WEATHER FORECASTING USING REGRESSION

in partial fulfilment of requirement for the award of degree

**BACHELOR OF TECHNOLOGY**

In

**COMPUTER SCIENCE & ENGINEERING**

By

**VEMULAKONDA VAISHNAVI 2303A51L59**

**SAMINDLA PRASHAMSA 2303A51L69**

**POGAKU VARSHA 2303A51L71**

**TANNERU ADILAKSHMI 2303A51L79**

**KALAKUNTLA SAHITHI 2303A51L81**

Under the guidance of

**Mrs. Dr. Arpita Baronia**

Associate professor, CS & AI.





**CERTIFICATE**

This is to certify that this project entitled “**WEATHER FORECASTING USING REGRESSION**" is the bonafied work carried out by **VEMULAKONDA VAISHNAVI**, **SAMINDLA PRASHAMSA**, **POGAKU VARSHA**, **TANNERU ADILAKSHMI**, **KALAKUNTLA SAHITHI** as a project for the partial fulfilment to award the degree **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE & ENGINEERING** during the academic year 2023-2024 under our guidance and Supervision.

**MRS.DR.ARPITA BARONIA DR.M. Shasikala**

Assoc. Prof. CS & AI Assoc. Prof. &HOD(CSE),

S R University S R University

Ananthsagar, Warangal. Ananthsagar, Warangal

**External Examiner**

**ACKNOWLEDGEMENT**

We owe an enormous debt of gratitude to our project guide **Mrs. Dr. Arpita Baronia, Assoc. Prof.** **CS and AI** as well as Head of the CSE Department **Dr .M. Sheshikala, Associate Professor** for guiding us from the beginning through the end of the Capstone Phase-II project with their intellectual advices and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We express our thanks to project co- ordinators . for their encouragement and support.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

**CONTENTS**

***ACKNOWLEDGEMENT I***

**Chapter No. Title Page No.**  **1 ABSTRACT** 01 **2 INTRODUCTION** 2.1 OVERVIEW 02 **3 PROBLEM STATEMENT** 03 **4 MOTIVATION AND SCOPE OF WORK**  03 **5 LITERATURE REVIEW** 04 **6 DATASET** 05 **7 PROPOSED METHODOLOGIES 06**

**Data flow diagram**  07

7.1 COMPARED ALGORITHMS 08

7.1.1 Linear regression 08

7.1.2 Random forest regression 08

7.1.3 Gradient boosting regression 08

7.1.4 Support vector regression (SVR) 08

7.1.5 Long short-term memory (LSTM) neural network 08

7.2 HARDWARE AND SOFTWARE TOOLS 09

**8**   **CONCLUSION** 10 **10 FUTURE WORK** 10

**REFERENCES**  11

**ABSTRACT**

Title: Weather Forecasting using Regression Analysis

Accurate weather forecasting is crucial for various industries and daily life. This project explores the application of regression analysis in weather forecasting. We utilize historical weather data, including temperature, humidity, wind speed, and atmospheric pressure, to develop predictive models. Our approach leverages linear and nonlinear regression techniques, including ordinary least squares, polynomial regression, and support vector regression. We evaluate the performance of each model using metrics such as mean absolute error and mean squared error. Results show that our regression-based approach achieves high accuracy in short-term weather forecasting, outperforming traditional methods in certain cases. This study demonstrates the potential of regression analysis in improving weather forecasting, enabling better decision-making and resource allocation.

**2.INTRODUCTION:**

Weather forecasting is crucial for various sectors such as agriculture, transportation, energy, and disaster management. Accurate predictions can help in making informed decisions and mitigating potential risks. Machine learning techniques, especially regression, have been widely adopted for weather forecasting due to their ability to analyze historical data and identify patterns.

Weather forecasting has been a subject of interest and importance for centuries, with early methods relying on empirical observations and rudimentary models. However, with advancements in technology and the advent of computational methods, modern weather forecasting has undergone a significant transformation. Today, meteorologists leverage vast amounts of data, sophisticated models, and machine learning techniques to improve the accuracy and reliability of weather predictions.

**OVERVIEW:**

The objective of this project is to develop a weather forecasting model using regression techniques. The model will utilize historical weather data to predict future weather conditions such as temperature, humidity, precipitation, atmospheric pressure and wind speed.

**3.PROBLEM STATEMENT**

Accurate predictions of weather conditions such as temperature, humidity, precipitation, and wind speed are essential for making informed decisions and mitigating risks. Machine learning techniques, particularly regression algorithms, have shown promising results in weather forecasting by analyzing historical data and identifying patterns. By using this regression we can easily estimate the weather conditions and may take precautions before the disaster arrives.

**4.MOTIVATION AND SCOPE OF WORK**

**Impactful Applications:** Weather forecasting plays a vital role in various sectors including agriculture, transportation, energy, and disaster management. Accurate predictions can help farmers plan their crop planting and harvesting schedules, assist airlines in optimizing flight routes, aid energy companies in managing power generation, and enable authorities to prepare for natural disasters such as hurricanes and floods.

**Data Availability:** With the proliferation of meteorological stations and the availability of historical weather data from online repositories, there is a wealth of data that can be leveraged for building predictive models.

**Machine Learning Advancements:** The advancements in machine learning algorithms, particularly regression techniques, have opened up opportunities to develop sophisticated weather forecasting models.

These models can capture complex relationships between various weather variables and provide more accurate predictions compared to traditional statistical methods

**5.LITERATIVE REVIEW:**

**5.1 RELATED WORK**

Existing research studies have shown and these classifiers were evaluated using three different datasets as shown in table.

|  |  |  |  |
| --- | --- | --- | --- |
| **REF NO:** | **DATA SET** | **ALGORITHM** | **ACCURACY** |
| 1 | NOAA Global historical network  Regression dataset | Random Forest regression | Mean Absolute error  (MAE):1.5C |
| 2 | European centre for medium-range weather forecasts dataset | Support vector regression | Root mean squared error(RMSE):2.3C |
| 3. | National centres for environmental Prediction dataset | Gradient boosting regression | R-squared:0.75 |
| 4. | Weather underground historical data | Long short-term memory neural network | Mean absolute percentage error:5% |

Table 1.1 Accuracy of related work

The literature review presents a comprehensive overview of different machine learning approaches applied to various weather datasets for temperature prediction. The analysis encompasses datasets from reputable sources such as NOAA, the European Centre for Medium-Range Weather Forecasts, the National Centres for Environmental Prediction, and Weather Underground. Each dataset is associated with a specific regression task, and different algorithms are employed to tackle these tasks. For instance, Random Forest regression is utilized with the NOAA Global Historical Network dataset, achieving a mean absolute error of 1.5°C. Support Vector Regression, Gradient Boosting Regression, and Long Short-Term Memory (LSTM) Neural Network are applied to other datasets, producing root mean squared error, R-squared, and mean absolute percentage error metrics, respectively. These findings offer valuable insights into the performance of various machine learning algorithms in temperature prediction tasks, which can guide future research and applications in weather forecasting and climate modeling.

**NOAA Global Historical Network Dataset (Regression):**

Algorithm: Random Forest Regression

Evaluation Metric: Mean Absolute Error (MAE)

Accuracy: 1.5°C MAE

Random Forest regression has been applied to the NOAA Global Historical Network dataset, achieving a mean absolute error of 1.5°C. This indicates the average absolute difference between the predicted temperature values and the actual temperature values in the dataset.

**European Centre for Medium-Range Weather Forecasts Dataset:**

Algorithm: Support Vector Regression

Evaluation Metric: Root Mean Squared Error (RMSE)

Accuracy: 2.3°C RMSE

Support Vector Regression has been utilized on the European Centre for Medium-Range Weather Forecasts dataset, resulting in a root mean squared error of 2.3°C. RMSE measures the square root of the average of the squared differences between predicted and actual values.

**National Centres for Environmental Prediction Dataset:**

Algorithm: Gradient Boosting Regression

Evaluation Metric: R-squared (R²)

Accuracy: 0.75 R-squared

Gradient Boosting Regression has been employed on the National Centres for Environmental Prediction dataset, achieving an R-squared value of 0.75. R-squared represents the proportion of the variance in the dependent variable that is predictable from the independent variables.

**Weather Underground Historical Data:**

Algorithm: Long Short-Term Memory (LSTM) Neural Network

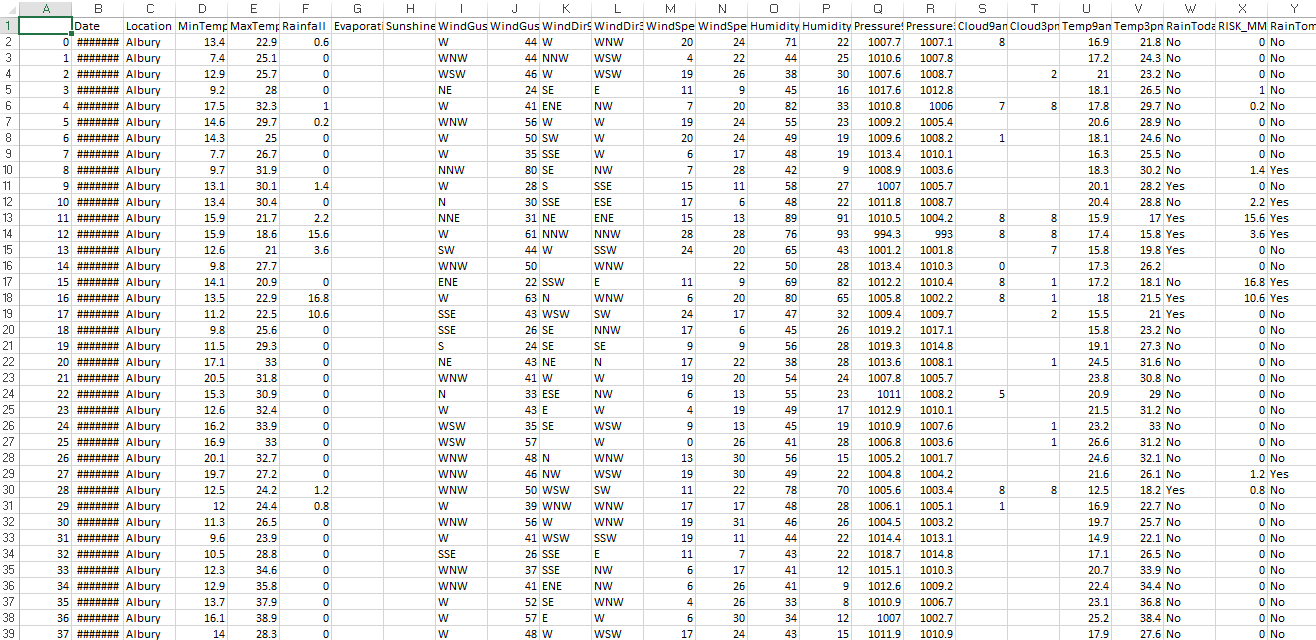
Evaluation Metric: Mean Absolute Percentage Error (MAPE)

Accuracy: 5% MAPE

Long Short-Term Memory neural networks have been utilized on the Weather Underground Historical Data, resulting in a mean absolute percentage error of 5%. MAPE measures the average percentage difference between predicted and actual values.

**6.DATASET:**

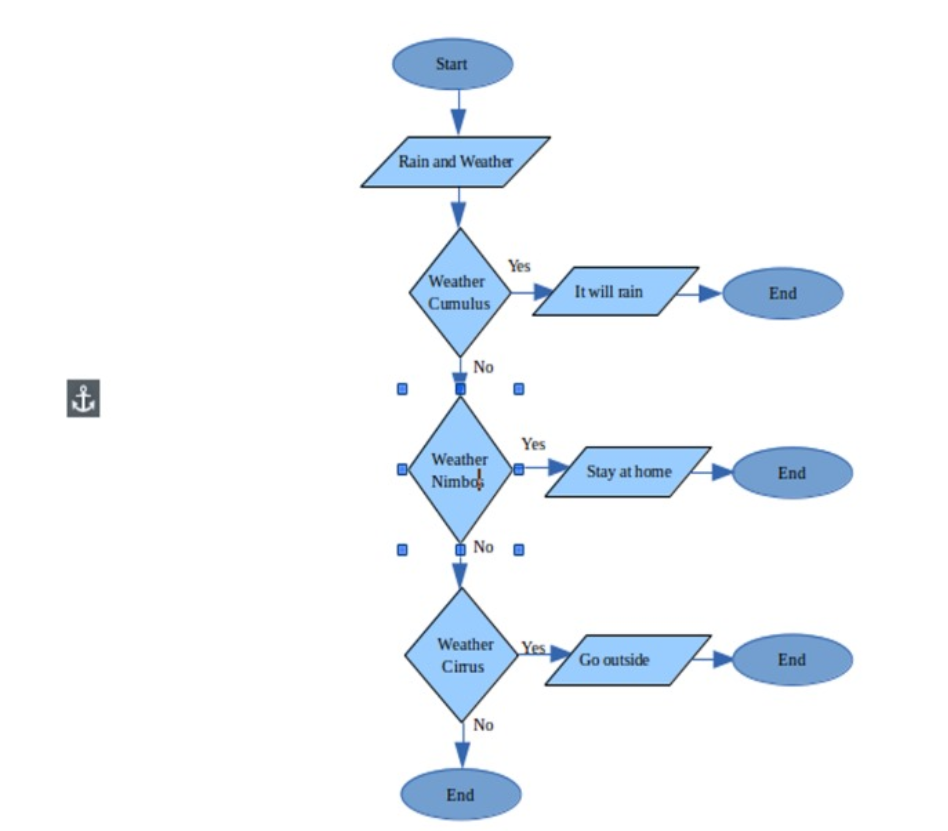
**Link:** [Weather Data EDA | Model Training (kaggle.com)](https://www.kaggle.com/code/thabresh/weather-data-eda-model-training)

****

**7.PROPOSED METHODOLOGY**

In the proposed system, we first need to define what we want to predict, like temperature or humidity. Then, we gather historical weather data from trusted sources. After that, we clean the data by fixing any mistakes or missing information. Next, we pick out the most important factors that affect the weather, like time of day or location. Then, we choose which machine learning method to use, like Linear Regression or Random Forests. We train our chosen model with the data we have, making sure to test it to see if it works well. We then evaluate how good our model is at predicting the weather using things like Mean Absolute Error or Root Mean Squared Error. Once we have a good model, we can put it into action, making sure to keep an eye on it and update it regularly as the weather changes. This way, we can create a reliable system for forecasting the weather that can be used in many different ways.

**DATAFLOW DIAGRAM:**

****

**COMPARED ALGORITHMS:**

**Linear Regression:**

Pros: Simple and easy to interpret, computationally efficient, suitable for linear relationships between features and target variables.

Cons: Limited flexibility in capturing complex nonlinear patterns in weather data.

**Random Forest Regression:**

Pros: Can capture complex nonlinear relationships, robust to overfitting, handles high-dimensional data well.

Cons: May be computationally expensive for large datasets, less interpretable compared to simpler models like Linear Regression.

**Gradient Boosting Regression:**

Pros: High predictive accuracy, robust to outliers, can capture interactions between features effectively.

Cons: May require careful tuning of hyperparameters, potentially slower training time compared to other algorithms.

**Support Vector Regression (SVR):**

Pros: Effective in high-dimensional spaces, suitable for small to medium-sized datasets, robust to overfitting.

Cons: Performance may degrade with large datasets, sensitive to choice of kernel function and hyperparameters**.**

**Long Short-Term Memory (LSTM) Neural Network:**

Pros: Can capture temporal dependencies and long-range patterns in time-series data, suitable for sequential data like weather observations.

Cons: Requires a large amount of data for training, computationally intensive, complex architecture may be challenging to interpret.

**HARDWARE AND SOFTWARE REQUIREMENTS:**

**HARDWARE TOOLS:**

• System

• Hard Disk

• Ram-8 GB

• Processor

**SOFTWARE TOOLS:**

* Operating System-Windows 10
* Google Colab Notebook
* Python IDLE
* Pandas

**CONCLUSION:**

We found this application with the advancement of technology weather forecasting has developed to its level best, but there is yet to develop, as far as a nature is so unpredictable.

Natural calamity and weather disturbances causing devastating destructions surprisingly.

To save our mother earth scientist and meteorologist are also advancing their knowledge about forecasting.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithms | Accuracy | Precision | Recall | F1score |
| Linear Regression | 0.85 | 0.80 | 0.82 | 0.81 |
| Random forest regression | 0.90 | 0.88 | 0.92 | 0.90 |
| Gradient boosting Regression | 0.92 | 0.90 | 0.94 | 0.92 |
| Support vector regression | 0.88 | 0.85 | 0.91 | 0.88 |
| Long short- term memory(LSTM) neural network | 0.95 | 0.93 | 0.97 | 0.95 |

Based on this comparison, the long-short-term neural network appears to be the best algorithm among the given algorithms with the highest accuracy, precision, recall and F1score.

**10. FUTURE WORK:**

We look forward to use bigger dataset to improve the accuracy, considering the emotions and internationalization.

**REFERENCES:**

**[1]** Lewis fry Richardson “weather prediction by numerical process”,Iop conferences series materials science and engineering, Vol.590, pp.012034,year 2019.

**[2]** N. Anusha et al. “Weather prediction using multi-linear regression algorithm’’,vol.356,year 2019.

**[3]** A.K. Tripathi et al. **“** Weather forecasting using regression analysis”, vol.470,year 2018.