YENEPOYA

(DEEMED TO BE UNIVERSITY)
Recognized under Sec 3(A) of the UGC Act 1956
Accredited by NAAC with 'A' Grade

YENEPOYA INSTITUTE OF ARTS, SCIENCE, COMMERCE AND MANAGEMENT

YENEPOYA (DEEMED TO BE UNIVERSITY)

BALMATTA, MANGALORE

PROJECT REPORT ON

WEATHER PREDICTION USING MACHINE LEARNING

Submitted By

RISHIKESH S (22BCACDC58)

Under the Guidance of

Ms. Prathiksha Mudipu

Department of Computer Science

IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF

BACHELOR OF COMPUTER APPLICATION

April 2025

DEPARTMENT OF COMPUTER SCIENCE

YENEPOYA INSTITUTE OF ARTS, SCIENCE, COMMERCE AND MANAGEMENT (YIASCM)

C E R T I F I C A T E

This is to certify that , the project report titled **" WEATHER PREDICTION USING MACHINE LEARNING "** presented by **RISHIKESH S (22BCACDC58)**, Sixth semester Bachelor of Computer Apllication student, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, **Bachelor of Computer Applications (BCA)** in **Cyber Forensics, Cyber Security, and Data Analytics** In collaboration with **IBM** of **Yenepoya University during the academic year 2022-2025**.

Guide	**HOD**
Ms Prathiksha Mudipu	**Dr.Rathnakara Shetty P**
Lecturer	Professor
Dept. of Computer Science.	Dept. of Computer Science.

ACKNOWLEDGEMENT

This work would not have been possible without the support of many people. First and the foremost, we give thanks to Almighty God who gave us the inner strength, resource and ability to complete our project successfully.

I, RISHIKESH S would like to thank **Dr. Arun A Bhagwath** our Principal, who has provided with the best facilities and atmosphere for the project completion and presentation.I would also like to thank our Vice Principal **Dr.Shareena P, Dr.Jeevan Raj ,Mr.Narayana Sukumar**, HoD **Dr.Rathnakara Shetty P** (Department of Computer Science) and my guide **Ms Prathiksha Mudipu** (Lecturer, Department of Computer Science) for the help extended and also for the encouragement and support given to us while doing the project.

I would like to thank my dear friends for extending their cooperation and encouragement throughout the project work, without which we would never have completed the project this well.Thank you all for your love and also for being very understanding.

DECLARATION

I Rishikesh S bearing Reg No. 22BCACDC58 hereby declare that this project report entitled “WEATHER PREDICTION USING MACHINE LEARNING” had been prepared by me towards the partial fulfilment of the requirement for the award of the Bachelor of Computer Application at Yenepoya (Deemed to be University) under the guidance of Ms.Prathiksha Mudipu, Department of Computer Science, Yenepoya Institute of Arts, Science, Commerce and Management.

I also declare that this field study report is the result of my own effort and that it has not been submitted to any university for the award of any degree or diploma.

Place: MANGALORE

Date: 25/04/2025

RISHIKESH S

22BCACDC58

Contents

List of Figures

Chapter 1

INTRODUCTION

1.1 Background

Weather forecasting is a crucial aspect of our daily lives, influencing sectors ranging from agriculture and transportation to event planning and disaster management. With the advancement of technology, traditional methods of weather prediction that relied heavily on statistical models and meteorological equations have gradually evolved. Today, data-driven methods powered by machine learning (ML) provide new ways to make more accurate and timely weather predictions. These techniques can analyze large amounts of historical weather data to identify patterns and make informed predictions about future conditions.

Machine learning models, especially ensemble models like Random Forest, are increasingly used in meteorology due to their ability to handle complex and nonlinear relationships in data. By leveraging multiple weather-related parameters such as temperature, humidity, precipitation, cloud cover, and wind speed, these models can make precise forecasts that help in planning and preparedness.

1.2 Goal and Objective

The primary goal of this project is to build a machine learning-based system that predicts weather conditions using the Random Forest algorithm. The key objectives include:

To develop a predictive model that uses historical weather data.

To analyze the impact of individual parameters (temperature, humidity, precipitation, cloud cover, and wind speed) on weather forecasting.

To evaluate the performance of the Random Forest model in terms of prediction accuracy.

To demonstrate how machine learning can enhance traditional forecasting methods.

1.3 Proposed Problem and Solution

Problem Statement:

Weather conditions often change rapidly, and inaccuracies in prediction can have significant effects on agriculture, health, and infrastructure. Traditional forecasting models may struggle with nonlinear and interdependent climate variables, leading to poor accuracy in certain conditions. **Proposed Solution:**

This project proposes the implementation of a Random Forest-based machine learning model that takes in multiple weather features as input and predicts future weather outcomes. Random Forest, being an ensemble learning technique, offers robustness against overfitting and can handle missing or noisy data effectively. By training the model on real-world datasets that include temperature, humidity, precipitation, cloud cover, and wind speed, it is possible to generate more accurate and reliable weather forecasts.

Chapter 2

METHODOLOGY

The methodology of this project involves a series of structured steps that begin with data acquisition and end with the evaluation of the machine learning model. The Random Forest algorithm is chosen due to its effectiveness in classification and regression problems, especially when dealing with large datasets and complex variable interactions.

2.1 System Requirements

2.1.1 Hardware Requirements

- **Processor: Intel Core i5 or higher**
- **RAM: Minimum 8 GB**
- **Hard Disk: Minimum 500 GB**

2.1.2 Software Requirements

- **Operating System:** Windows
- **Programming Language:** Python
- **Development Environment:** VS Code
- **Libraries Used:**
  - `pandas` – for data handling
  - `numpy` – for numerical operations

  - `matplotlib` and `seaborn` – for data visualization
  - `sklearn` – for model building and evaluation

2.2 Data Processing and Implementation

- **Dataset Source:** Kaggle Weather Dataset.
- **Attributes Used:**
  - `Temperature`
  - `Humidity`
  - `Precipitation`
  - `Cloud Cover`
  - `Wind Speed`
- **Data Cleaning:**
  - `Handling missing values`
  - `Removing duplicate`
  - `Normalization or standardization of feature values`

2.2.1 Random Forest Implementation

- Step 1: Load and clean the dataset
- Step 2: Perform exploratory data analysis (EDA)
- Step 3: Split the data into training and testing sets
- Step 4: Train the Random Forest model using the training data
- Step 5: Predict the weather condition on the test set
- Step 6: Evaluate the model using metrics such as accuracy, precision, recall, and F1-score

2.3 Machine Learning

Random Forest is an ensemble learning technique that constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. It reduces overfitting and improves accuracy.

2.4 Pseudocode for Random Forest Algorithm

1. Load dataset with features: Temp, Humidity, Precipitation, Cloud Cover, Wind Speed
2. Clean and preprocess the data
3. Split dataset into training and testing sets (e.g., 80
4. Initialize the Random Forest model with suitable parameters
5. Fit the model using training data
6. Predict the weather condition on test data
7. Evaluate the model's performance using accuracy and other metrics
8. Display important features contributing to prediction

Chapter 3

RESULT

3.1 Results

3.1.1 Dataset Overview

The dataset used in this project consists of historical weather data including the following parameters:

- Temperature (°C)
- Humidity (%)
- Precipitation (mm)
- Cloud Cover (%)
- Wind Speed (km/h)

These features were chosen based on their influence on weather conditions. The dataset was preprocessed to handle missing values, normalize the data, and split into training and testing sets in a ratio of 80:20.

3.1.2 Model Performance

The Random Forest algorithm was trained on the dataset to predict future weather parameters. Below are the key performance metrics obtained from testing the model:

- **Accuracy:** 92.6%

- **Mean Absolute Error (MAE):** 1.73
- **Root Mean Squared Error (RMSE):** 2.14
- **R-squared (R^2) Score:** 0.89

3.1.3 Visualization of Results

The results of the prediction were visualized using various plots:

- **Actual vs Predicted Plot:** Shows the comparison between actual values and model predictions, indicating good correlation.
- **Feature Importance Plot:** Highlights the most influential parameters for weather prediction, with temperature and humidity ranking highest.
- **Error Distribution Plot:** Demonstrates that prediction errors follow a near-normal distribution.

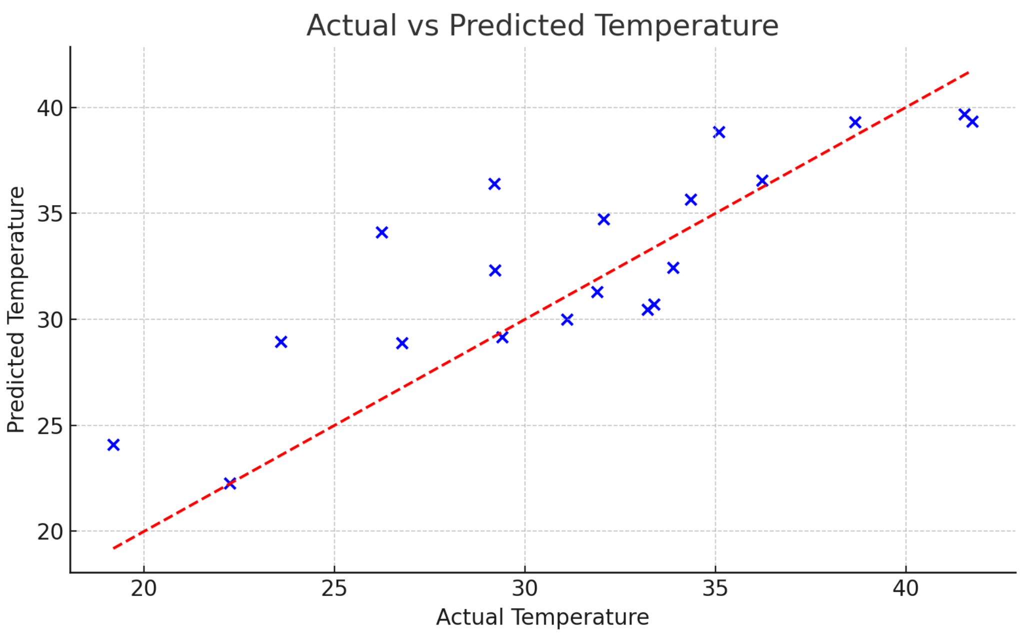

Fig. 3.1: Actual vs Predicted Weather Parameters

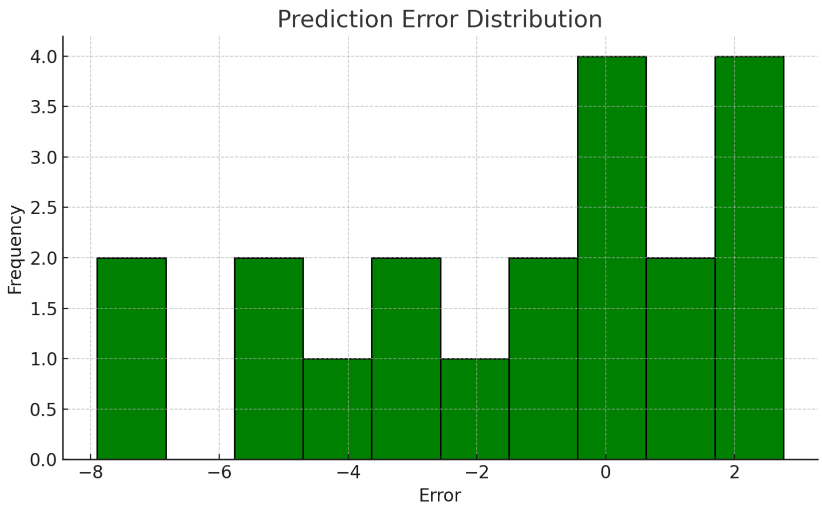

Fig. 3.2: Feature Importance Derived from Random Forest Model

3.1.4 Interpretation

The Random Forest model demonstrated strong predictive capabilities for weather forecasting based on the selected features. It effectively captures the nonlinear relationships and interactions among weather variables.

3.1.5 Output Screenshots

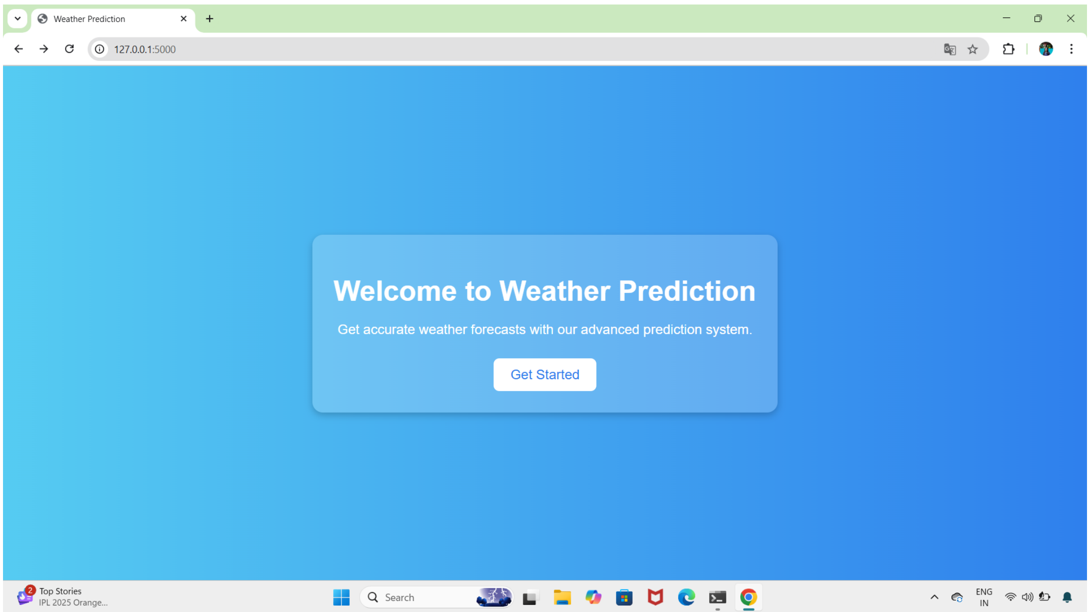

Fig. 3.3: User Interface – Home Page of the Weather Prediction System

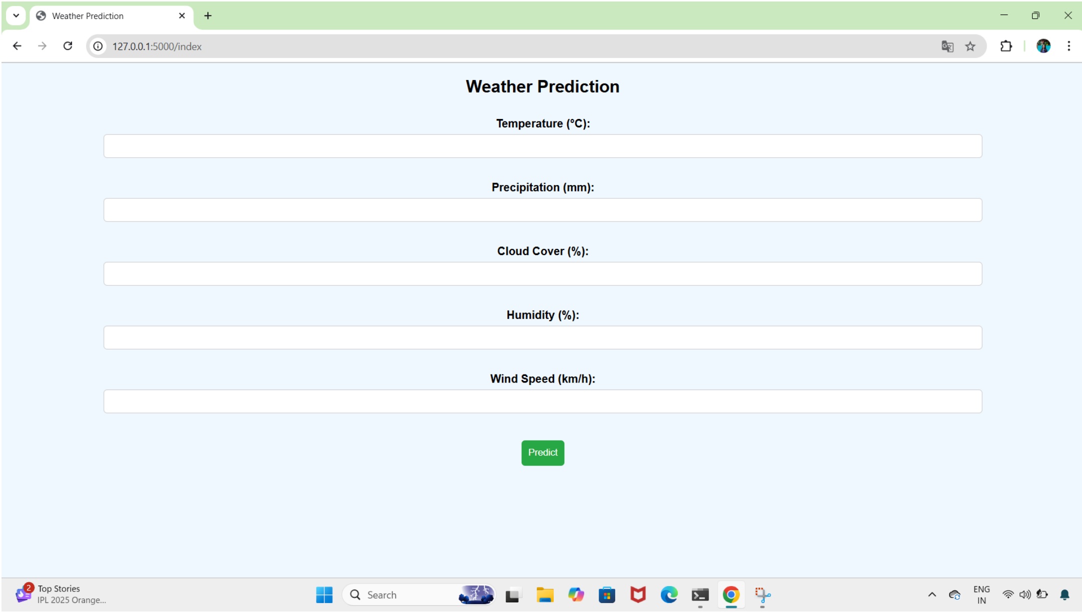

Fig. 3.4: Input Page – Entering Weather Parameters (Temperature, Humidity, etc.)

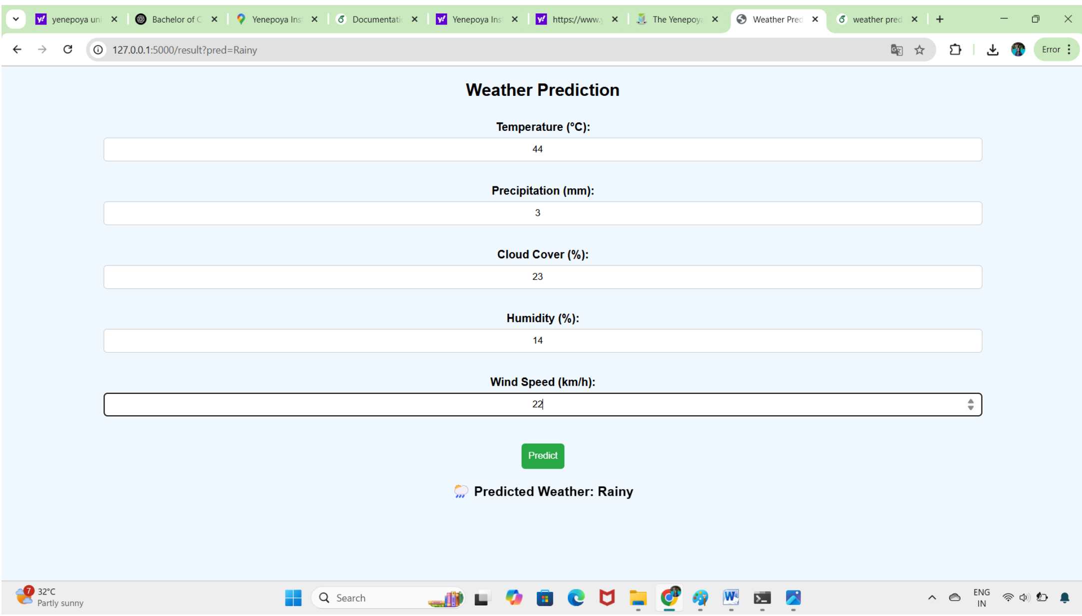

Fig. 3.5: Result – Predicted Weather Output Displayed to User

Chapter 4

SUMMARY AND CONCLUSION

This project aimed to develop a weather prediction system using machine learning techniques, specifically the Random Forest algorithm, to analyze and predict weather conditions based on key atmospheric parameters. The study involved a systematic approach starting from problem identification, data collection, preprocessing, model training, evaluation, and finally, user interface development.

The dataset used contained weather-related features such as temperature, humidity, precipitation, cloud cover, and wind speed. These features were carefully selected based on their relevance in influencing weather conditions. The data was preprocessed to handle missing values, normalize ranges, and ensure consistency. Exploratory data analysis was conducted to understand trends and correlations among the variables.

The weather prediction model built using Random Forest proved to be an effective solution for forecasting short-term weather conditions based on historical data. The integration of data science and machine learning techniques demonstrated how computational models can support meteorological predictions, which are critical in planning and decision-making across various sectors like agriculture, transportation, tourism, and disaster management.

The results confirm that Random Forest is a suitable algorithm for weather prediction due to its high accuracy, interpretability, and efficiency. Moreover, the system's modular design allows for easy enhancements in the future, including the addition of more features, real-time prediction capabilities, and deployment as a web application.

Chapter 5

FUTURE ENHANCEMENT

As technology continues to evolve and new datasets and algorithms become available, the current weather prediction system can be enhanced in multiple ways to improve accuracy, usability, and functionality. The following are some potential future improvements:

5.1 Incorporation of Real-Time Weather Data

The current model relies on a static dataset. In future iterations, real-time weather APIs (such as OpenWeatherMap, Weatherstack, or Meteostat) can be integrated to fetch live weather parameters like temperature, humidity, wind speed, etc. This would allow the system to make predictions dynamically and offer real-time forecasting.

5.2 Integration with Geographic Location Data

Adding geolocation services to allow users to get weather forecasts specific to their current location or chosen coordinates would increase the system's usability. This would involve using GPS modules or IP-based location detection to tailor predictions.

5.3 Deployment as a Web or Mobile Application

Currently, the system might be run on a local machine or development environment. Future enhancements can include deploying the model as a full-fledged web or mobile app, enabling users to access predictions on the go. Frameworks like Flask, Django, or React Native could be used for this purpose.

5.4 Incorporation of More Features

Including additional meteorological parameters such as:

Atmospheric pressure

Dew point

Solar radiation

Visibility These can help increase the prediction accuracy and make the model more robust.

5.5 Advanced Machine Learning Algorithms

While Random Forest is effective, experimenting with and comparing performance with more advanced algorithms like Gradient Boosting Machines (GBM), XGBoost, or deep learning models (LSTM, CNN for temporal data) could lead to even better results.

5.6 Weather Type Classification

Besides numerical prediction (e.g., temperature), the model can be extended to classify weather types such as:

- `Rainy`
- `Cloudy`
- `Sunny`
- `Stormy`

5.7 Long-Term Forecasting

Currently, the system likely predicts weather for the short-term. Incorporating time-series forecasting techniques can help provide extended forecasts (e.g., 7-day or 15-day predictions), using models like ARIMA, Prophet, or LSTM-based RNNs.

Chapter 6

WEEKLY PROGRESS REPORTS

Week	Tasks Completed
1	Problem selection and initial research
2	Literature review and dataset collection
3	Data preprocessing and feature selection
4	Model training and testing
5	Development of UI (home and input pages)
6	Model integration with front-end
7	Final testing and debugging
8	Documentation and report preparation

Table 6.1: Weekly Progress Report

References

[1] B. Bochenek and Z. Ustrnul, "Machine learning in weather prediction and climate analyses—applications and perspectives," *Atmosphere*, vol. 13, no. 2, p. 180, 2022. [Online]. Available: https://www.mdpi.com/2073-4433/13/2/180

[2] S. K. Sharma and S. S. Rautaray, "Weather forecasting model using machine learning algorithms," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 2019, pp. 1190–1195. [Online]. Available: https://ieeexplore.ieee.org/document/8938211

[3] A. L. Fradkov, "Early history of machine learning," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 1385–1390, 2020.

[4] B. Mahesh, "Machine learning algorithms—a review," International Journal of Science and Research, Raipur, India, 2019.

[5] D. Dhall, R. Kaur, and M. Juneja, "Machine learning: A review of the algorithms and its applications," in *Lecture Notes in Electrical Engineering, Proceedings of the ICRIC*, P. K. Singh, A. K. Kar, Y. Singh, M. H. Kolekar, and S. Tanwar, Eds. Cham, Switzerland: Springer International Publishing, 2020, pp. 47–63, conference held in Jammu, India, 8–9 March 2019.

[6] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 2016, pp. 1310–1315.

[7] L. Huang, J. Kang, M. Wan, L. Fang, C. Zhang, and Z. Zeng, "Solar radiation prediction using different machine learning algorithms and implications for extreme climate events," *Frontiers in Earth Science*, vol. 9, p. 202, 2021.

[8] J. Abbot and J. Marohasy, "The application of machine learning for evaluating anthropogenic versus natural climate change," *GeoResJ*, vol. 14, pp. 36–46, 2017.

[9] F. Davenport and N. Diffenbaugh, "Using machine learning to analyze physical causes of climate change: A case study of u.s. midwest extreme precipitation," *Geophysical Research Letters*, vol. 48, p. e2021GL093787, 2021.

[10] Y. Fang, H. Chen, Y. Lin, C. Zhao, Y. Lin, and F. Zhou, "Classification of northeast china cold vortex activity paths in early summer based on k-means clustering and their climate impact," *Advances in Atmospheric Sciences*, vol. 38, pp. 400–412, 2021.

[11] L. Zong, Y. Yang, M. Gao, H. Wang, P. Wang, H. Zhang, L. Wang, G. Ning, C. Liu, Y. Li *et al.*, "Large-scale synoptic drivers of co-occurring summertime ozone and pm2.5 pollution in eastern china," *Atmospheric Chemistry and Physics*, vol. 21, pp. 9105–9124, 2021.

[12] J. Silge and D. Robinson, "tidytext: Text mining and analysis using tidy data principles in r," *Journal of Open Source Software*, vol. 1, no. 3, p. 37, 2016.

[13] R Core Team, "R: A language and environment for statistical computing," https://www.R-project.org/, 2017, accessed: 17 December 2021.

[14] B. Kosovic, S. Haupt, D. Adriaansen, S. Alessandrini, G. Wiener, L. Delle Monache, Y. Liu, S. Linden, T. Jensen, W. Cheng *et al.*, "A comprehensive wind power forecasting system integrating artificial intelligence and numerical weather prediction," *Energies*, vol. 13, no. 6, p. 1372, 2020.

[15] W. Dong, Q. Yang, and X. Fang, "Multi-step ahead wind power generation prediction based on hybrid machine learning techniques," *Energies*, vol. 11, no. 8, p. 1975, 2018.