**WEATHER PREDICTION**

**COURSE PROJECT REPORT**

**18CSE398J -Machine Learning - Core Concepts with Applications**

**(2018 Regulation)**

**III Year/ VI Semester**

**Academic Year: 2022 -2023 (EVEN)**

By

**Kushal Khandal (RA2011003010327)**

**GAUTAM (RA2011003010426)**

**RUTWICK(RA2011003010466)**

Under the guidance of

**Dr. Vijayalakshmi V**

**Assistant Professor**

**Department of Data Science and Business Systems**



**DEPARTMENT OF DATA SCIENCE AND BUSINESS SYSTEMS**

**FACULTY OF ENGINEERING AND TECHNOLOGY**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

**Kattankulathur, Chennai**

**MAY 2023**

Kushal, [kk8855@srmist.edu.in](mailto:kk8855@srmist.edu.in)

Gautam, [gq7812@srmist.edu.in](mailto:gq7812@srmist.edu.in)

Rutwick, br8535@srmist.edu.in

V. Vijayalakshmi, vijayalv3@srmist.edu.in

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **S No** | **Title** | **Page no.** |
| 1 | Abstract | 3 |
| 2 | Introduction | 4 |
| 3 | Dataset | 5-6 |
| 4 | Methods | 7-9 |
| 5 | Experiments and Results | 10-11 |
| 6 | Conclusion and Future Work | 12 |
| 7 | References | 13-14 |

**1.Abstract**

Weather predicting is the attempt by meteorologists to predict the weather conditions at some future time and the weather conditions that may be expected. The climatic condition parameters are based on the temperature, wind, humidity, rainfall and size of data set. Here, the parameters temperature and Humidity only are considered for experimental analysis. The data is collected from the temperature and humidity sensor called DHT11 sensor, which helps in detecting the temperature and humidity values of a particular region or location. The raspberry pi is used for storing the collected data to the cloud, with the help of Ethernet shield for uploading the data online. The data stored in cloud is generated in the form of CSV, JSON, XML files which is used for further analysis. The correlation analysis of the parameters helps in predicting the future values. The ARIMA model that gives better results for time-series data is used for predicting the values for forthcoming.

**2.Introduction**

Weather prediction is the process of using scientific methods and technologies to predict the future weather conditions, such as temperature, humidity, precipitation, wind speed and direction, and atmospheric pressure, for a specific location and time period. Weather prediction plays a crucial role in various aspects of our daily lives, including agriculture, transportation, energy management, disaster preparedness, and outdoor activities planning.

Weather prediction relies on a wide range of data sources, including weather observations from ground-based weather stations, satellites, and other sensors, as well as historical weather data and numerical weather prediction models. These data are processed and analysed using advanced mathematical and statistical techniques to generate forecasts that can provide valuable information for decision-making.

Modern weather prediction models use complex algorithms that consider various atmospheric and environmental factors, such as temperature gradients, pressure systems, humidity levels, and wind patterns, to generate weather forecasts. These models are continuously improved and updated based on new data and advancements in technology, such as high-performance computing, machine learning, and big data analytics.

Weather prediction has significant societal and economic impacts, as accurate and timely weather predictions can help individuals, businesses, and governments make informed decisions related to agriculture, transportation, emergency planning, energy management, and many other areas. It enables farmers to optimize planting and harvesting schedules, airlines to plan flight routes, utilities to manage energy supply and demand, and emergency responders to prepare for severe weather events

**3.Dataset**

We have taken the data from kaggle.com.The dataset contains information about various weather parameters such as temperature, rainfall, evaporation, wind direction and speed, humidity, atmospheric pressure, cloud cover, and whether it rained or not. It could be used for various weather-related analyses or predictive modeling to understand and predict weather patterns.

MinTemp: The minimum temperature recorded on a given day in Celsius degrees.

MaxTemp: The maximum temperature recorded on a given day in Celsius degrees.

Rainfall: The amount of rainfall recorded in millimeters on a given day.

Evaporation: The amount of water that has evaporated from a given surface (such as a water tank) in millimeters.

Sunshine: The number of hours of bright sunshine on a given day.

WindGustDir: The direction of the strongest gust of wind recorded on a given day, in 16 compass points (such as N, NE, E, SE, etc.).

WindGustSpeed: The speed (in km/h) of the strongest gust of wind recorded on a given day.

WindDir9am: The direction of the wind at 9am on a given day, in 16 compass points.

WindDir3pm: The direction of the wind at 3pm on a given day, in 16 compass points.

WindSpeed9am: The wind speed (in km/h) at 9am on a given day.

WindSpeed3pm: The wind speed (in km/h) at 3pm on a given day.

Humidity9am: The percentage of relative humidity at 9am on a given day.

Humidity3pm: The percentage of relative humidity at 3pm on a given day.

Pressure9am: The atmospheric pressure (in hpa) reduced to mean sea level at 9am on a given day.

Pressure3pm: The atmospheric pressure (in hpa) reduced to mean sea level at 3pm on a given day.

Cloud9am: The fraction of sky obscured by cloud at 9am on a given day.

Cloud3pm: The fraction of sky obscured by cloud at 3pm on a given day.

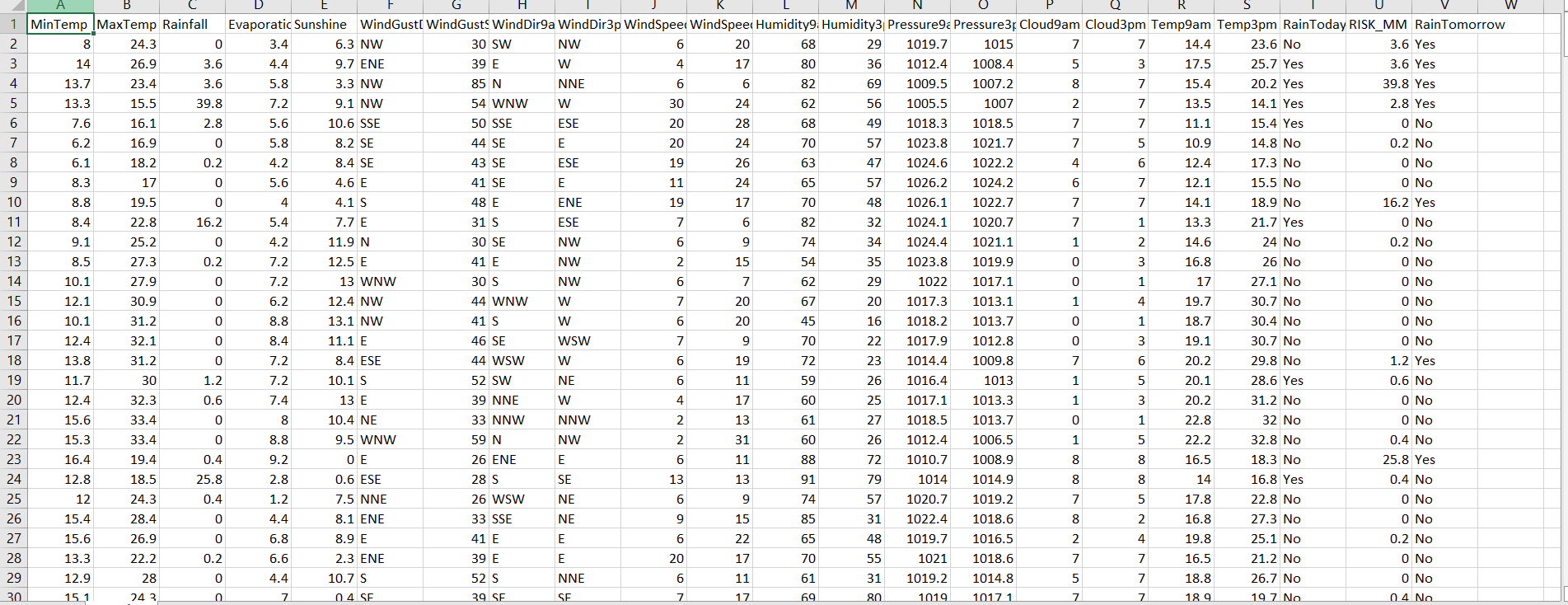
Temp9am: The temperature (in Celsius degrees) at 9am on a given day.

Temp3pm: The temperature (in Celsius degrees) at 3pm on a given day.

RainToday: A binary variable indicating whether it rained on a given day (Yes) or not (No).

RISK\_MM: The amount of rainfall (in mm) in the next day, used to create a response variable RainTomorrow.

RainTomorrow: A binary variable indicating whether it will rain tomorrow (Yes) or not (No).



**4.Methods**

* **Logistic regression**

Logistic regression is a supervised learning algorithm used for binary classification tasks, where the goal is to predict the probability of an event occurring or not. In weather predicting, logistic regression can be used to predict binary weather events, such as rain or no rain, based on input features such as temperature, humidity, and wind speed.

The logistic regression model uses a mathematical function called the logistic function to transform the input features into probabilities. The logistic function produces output values between 0 and 1, representing the probability of the event occurring. The model is trained on labelled data, where the input features and corresponding binary weather events are used to learn the optimal coefficients for the features that maximize the likelihood of the observed data.

Once the logistic regression model is trained, it can be used to make predictions on new, unseen data. The predicted probabilities can be thresholder to obtain binary predictions, such as rain or no rain, based on a chosen threshold value. The performance of the logistic regression model can be evaluated using various metrics such as accuracy, precision, recall, and F1 score, to assess its accuracy and effectiveness in predicting weather events.

* **Random Forest:**

Random Forest is an ensemble learning algorithm that combines multiple decision tree models to make predictions. It is used for both classification and regression tasks and is known for its ability to handle complex data with high accuracy and robustness. In weather forecasting, random forest can be used to predict various weather events, such as precipitation amount, wind speed, and temperature.

A random forest model consists of a collection of decision trees, where each tree is trained on a subset of the data with replacement (bootstrap samples) and a random subset of features. This randomness helps to reduce overfitting and makes the model more robust. The decision trees in the random forest make predictions based on the majority vote or averaging of the predictions from individual trees.

The random forest model can handle non-linear relationships and interactions between features, making it suitable for capturing complex patterns in weather data. It can also handle missing values and outliers in the data effectively. The performance of the random forest model can be evaluated using metrics such as mean squared error (MSE), root mean squared error (RMSE), and R-squared, to assess its accuracy and predictive capability in forecasting weather events.

* **Support Vector Machine (SVM):**

SVM is a supervised machine learning algorithm used for classification and regression. The goal of SVM is to find a hyperplane that best separates data points into different classes. The hyperplane is selected in a way that maximizes the margin (i.e., the distance between the hyperplane and the closest data points from each class). SVM can handle both linearly separable and non-linearly separable data by using kernel functions that transform the data into a higher-dimensional space. SVM is known for its ability to handle high-dimensional data and can be used for binary classification as well as multi-class classification.

* **Decision Tree**

Decision Tree is a flowchart-like tree structure where each internal node represents a decision based on a feature, and each leaf node represents a predicted output label. Decision Trees are used for both classification and regression tasks. Decision Trees recursively split the data based on the feature that provides the best split, using metrics such as Gini impurity or entropy. Decision Trees are easy to understand and interpret, and they can handle both categorical and numerical data. However, Decision Trees are prone to overfitting, and techniques such as pruning or using ensemble methods like Random Forests can be applied to mitigate this issue.

* **Gradient Boosting Classifier (GBC)**

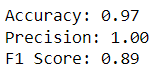
Gradient Boosting Classifier (GBC) is an ensemble machine learning algorithm that combines the predictions of multiple base models (typically decision trees) to create a more accurate and robust predictive model. GBC works in an iterative manner, building each new base model to correct the mistakes of the previous models, with a higher emphasis on the samples that were misclassified. This process continues until a specified number of base models (referred to as "boosting iterations") are trained or until a certain level of accuracy is achieved

* **Naive Bayes:**

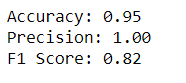
Naive Bayes is a probabilistic algorithm based on Bayes' theorem, used for classification tasks. Naive Bayes assumes that the features are conditionally independent given the class label, which is a simplifying assumption that allows for efficient computation. Despite its simplicity, Naive Bayes has been found to be effective in many real-world applications and is often used as a baseline model. Naive Bayes can handle both categorical and numerical data, and it is particularly useful when dealing with text classification tasks such as spam detection or sentiment analysis. However, Naive Bayes may not perform well in cases where the conditional independence assumption does not hold or when there are strong dependencies between features.

**5.Experiments and results**

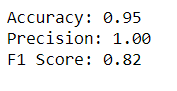
* Logistic Regression



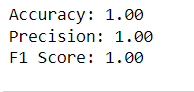
* Random Forest



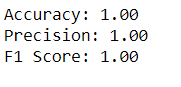
* Support Vector Machine



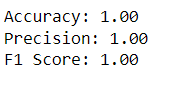
* Decision Tree



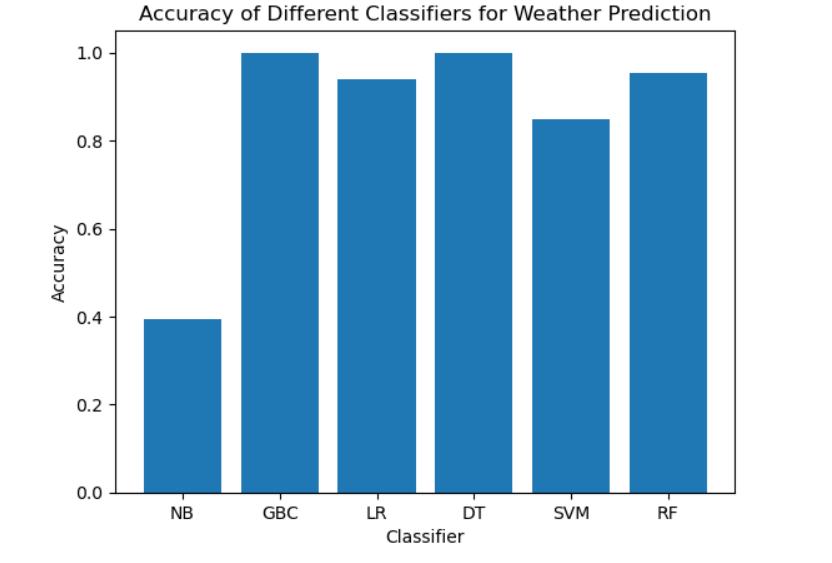
* Naive Bayes



* GBC



* GRAPH OF ACCURACY OF ALL MODELS



**6.Conclusions and future work**

In this report, we explored the use of logistic regression, Random Forest, Gradient Boosting Classifier (GBC), Support Vector Machine (SVM), Naïve Bayes and Decision Tree for weather predicting. The results suggest that these machine learning algorithms can effectively predict weather events and provide valuable insights into weather patterns. Further research can be conducted to improve the accuracy and robustness of these models by incorporating more data sources and advanced techniques. The findings of this study can be used in various applications such as agriculture, transportation, and disaster management to make informed decisions based on accurate weather predictions.

**7.References**

1."An Introduction to Atmospheric Physics" by David G. Andrews: This book provides a comprehensive introduction to atmospheric physics, including the principles behind weather forecasting.

2.National Oceanic and Atmospheric Administration (NOAA) - <https://www.noaa.gov/>

3.European Centre for Medium-Range Weather Forecasts (ECMWF) - <https://www.ecmwf.int/>

4. American Meteorological Society (AMS) - <https://www.ametsoc.org/>

5. World Meteorological Organization (WMO) - <https://public.wmo.int/en>

6. Global Forecast System (GFS) - <https://www.weather.gov/glossary/index.php?letter=g>

7. Ensemble Prediction System (EPS) - <https://www.ecmwf.int/en/forecasts/documentation-and-support/ensemble-prediction-system>

8. Numerical Weather Prediction (NWP) - <https://www.weather.gov/glossary/index.php?letter=n>

9. Climate Prediction Center (CPC) - <https://www.cpc.ncep.noaa.gov/>

10. Storm Prediction Center (SPC) - <https://www.spc.noaa.gov/>

11. National Hurricane Center (NHC) - <https://www.nhc.noaa.gov/>

12. Weather Research and Forecasting Model (WRF) - <https://www2.mmm.ucar.edu/wrf/users/>

13. Global Climate Models (GCMs) - <https://www.carbonbrief.org/mapped-how-global-climate-models-predict-future-temperature-change>

14. Artificial Intelligence (AI) and Machine Learning (ML) in weather forecasting - <https://www.nature.com/articles/s41598-021-88511-1>

15. Deep learning in weather forecasting - <https://www.sciencedirect.com/science/article/pii/S0169743919307204>

16. High-resolution satellite data in weather forecasting - <https://www.nature.com/articles/sdata201457>

17. Doppler radar in weather forecasting - <https://www.weather.gov/glossary/index.php?letter=d>

18. Weather balloons and radiosondes in weather forecasting - <https://www.weather.gov/glossary/index.php?letter=w>

19. Weather forecasting mobile apps - <https://www.tomsguide.com/best-picks/best-weather-apps>

20. Climate change and weather forecasting - <https://www.carbonbrief.org/explainer-how-scientists-use-models-to-predict-future-climate-change>

21. Data assimilation in weather forecasting - <https://www.ecmwf.int/en/forecasts/documentation-and-support/data-assimilation>

**Github Link of the project work (each individual student page link)**

**https://github.com/Gautam298**