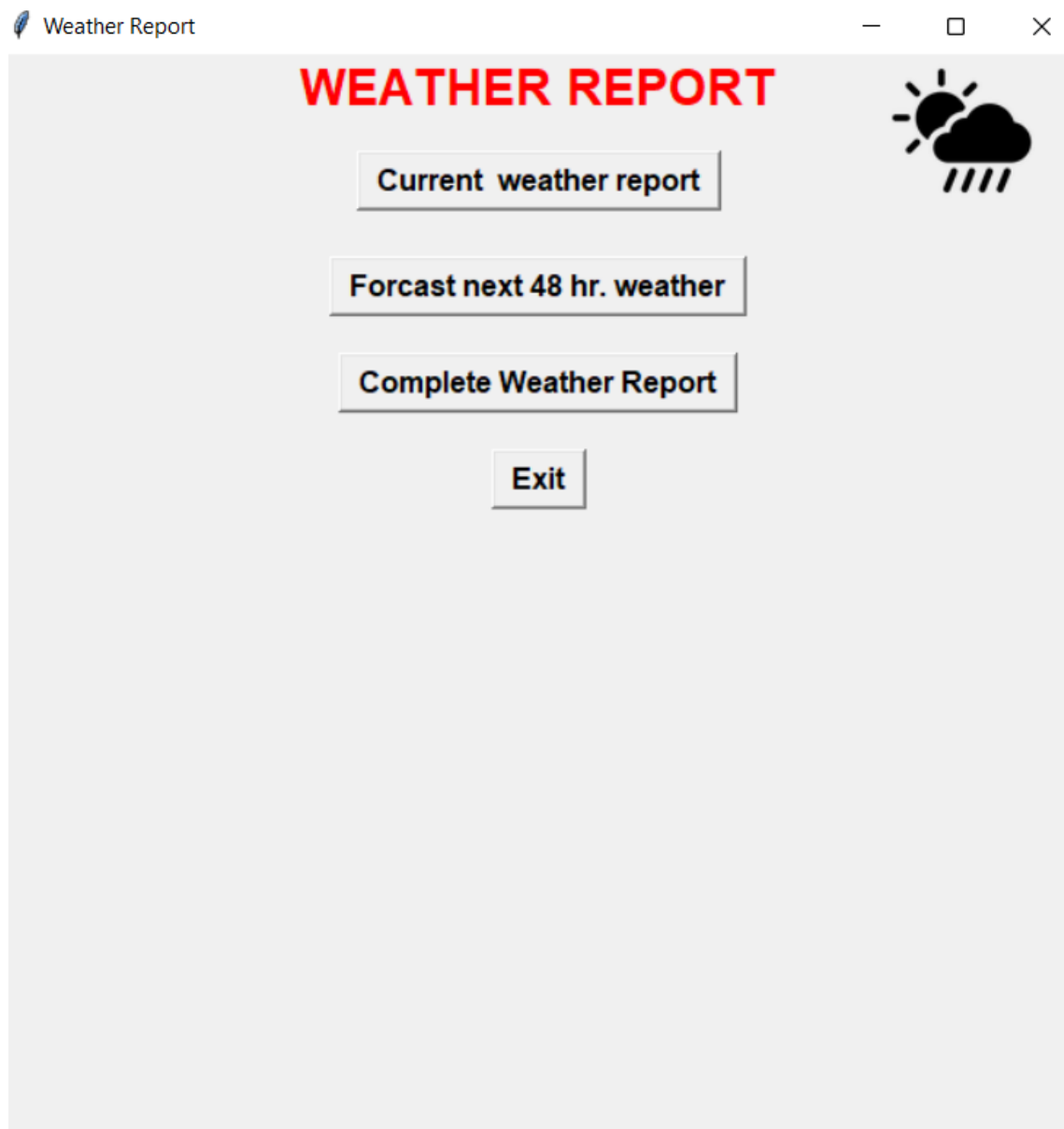
Select City

delhi
kanpur
lucknow
bangalore
mumbai
jaipur
hyderabad
nagpur
pune

Submit

Temperture

Humidity

Feels Like

Description

Pressure

Dew point

Visibility

Clouds

CURRENT WEATHER REPORT

<--

Select City   nagpur

Submit

Temperture	13 °C
Humidity	76 %
Feels Like	12.34 °C
Description	haze
Pressure	1013 hPa
Dew point	8.87 °C
Visibility	3000 m
Clouds	5 %

FORECAST WEATHER

<--

Select City   delhi

Submit

2022-01-20 02:30:00	8.09 °C
2022-01-20 03:30:00	7.94 °C
2022-01-20 04:30:00	7.66 °C
2022-01-20 05:30:00	6.78 °C
2022-01-20 06:30:00	8.62 °C
2022-01-20 07:30:00	8.56 °C
2022-01-20 08:30:00	9.76 °C
2022-01-20 09:30:00	11.58 °C
2022-01-20 10:30:00	13.42 °C
2022-01-20 11:30:00	14.57 °C
2022-01-20 12:30:00	15.49 °C
2022-01-20 13:30:00	17.62 °C
2022-01-20 14:30:00	17.72 °C
2022-01-20 15:30:00	17.72 °C
2022-01-20 16:30:00	17.64 °C

COMPLETE WEATHER REPORT

Select City nagpur

Submit

Temperture	13 °C
Humidity	76 %
Feels Like	12.34 °C
Description	haze
Pressure	1013 hPa
Dew Point	8.87 °C
Visibility	3000 m
Clouds	5 %

Next 48 hr. Forcast

2022-01-20 02:30:00	13.33 °C
2022-01-20 03:30:00	13.17 °C
2022-01-20 04:30:00	13.62 °C
2022-01-20 05:30:00	12.46 °C
2022-01-20 06:30:00	12.53 °C
2022-01-20 07:30:00	13.88 °C
2022-01-20 08:30:00	16.2 °C
2022-01-20 09:30:00	18.4 °C
2022-01-20 10:30:00	20.59 °C
2022-01-20 11:30:00	22.7 °C
2022-01-20 12:30:00	24.23 °C
2022-01-20 13:30:00	25.12 °C

Temp vs Time Graph
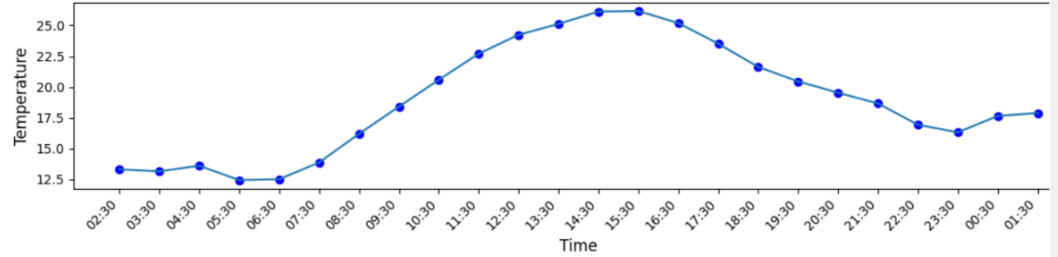
First 24 hours

Next 48 hr. Forcast

2022-01-20 02:30:00	13.33 °C
2022-01-20 03:30:00	13.17 °C
2022-01-20 04:30:00	13.62 °C
2022-01-20 05:30:00	12.46 °C
2022-01-20 06:30:00	12.53 °C
2022-01-20 07:30:00	13.88 °C
2022-01-20 08:30:00	16.2 °C
2022-01-20 09:30:00	18.4 °C
2022-01-20 10:30:00	20.59 °C
2022-01-20 11:30:00	22.7 °C
2022-01-20 12:30:00	24.23 °C
2022-01-20 13:30:00	25.12 °C
2022-01-20 14:30:00	26.12 °C
2022-01-20 15:30:00	26.16 °C
2022-01-20 16:30:00	25.18 °C
2022-01-20 17:30:00	23.51 °C
2022-01-20 18:30:00	21.63 °C
2022-01-20 19:30:00	20.46 °C
2022-01-20 20:30:00	19.54 °C
2022-01-20 21:30:00	18.67 °C
2022-01-20 22:30:00	16.94 °C
2022-01-20 23:30:00	16.33 °C
2022-01-21 00:30:00	17.66 °C
2022-01-21 01:30:00	17.89 °C
2022-01-21 02:30:00	17.29 °C

Temp vs Time Graph

First 24 hours
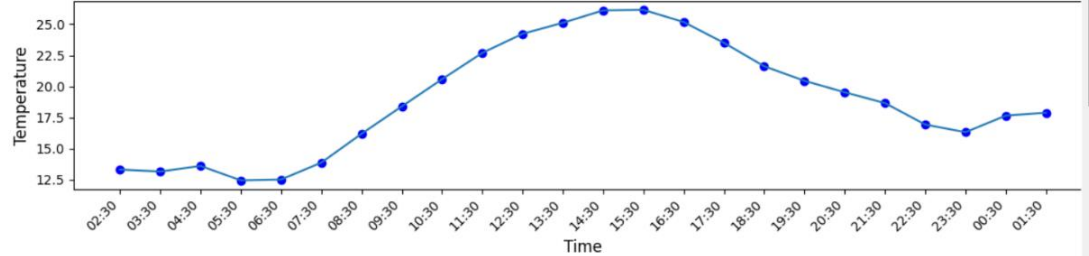
Next 24 hours
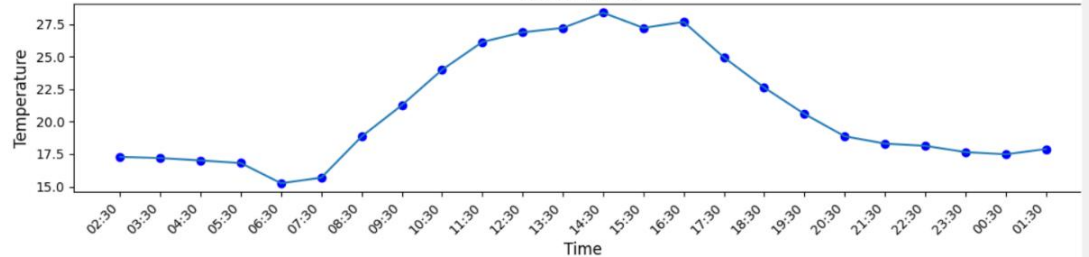
Show desktop

5.6 Contribution of Project to the Society

Weather affects aviation, air quality, health, ground and marine transportation, defense, agriculture, fisheries, water, energy, construction, tourism, and many other sectors of the economy. As society becomes more sensitive to weather, the importance of weather prediction for the protection of lives and property and continued economic growth increases. For example, the U.S. population that resides within 50 miles of the nation's coastlines and is most threatened by hurricanes and flooding is growing rapidly. Such population growth in these and other high-risk areas significantly increases the need for improved weather predictions and warnings to minimize risks to life and property. Another consideration is that the new economic concept of "just-in-time manufacturing" uses computer-timed and -directed supply systems to eliminate the warehousing of parts and products at ports and factories. However, even minor weather disruptions of land, sea, and air-supply-system pathways caused by snow, ice, and high-wind weather systems can now have large, leveraged impacts on these production systems, whereas previously they had little effect.

The goal of weather prediction is to provide information people and organizations can use to reduce weather-related losses and enhance societal benefits, including protection of life and property, public health and safety, and support of economic prosperity and quality of life. In economic terms, the benefit of the investment in public weather forecasts and warnings is substantial: the estimated annualized benefit is about $31.5 billion, compared to the $5.1 billion cost of generating the information. The annual impacts of adverse weather on the national highway system and roads are staggering: 1.5 million weather-related crashes with 7,400 deaths, more than 700,000 injuries, and $42 billion in economic losses (BTS, 2007). In addition, $4.2 billion is lost each year as a result of weather-related air traffic delays (NOAA, 2010). Weather is also a major factor in the complex set of interactions that determine air quality; more than 60,000 premature deaths each year are attributed to poor air quality (Schwartz and Dockery, 1992).

Better forecasts and warnings are reducing these numbers, but much more can be done.

5.7 Conclusion, Limitations and Future Work

Conclusion

The weather prediction done using Random Forest algorithm is very essential for improving the future performance for the people. For predicting the weather, the Random Forest algorithm was applied to the datasets of the weather. We made a model to predict the weather using some selected input variables collected from Kaggle. The problem with current weather scenario is that we are not able to prepare our self and not able to do some important works. So, for knowing the weather scenario at high accuracy considering every factor that affects in the weather scenario, this model is created.

Limitations

This weather reporting system is limited to current weather report and next 48 hours hourly temp prediction. In terms of cities it is limited to 8 top cities of India namely Delhi, Kanpur, Nagpur, Pune, Jaipur, Hyderabad, Bombay and Bengaluru.

Weather forecasts are becoming more detailed, more accurate and are providing the information needed to make sound decisions to protect life and property. Technological advances, such as apps, are making weather information more accessible and immediately alerting those in harm's way.

Future Work

The current version of Weather Prediction that we have developed is still premature. This implies that there are still many limitations that can be resolved and improved. One of the biggest limitation right now is, that the location has to be chosen from the list of places the application is bound to. This can be improved if we use web scraping tools to automatically get the weather data, for various locations, from the internet and then input it into the database. Another enhancement that can be done is the automatic validation of longitude and latitude coordinates. Another improvement that can be done is to beautify the Ul to make it more appealing to the younger generation.

Future enhancements will make our Weather Prediction more flexible, user friendly and thus it will be more appealing to a wide range of audience.

6. REFERENCES

- Tanni Dhoom , Munmun Biswas , Sayantanu Barua, " Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data ", International Journal of Computer Applications, Issue 3 December 2018

- https://www.kaggle.com/hiteshsoneji/historical-weather-data-for-indian-cities

- https://openweathermap.org/

- https://www.javatpoint.com/machine-learning-random-forest-algorithm

- https://www.upgrad.com/blog/data-preprocessing-in-machine-learning/

- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition