Data-Driven Weather Prediction: Integrating Deep Learning and Ensemble Models for Robust Weather Forecasting

Hassan Al Sukhni[1,*] , Fatma Sakr[2] , Fadi yassin Salem Al jawazneh[3], Mutasem K. Alsmadi[4] , Magdy Abd-Elghany[5] , Ibrahim A. Gomaa[5] , Shaimaa Abdallah[6]

[1]Cybersecurity Department, Faculty of Science and Information Technology, Jadara University, Irbid, Jordan
[2]Department of Computer Science ,Faculty of Computers and Artificial Intelligence , Benha University, Benha, Egypt
[3]Faculty of Information Technology, Applied Science Private University, Amman, Jordan
[4]Department of MIS, College of Applied Studies and Community Service, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam, Saudi Arabia
[5]Computer Science Department, Alobour high institute for computer and informatics, Cairo, Egypt
[6]Computer Science Department, Higher Technology Institute , 10th of Ramadan City, 12311, Egypt

Emails: h.alsukhni@jadara.edu.jo; fatma.saqr@fci.bu.edu.eg; mksalsmadi@gmail.com; magdy_zeid83@oi.edu.eg; Ibraheemg@oi.edu.eg; Earth_sky_2014@yahoo.com

Abstract

Accurate weather forecasting is critical for sectors like agriculture, transportation, disaster management, and public safety. This paper presents a comprehensive methodology integrating traditional machine learning models, deep learning techniques, and ensemble learning approaches to enhance the precision and reliability of weather predictions. Using a combination of four datasets—two for classification and two for regression—the study evaluates various machine learning models such as Decision Trees, Support Vector Machines, and K-Nearest Neighbors, alongside ensemble methods like Bagging and AdaBoost. Additionally, deep learning models, particularly the Multilayer Perceptron (MLP), are employed to handle complex weather patterns. The Random Forest Regressor and Bagging Regressor consistently outperformed other models in terms of accuracy, precision, and F1-score. By comparing the performance of these models across different weather datasets, this research demonstrates the effectiveness of cross-validation and the importance of optimizing hyperparameters. The findings contribute valuable insights into enhancing the robustness and efficiency of weather forecasting systems, with potential applications in environmental monitoring, event planning, and climate change analysis.

The findings indicate that Random Forest Regression consistently outperformed the other machine learning models evaluated. For ensemble learning, the Bagging Regressor was the top performer. In deep learning, the Multilayer Perceptron without cross-validation delivered outstanding performance. For the classification datasets, Random Forest achieved the highest accuracy, precision, and F-score. Our study also highlights the importance of cross-validation to prevent overfitting and ensure model robustness, as well as the impact of class imbalance on overall performance metrics.

Keywords: Machine Learning; Deep Learning; Artificial Neural network (ANN); Ensemble learning; Multi-Layer Perceptron

DOI: https://doi.org/10.54216/JCIM.150220
Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

1 introduction

Since the beginning of the nineteenth century, individuals have been making casual forecasts regarding the weather for countless years. Weather forecasting involves using scientific and technological methods to predict atmospheric conditions for a specific location and time period. Initially, weather forecasts were manually created, relying primarily on changes in barometric pressure, current weather patterns, and cloud cover. Nowadays, forecasting utilizes computer-based models that incorporate a wide range of atmospheric variables. The process involves collecting objective data about the current atmospheric conditions at a given location and applying meteorological principles to anticipate future weather. Additionally, human expertise is essential to select the most appropriate forecast model to base the predictions on.

The problem is the need for accurate and efficient weather forecasting methods to support various sectors such as outdoor event planning, agriculture, transportation, energy production, disaster preparedness, and environmental monitoring. Reliable forecasts are necessary for informed decision-making, resource optimization, and ensuring safety and productivity in these industries. Holmstrom et al. presented an approach to forecast the highest and lowest temperatures for the subsequent seven days by utilizing historical data from the preceding two days. They utilized a linear regression model alongside an adapted version of functional linear regression. The research concluded that while their models weren't as precise as professional weather forecasting services for forecasts within a seven-day span, they exhibited enhanced performance for longer-term predictions and forecasts made for later days. Traditional weather forecasting methods depend on intricate physical models, whereas machine learning (ML) takes a data-focused approach, detecting patterns and connections in historical weather data in order to forecast future circumstances. Machine learning models can process large volumes of data in real time, enabling Consistent and accurate forecasts. Machine learning can substantially improve the precision of weather prediction models.

In weather forecasting, ensemble learning techniques play a vital role in improving predictive accuracy by combining multiple models. Methods such as bagging, boosting, and stacking merge predictions from various individual models to mitigate biases and errors. One significant application of ensemble learning in this field is probabilistic weather forecasting, where multiple forecasts from different models are generated and combined to provide probabilistic predictions.

Deep learning, a subset of machine learning, has garnered significant attention due to its flexibility and its design, which is inspired by the human brain's functioning. It consists of layers of virtual neurons. Each neuron's role is to sum the inputs it receives and determine whether to send an output signal to the next layer of neurons. Neurons in one layer connect to all neurons in the layers above and below. By learning the optimal weights for these connections, neural networks can tackle a wide variety of problems, mimicking the way the human brain works. Despite the simplicity of the neural network concept, the extensive number of connections allows for the representation of highly complex problems. This approach enhances the reliability and usefulness of weather predictions, aiding in better decision-making for various stakeholders.

We used 4 dataset 2 of them are classification and the other are regression .As for the classification datasets, the analysis compared the strengths and weaknesses of different models, such as SVM, Naïve Bayes, and Random Forest, across multiple datasets. The findings highlighted the importance of assessing model performance on diverse datasets to identify the most suitable classification technique. The consistent superior performance of the Random Forest Regression model and the advantages of ensemble techniques, like the Bagging Regressor, offer valuable insights for improving the accuracy and reliability of regression models. For Dataset 1, the algorithm with the best overall performance was the Support Vector Machine (RBF) with the highest accuracy of 73.12 percent. For the second dataset in the classification case study, the Random Forest algorithm outperformed the other models, reaching an accuracy of 95 percent.In the regression datasets, the Random Forest Regression model consistently demonstrated the highest R-squared values across both datasets, indicating its robustness for real-world regression problems.

The primary aim of our study is to improve the accuracy and dependability of weather forecasting systems by investigating advanced meteorological data analysis techniques. This research encompasses a variety of methods, which were ensemble learning including Bagging and AdaBoost , deep learning including Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and Multilayer Perceptrons (MLP) networks and finally machine learning including regression models. To achieve this, we will employ diverse datasets, such as historical weather data, real-time atmospheric observations, and climate models. Our goal

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

is to thoroughly evaluate each technique's performance across different datasets, enabling us to fine-tune our forecasting system for various weather conditions. Through this comprehensive analysis, we seek to enhance weather forecasting systems, ensuring improved precision and relevance for agricultural operations, climate change assessments, and everyday decision-making.

This paper is organized to carefully explore the different facets of our search. We discuss related work in the second section, giving a thorough rundown of current studies and approaches relevant to weather forecasting and prediction models. The third section provides specifics about our research, including the sources of the data, the preprocessing procedures, and the machine learning and deep learning methods we used. We assess the performance of the various models and talk about the consequences in the fourth section, where we give our findings and debate. The paper is finally concluded in the final section, which summarizes our results, emphasizes the importance of our work, and offers some possible directions for further study.

2 Related Work

Weather prediction data is collected from various sources, including radar, ships, and ground observations, comprising both critical and non-critical information in an unstructured format. Pandey et al. explored the use of big data in weather forecasting, utilizing the Hadoop system to process this unstructured data. Numerous techniques was developed such as fuzzy logic and the adaptive neuro-fuzzy inference system were employed to achieve precise weather predictions based on mean square error.

Numerical weather prediction models are known to require powerful computers to solve complex scientific equations for climate-based predictions. In contrast, Hewage et al. proposed a lightweight, data-driven weather prediction model that employs temporal modeling techniques such as long short-term memory (LSTM) and temporal convolutional networks (TCN). Morever, they utilized the arbitrage of forecasting expert (AFE) as a dynamic ensemble technique. By carefully synthesizing existing research, this section seeks to outline the evolution of weather forecasting methods, highlighting significant discoveries and technological advancements that collectively shape the current state of personalized weather predictions. In,[15] The purpose of this paper is to raise awareness of the use of time series analysis across various domain. Additionally, it aims to discover the utility of different time series forecasting approaches. The author has employed a wide range of methods for time series forecasting, which can be divided into five main categories:

- Regression methods
- Soft computing techniques, such as Recurrent Neural Network and Convolutional Neural Networks (CNNs)
- Stochastic approaches, where seasonality and trend are crucial characteristics of time series data, such as Autoregressive Moving Average (ARMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) models
- Fuzzy Logic Forecasting, which is used for multivariate time series data
- Autoregressive Integrated Moving Average (ARIMA) models, which have three parameters

For evaluation purposes, the author has performed three different types of time series forecasting: short-period, medium-period, and long-period forecasting. The author applied the above methods to the three discrete datasets, and observed the results to figure out:

- For short-term forecasting, the top-performing models were LSTMs, CNNs, weighted MVFTS, and CBLSTM, with ARIMA also delivering good results. Regression models, however, did not perform well for short-term predictions.
- Medium-term forecasting, regression models continued to underperform, whereas ARIMA, weighted MVFTS, and CBLSTM showed excellent results. CNNs and LSTMs also achieved moderate success.

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

- For long-term forecasting, the results were identical with slight difference to the medium-term, with ARIMA, weighted MVFTS, CBLSTM, and CNN consistently performing the best.

In,[1] The paper presents a survey on the methods used for weather prediction and discusses the challenges in forecasting. This paper applied various machine learning algorithms such as Naive Bayes and Decision Tree, as well as deep learning techniques like Convolutional Network (CNN) and Recurrent Neural Network (RNN). It also utilized common data mining techniques such as classification, clustering, and decision trees. The paper highlighted the importance of employing deep learning, data mining, and machine learning algorithms in weather forecasting. The paper highlighted that the main problems it addressed were the lack of available surveys on the current status of weather research and applications in this field, as well as the unavailability of suitable datasets to train and test the proposed methods.

In[9] ,The matter being addressed is the necessity to raise the global ensemble weather forecasts' precision and dependability, particularly in the case of extreme weather occurrences. The intricate and chaotic structure of Earth causes traditional numerical weather prediction methods to struggle, which results in forecast uncertainties. The goal is to use deep learning techniques to reduce computational expenses and increase forecast precision and reliability. Three datasets were employed by the researchers: testing (2016–2017), validation (2014–2015), and training (2019–2013). To post-process the ensemble weather forecasts, they used deep neural networks. Results revealed lower computing costs with fewer trajectories and a 14 percent improvement in ensemble forecast skill (CRPS), especially for harsh weather.

In,[8] The resource-intensive and frequently erroneous classical weather forecasting methods are discussed in the paper along with their shortcomings. It suggests a method for efficiently producing reliable forecasts by training basic machine learning models with historical data from several weather stations.The study uses two main datasets: a training set from July 2018 and a test set from September 2018, focusing on predicting hourly temperatures for Nashville, Tennessee. By leveraging data from surrounding cities, the models achieve competitive accuracy with traditional methods. According to the results, these models are less resource-intensive and can accurately anticipate the weather. Additionally, to increase forecast accuracy, the study recommends incorporating inexpensive IoT devices in the future.

In,[10]Since weather patterns are dynamic and meteorological data is complicated traditional weather forecasting systems are not very accurate. To improve short-range forecasting, create a dependable weather prediction model utilizing a hybrid C5.0 machine learning strategy. utilized meteorological data spanning the years 1980 to 2020 from the MERRA database, with an emphasis on variables such as temperature, humidity, pressure, rainfall, and wind speed. applied K-means clustering in conjunction with the C5.0 algorithm to forecast the weather. In comparison to other algorithms, the model's predicted accuracy was 90.18 percent, and its MAE and RMSE were lower. outperformed alternative methods with gains in 1.8 to 6.9 percent, 2.1 to 7.2 percent, and 1.7 to 4.6 percent, respectively, in accuracy, sensitivity, and specificity.

In[11] , The main purpose of this paper is to explore the effectiveness of data-driven models, convolutional neural networks, and ensemble neural network forecasts in improving weather and climate predictions. The training process utilized a total of 32 slightly different DLWP models generated by eight training cycles with four members drawn from the checkpoints of each cycle. Each training cycle contributed four members to the ensemble, resulting in a total of 32 models used in training. The author has used convolutional neural networks (CNNs) to generate data-driven global forecasts. Specifically, the Deep Learning Weather Prediction (DLWP) model employed CNNs on a cubed sphere grid to produce global forecasts efficiently. By using a cubed sphere grid, the DLWP model is able to generate global weather forecasts efficiently and without some of the distortions or numerical issues that can arise from using a traditional latitude-longitude grid. The ECMWF ensemble outperformed the DLWP ensemble in terms of probabilistic measures of ensemble skill, such as the continuous ranked probability score (CRPS) and the ranked probability skill score (RPSS). Overall, the challenges faced by the DLWP ensemble include inadequate representation of uncertainty, Suboptimal spread of ensemble members and lower performance compared to established NWP models like ECMWF in terms of probabilistic skill scores.

As mentioned in[3] , the primary limitation of this paper is the small size of the dataset, which may not be sufficient for a robust analysis. Additionally, the dataset was collected manually and does not include all relevant publications in the field. The main objective of this paper was to apply machine learning techniques to numerical weather prediction. As mentioned, the paper utilized a single dataset comprising 500 scientific

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

articles related to machine learning in numerical weather prediction and climate analysis. The authors explored a variety of machine learning techniques for numerical weather forecasting, including both supervised and unsupervised learning methods.

In the supervised learning group, the authors evaluated various techniques such as Decision Trees, Random Forests, XGBoost, Artificial Neural Networks (ANNs), Deep Learning, and Support Vector Machines (SVM).In the unsupervised learning group, they utilized K-Means Clustering and Principal Component Analysis (PCA). Based on the authors' experience, it appears that all of these methods can be applied to successful numerical weather prediction and climate analyses. However, the most powerful algorithms were ANNs and Deep Learning, suggesting these techniques may be particularly well-suited for this domain.

In,[5]The field of weather analysis has seen significant advancements through the application of connectionist learning paradigms. This paper builds upon the foundational work of Maqsood, Khan, and Abraham[1], who conducted a comparative study of neural network models for forecasting weather patterns in Vancouver, Canada. Their research demonstrated the superior forecasting capabilities of Radial Basis Function Networks (RBFN) over Multi-Layered Perceptrons (MLP) and Elman Recurrent Neural Networks (ERNN), using a dataset of daily maximum and minimum temperatures and wind speed.

The aim of our research is to extend these findings by employing an ensemble of connectionist learning paradigms to improve the accuracy and reliability of weather forecasts. Ensembles have been shown to enhance predictive performance by combining the strengths of diverse models, thereby reducing the likelihood of overfitting to a particular dataset. Our approach is inspired by the success of ensemble methods in various domains, including the work of Dietterich[2], who highlighted their effectiveness in reducing errors in complex tasks.

In pursuit of this goal, our paper seeks to address the challenges inherent in weather prediction, which remains a complex and chaotic domain, as noted by Lorenz[3]. By integrating multiple connectionist models, we aim to capture the non-linear and dynamic nature of weather systems more effectively than single-model approaches. In,[6] The reliability of power distribution systems against environmental factors is a critical area of research, particularly concerning weather-related outages. The paper "ADABOOST+: An Ensemble Learning Approach for Estimating Weather-Related Outages in Distribution Systems" builds upon the significant body of work that has explored various predictive models for outage estimation. The ensemble learning approach, particularly the AdaBoost algorithm, has been recognized for its robustness and accuracy in handling complex prediction tasks.

This research is inspired by the pioneering work of Freund and Schapire1, who introduced the AdaBoost algorithm, demonstrating its effectiveness in improving the performance of weak learners. The application of AdaBoost in power systems was further explored by Kankanala, Das, and Pahwa2, who proposed the ADABOOST+ model to estimate wind and lightning-related outages. Their model showed a marked improvement in accuracy over traditional regression and neural network models, using actual data from four cities in Kansas.

The objective of our paper is to enhance the ADABOOST+ model by incorporating additional environmental factors and extending the dataset to a broader geographical area. By doing so, we aim to provide a more comprehensive and accurate tool for utility companies to predict and mitigate the impact of weather-related outages on power distribution systems. In,[7]The accurate forecasting of maximum air temperatures is a pivotal challenge in meteorology, especially given the implications of extreme heat events on public health and energy consumption. This paper, "A Novel Ensemble Learning for Post-Processing of NWP Model's Next-Day Maximum Air Temperature Forecast in Summer Using Deep Learning and Statistical Approaches," contributes to the growing body of literature that seeks to refine the predictive capabilities of Numerical Weather Prediction (NWP) models.

Our work is particularly influenced by the ensemble learning techniques that have been successfully applied in various meteorological forecasting scenarios. The concept of ensemble learning was notably advanced by Zhou[1], who demonstrated that an ensemble of multiple learning algorithms could achieve better predictive performance than any of the constituent algorithms alone. This principle has been applied to weather forecasting by researchers such as Rasp and Lerch[2], who utilized deep learning models to post-process ensemble forecasts, significantly improving the prediction of precipitation.

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

In the context of temperature forecasting, the study by Cho and Yoo[3] is of particular relevance. They proposed a novel multi-model ensemble that aggregates individual post-processing models based on skill scores for the Local Data Assimilation and Prediction System (LDAPS) model's next-day maximum air temperature forecast data in South Korea. Their approach underscores the potential of ensemble methods in enhancing the accuracy of temperature predictions.

An ensemble of data-driven weather prediction models for practical sub-seasonal forecasting is presented by the authors in.[12] The ECMWF ocean model and hybrid data-driven models are combined by the authors to forecast global weather at a 1-degree resolution for a maximum of four weeks. In terms of 2-meter temperature prediction, the ensemble outperforms the raw ECMWF extended-range ensemble by an average of 4–17 percent. At four weeks, statistical bias corrections increase the accuracy of the ECMWF ensemble by around 3 percent. The dependability of the ensemble is assessed using CRPS scores, rank histograms, and geographic biases.

The study shows that a multi-model ensembling method using data-driven models can provide near-state-of-the-art sub-seasonal forecasts. It highlights the requirement of utilizing a variety of model designs and tackles the shortcomings of current models, such as their inability to provide extended-range forecasting. ECMWF ERA5 reanalysis, which uses surface parameters and atmospheric variables as inputs, is included in the training data.

The competitive performance of the ensemble highlights the significance of ensemble reliability for probabilistic forecasting, since it is on par with state-of-the-art models. The study emphasizes how merging traditional and data-driven NWP models can improve forecast accuracy. Prospective research avenues encompass tackling exceptional occurrences, enhancing precipitation predictions, and refining ensemble generation techniques.

In,[13] the authors introduced an AI-based global weather forecasting system called Pangu-Weather. Pangu-Weather outperforms traditional Numerical Weather Prediction (NWP) techniques in terms of accuracy and speed by utilizing deep neural networks that have been trained on 39 years' worth of weather data. The input and output method used by the system is 3D (latitude, longitude, height), and each model has been trained for a distinct prediction time interval. As a result, fewer iterations are needed to provide medium-range projections. When it comes to large-member ensemble forecasts and the prediction of extreme weather occurrences, Pangu-Weather performs better than other models. Pangu-Weather obtains impressive results by combining hierarchical temporal aggregation and the 3D Earth Specific Transformer architecture. This demonstrates how artificial intelligence (AI) can improve weather forecasts beyond what can be achieved with conventional NWP.

The authors emphasize training and validation using reanalysis datasets, particularly ERA5. They stress that increasing forecast accuracy requires a focus on data dimensionality and processing efficiency. Pangu-Weather performs best when using individual models for various time intervals and the 3D data method. The results show that Pangu-Weather can generate precise, high-resolution forecasts that closely resemble real data. On benchmark datasets, it demonstrates improvements in RMSE and beats operational IFS in terms of overall accuracy. Additional validation of the system's upper-air variable prediction accuracy comes from diagnostic investigations. In comparison to current NWP and AI-based techniques, Pangu-Weather, a ground-breaking AI-based weather forecasting system, performs better, as the article demonstrates. The approach's potential to improve weather prediction capabilities is highlighted by the authors.

The "WeatherBench" benchmark dataset is presented by the authors in[14] as a tool for data-driven, medium-range global weather prediction. In order to facilitate quick advancement in this area, the authors want to offer a benchmark with solid baselines. The 3-5 day temperature and pressure field estimates in the dataset, which is prepared for easy integration into machine learning models, are taken from the ERA5 archive. As baselines, the authors suggest deep learning models, simply physical forecasting models, and basic linear regression in order to enable easy comparisons between various techniques. The authors point out that deep learning and other data-driven methods have the potential to generate more accurate forecasts by directly learning from observations, while traditional weather models have limitations in some applications. The medium-range forecast scenario, which is the focus of the benchmark, is significant for societal applications but presents a difficult problem for data-driven approaches due to its complicated atmospheric dynamics.

The aim of our research is to build upon these methodologies by integrating deep learning and statistical approaches to post-process Numerical Weather Prediction model forecasts. By doing so, we strive to create a robust framework that can more accurately predict next-day maximum air temperatures during summer

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

months, which are critical for planning and response to potential heatwaves. In[22] ,Ensemble weather forecasts provide a measure of uncertainty for each prediction by calculating the ensemble's spread. However, achieving a good spread-error relationship is challenging, and numerous methods have been explored, primarily within numerical weather prediction models. In this study, we aim to convert a deterministic neural network weather forecasting system into an ensemble forecasting system. We evaluate four methods for generating the ensemble: random initial perturbations, retraining the neural network, applying random dropout within the network, and using initial perturbations created through singular vector decomposition. While the latter method is commonly used in numerical weather prediction models, it has not yet been tested with neural networks. The ensemble mean forecasts from all four methods surpassed the unperturbed neural network forecasts, with retraining providing the greatest improvement. Nonetheless, the skill level of the neural network forecasts remains consistently lower than that of state-of-the-art numerical weather prediction models. In[21] , In this paper, we propose the RSEL (Random Subfeature Ensemble Learning) algorithm to enhance weather forecasting accuracy. The RSEL algorithm integrates random subfeature selection with an ensemble learning strategy based on classical machine learning algorithms, increasing feature diversity and mitigating the impact of randomly generated unstable outliers. Additionally, we design feature engineering schemes for weather forecast data to fully utilize spatial and temporal context. We test the RSEL algorithm by forecasting wind speed and direction, demonstrating improved accuracy and robustness compared to traditional methods. In[20] , Climate change research involves analyzing varying weather patterns over time. Rainfall forecasting predicts future rainfall amounts based on past data such as wind, humidity, and temperature. Recently, various machine learning (ML) techniques have been employed for rainfall forecasting, achieving different levels of performance for short-term and long-term predictions. Despite numerous ML methods proposed for enhancing rainfall forecasting, selecting the appropriate technique for specific rainfall durations remains unclear. This study proposes an ensemble learning approach to improve rainfall prediction effectiveness. Ensemble learning combines multiple ML classifiers, including Naïve Bayes, Decision Tree, Support Vector Machine, Random Forest, and Neural Network, using Malaysian data. We explore three algebraic combiners: average probability, maximum probability, and majority voting. Our analysis shows that fused ML classifiers using majority voting significantly enhance rainfall prediction performance compared to individual classifiers.

3 Methodology

Figure 1 , and Algorithm 1 the proposed framework for Weather Forecasting, which contains mainly eight significant steps

- Data collection
- Data preprocessing
- Data splitting
- Feature extraction
- Optimizers for Deep Learning
- Machine Learning Algorithms
- Ensemble Learning and Deep Learning Algorithms
- Prediction and evaluation metrics

3.1 Data collection

The collection of datasets is available for examining weather patterns and conditions in various locations thanks to this dataset collection. During a four-month period in 2014, the Mauna Loa Daily Temps Dataset provides a detailed look at temperature variations, sunrise and sunset timings in Hawaii. More extensive, long-term data on a range of meteorological parameters, such as precipitation, wind, humidity, and barometric pressure, for

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

Algorithm 1 Weather Prediction Methodology

Input: Historical weather data $\{D_{train}, D_{test}\}$, Machine Learning Algorithms $\{ML1, ML2, ..., ML_n\}$, Ensemble Learning Algorithms $\{EL_1, EL_2, ..., EL_m\}$
Output: Predicted weather conditions
procedure DATA COLLECTION
    Collect datasets such as Mauna Loa Daily Temps Dataset, Seattle Weather Dataset
end procedure
procedure DATA PREPROCESSING
    Handle missing data, clean and normalize datasets
    Apply scaling or standardization if required
end procedure
procedure DATA SPLITTING
    Split data into training set D_{train} and test set D_{test}
end procedure
procedure FEATURE EXTRACTION
    Extract meaningful features using PCA, LDA, or other techniques
    Select important features for training models
end procedure
procedure MODEL TRAINING AND OPTIMIZATION
    for each machine learning algorithm $ML_i \in \{ML_1, ..., ML_n\}$ **do**
        Train ML_i on D_{train}
        Optimize hyperparameters using Grid Search
        Evaluate using cross-validation
    end for
end procedure
procedure ENSEMBLE LEARNING
    for each ensemble learning algorithm $EL_j \in \{EL_1, ..., EL_m\}$ **do**
        Apply Bagging, Boosting, or Random Forest for ensemble
        Train ensemble model on top of base learners
    end for
end procedure
procedure PREDICTION AND EVALUATION
    Generate predictions using the trained model on D_{test} State-evaluate model using metrics like RMSE, Accuracy, etc.
end procedure

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

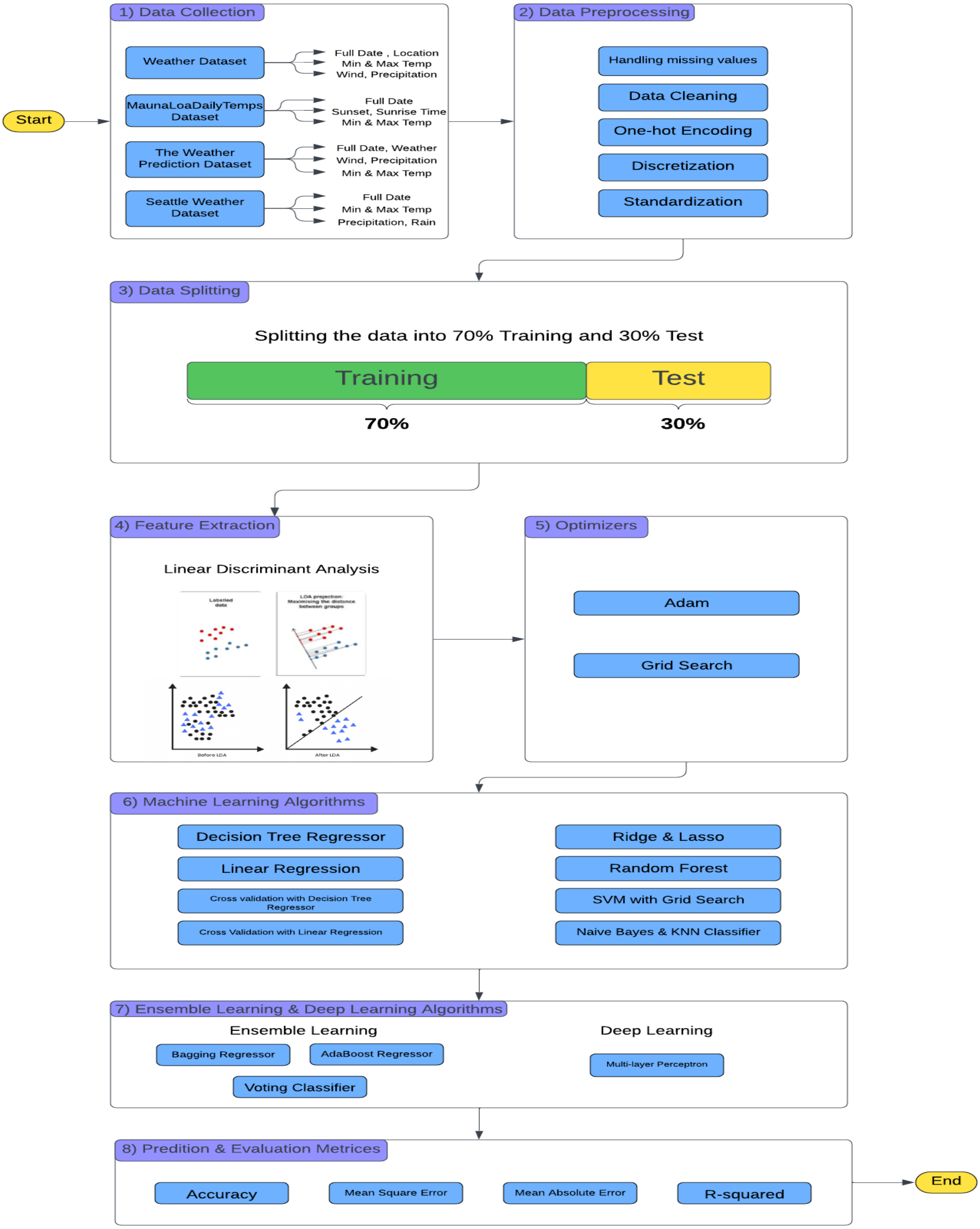

Figure 1: The Proposed System of Weather Forecasting.

larger geographic areas are probably available in The Weather Dataset and The Weather Prediction Dataset. By permitting comparisons and contrasts with data from Mauna Loa and other sites, the Seattle Weather

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

Table 1: Regression Datasets Information

Dataset Name Size	Year	Source
Weather 1750KB	2017	Github
MaunaLoaDailyTemps 49.8KB	2020	Github

Table 2: Classification Datasets Information

Dataset Name Size	Year	Source
The Weather Prediction 138KB	2019	Kaggle
Seattle Weather 144KB	2021	Kaggle

Dataset may provide insights into the distinct climate of the Pacific Northwest. When combined, these four datasets offer a strong basis for studying regional and global weather patterns, the effects of climate change, and the creation of more precise forecasting models. This collection of weather-related datasets contributes significantly to the scientific community's efforts to understand, simulate, and predict the dynamics of our planet's atmosphere. This data has numerous potential applications, including advancements in industries like as agriculture, transportation, and public health.

Tables 1 and 2 provides detailed information about the datasets used in our analysis.

1. **Weather Dataset** :[16] This weather dataset captured a wide range of meteorological characteristics from various locations and time periods that has been assembled from a network of weather stations. The main objective is to use this dataset to advance machine learning techniques to improve weather forecasting accuracy. This dataset offers a comprehensive look at the weather for January 3, 2016 to January 1, 2017, across the United States, with a particular emphasis on Alaska and Alabama. The meteorological information for a particular station, including measurements for precipitation, average temperature, maximum and lowest temperatures, wind direction, and wind speed, is represented by each row in the dataset. The stations are spread over a wide range of towns and cities, from larger, more populated areas like Phoenix, Arizona, and Tucson, Arizona, to smaller, more isolated places like Puntilla Lake, Alaska. Investigating the regional trends and examining the correlations between the various weather factors may provide significant new information. Additionally, forecasting models and weather-dependent activity planning may benefit from a comparison of the current and past weather data. All things considered, the Weather dataset offers a comprehensive and varied perspective of the US meteorological environment on a particular date, which has the potential to lead to significant discoveries and wise decision-making.

2. **MaunaLoaDailyTemps Dataset** :[17] This dataset covers the period from January 1, 2014, to an undisclosed date in April 2014, and includes comprehensive temperature and sunrise/sunset data for Mauna Loa, Hawaii. Together with the daily sunrise and sunset timings, the data also includes the lowest, maximum, and average temperatures. With the coldest temperatures in January and the warmest in April, the temperature data exhibits a normal seasonal pattern. Minimum temperatures vary, with January 6th being the lowest at 22°F and April 18 and 19th being the highest at 43°F. The highest recorded temperatures are 62°F on January 25 and a low of 36°F on January 29. There is a seasonal trend in the dawn and sunset hours as well; on January 1st, the earliest sunrise is at 6:54 AM, and on January 16th and 17th, the latest sunrise is at 7:00 AM. January 1st at 5:56 PM is the earliest sunset, and April 1st at 6:42 PM is the latest. Overall, this dataset offers a thorough examination of the daily high and low temperatures at Mauna Loa throughout a four-month period. Both vacationers arranging outdoor activities and researchers researching the climatic and environmental conditions in this area may find it helpful. With minimum, maximum, and average temperature values, this resource is quite detailed and beneficial for comprehending the local weather trends.

3. **The Weather Prediction Dataset** [18] : This dataset offers comprehensive meteorological data for many months in 2012 to 2015 for Seattle, Washington. Daily measurements of precipitation, maximum and

DOI: https://doi.org/10.54216/JCIM.150220
Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

lowest temperatures, wind speed, and weather conditions are all included in the data. The data's depiction of weather patterns reveals Seattle's normal temperate, wet environment. There is a lot of precipitation; most days there is rain or drizzle. During the winter, snowfall is also frequent, with some days seeing a substantial buildup. The coldest months are mid- to late-January, when temperatures vary from about -3°C to 16°C. According to the data, Seattle has a wide range of daily temperatures, with notable variations between the highs and lows.There are many different kinds of weather conditions included in the data, such as sun, rain, drizzle, and snow. This variation emphasizes how unpredictable Seattle's climate can be, even over a brief period of time. As a whole, this dataset offers a thorough overview of the weather in Seattle in the winter and early spring of 2012 till 2015. It might be helpful for analysis pertaining to regional climate, weather patterns, or seasonal trends.

4. **Seattle Weather Dataset** [19] : This is another dataset that also contains daily weather measurements for Seattle, Washington but here on a longer period ,starting from January 1, 1948 till December 31, 2017.The date, the amount of precipitation in inches (PRCP), the highest and minimum temperatures in Fahrenheit (TMAX and TMIN), and a boolean indicator indicating whether or not it rained that day (RAIN) are all included in the data. The dataset enables the examination of long-term trends and patterns by offering a thorough record of Seattle's weather over a 70-year span. With more than 25,000 unique data points, scientists could look into seasonal variations, pinpoint extreme weather occurrences, and assess how climate change is affecting the local weather. Fine-grained investigations are made possible by the daily observations' granularity, such as investigating the connection between precipitation and temperature extremes. For meteorologists, climatologists, urban planners, and anybody else interested in learning about the climate and weather history of the Seattle area, this historical dataset is a great resource. We can learn things from the analysis of this data that help in forecasting, designing infrastructure, and getting ready for changing environmental circumstances.

3.2 Data Preprocessing

Regarding the Data Preprocessing stage, we have used a thorough technique to clean up and set up the four weather datasets for further examination. This required Handling Missing Values, One-hot Encoding, Data Cleaning, Discretization, and Standardization, among other important procedures. To improve the precision and dependability of our ensuing weather analysis and model projections, this exhaustive preprocessing step is essential.

- **Handling Missing Values:** One of the most important steps in data preprocessing to guarantee a dataset's quality and dependability is handling missing values. Missing values can occur for a number of reasons, including incorrect data entry, broken equipment, or information that was not recorded. If ignored, they may create bias, lower a model's efficacy, and produce false conclusions. There are numerous approaches to dealing with missing values. Eliminating characteristics or records with missing data is a popular method, although it might result in a large loss of data. As an alternative, statistical approaches can be used to fill in missing values through imputation methods. More complex techniques, such as regression imputation or k-nearest neighbors, employ patterns in the data to anticipate and fill in missing values. Simple imputation might replace missing values with the column's mean, median, or mode. Complex methods such as multiple imputation generate numerous credible datasets and aggregate the outcomes for a more comprehensive study. Every approach has benefits and drawbacks. The choice of method is contingent upon the particular requirements of the analysis as well as the characteristics of the data. Handling missing values correctly contributes to preserving the dataset's integrity, which results in predictive models that are more trustworthy and accurate.

  For instance, we used a function for dropping duplicates to eliminate duplicate entries from our datasets, ensuring that the dataset was free of redundant items that would have distorted the study. We also used a function that return the number of missing values to check for missing values, which helped us find any columns that had data gaps that needed to be filled in.

- **Data Cleaning:** Data cleaning is a crucial step in data preprocessing that involves identifying and repairing flaws and inconsistencies in the dataset in order to improve its quality. This procedure guarantees

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

that the data is correct, complete, and appropriate for analysis. Data cleaning duties include deleting duplicate records, fixing errors in data entries, addressing missing information, and filtering out extraneous data. It also entails dealing with outliers, which might distort the analysis and lead to inaccurate results. Outliers are detected , eliminated or treated to mitigate their impact. Data cleaning frequently involves putting data into a standard format, ensuring that all variables are properly formatted and labeled. This phase is crucial because it prepares the dataset for further analysis and modeling while also ensuring that the insights generated from the data are legitimate and dependable.

  – **Removing Outliers using IQR:** Identifying and handling extreme values in a dataset is a typical task that can be accomplished by applying the Interquartile Range (IQR) to remove outliers. The intermediate 50 percent of the data is represented by the IQR, which is the range between the first quartile (Q1) and the third quartile (Q3). We compute Q1 and Q3, then subtract Q1 from Q3 to get the IQR in order to find outliers. Data points that fall below Q1 - 1.5IQR or above Q3 + 1.5IQR are commonly referred to be outliers. We can successfully discover values that are far higher or lower than the majority of the data by using these thresholds. To stop these outliers from distorting the outcomes of further study, they can then be eliminated or otherwise dealt with. This strategy is especially helpful because it is robust for datasets with non-normal distributions and is unaffected by the data's mean or standard deviation. By eliminating outliers with the IQR, prediction models become more accurate and the data integrity is preserved.

- **One-hot Encoding:** One method for converting categorical information into a format that machine learning algorithms may utilize to make better predictions is called one-hot encoding. Each category of a categorical feature is created into a new binary column throughout this process. In the original data, each column corresponds to a single, distinct category value. For instance, if the feature "Color" comprises three categories, namely "Red," "Green," and "Blue," then one-hot encoding will generate three new binary columns, one for every color. When a row's "Color" feature value is "Red," its values in the "Red" column and the "Green" and "Blue" columns are zero. This technique guarantees that machine learning models regard each category as a unique entity and successfully eliminates the ordinal link that the models can erroneously infer from numerical labels. We carried out one-hot encoding on particular category columns using the ColumnTransformer and OneHotEncoder from sklearn. 'Data.Temperature.Avg Temp' and 'Date.Full' were removed from one of the datasets in order to produce the resulting feature matrix X and target variable Y.

- **Discretization:** Discretization is the process of transforming continuous numerical information into discrete categories in data preprocessing. When continuous data needs to be divided into meaningful intervals, this method can help. This makes the data easier to understand and frequently more appropriate for specific kinds of analysis or machine learning techniques. It is possible to discretize a continuous variable like age, for instance, into age groups like "0-18," "19-35," "36-60," and "60+." By converting continuous data into categorical data, discretization aids in the reduction of model complexity, which can streamline the modeling procedure and enhance the interpretability of the outcomes. Additionally, it helps manage noise and outliers in the data, strengthening the analysis. Discretization techniques include quantile-based discretization, which divides data into intervals with an equal number of observations, and binning, which divides data into fixed-size intervals.In our regression datasets, We discretized the continuous target variable into quantiles for classification tasks. pd.qcut was used to divide the target variable into four quantiles and convert the regression problem into a classification problem.

- **Standardization:** The process of standardizing features in a dataset to give them a mean of zero and a standard deviation of a statistical technique in data preprocessing. By ensuring that every feature contributes equally to the analysis, this procedure keeps features with bigger scales from controlling the performance of the model. Since standardization normalizes the data range, it is especially crucial when the data characteristics have varying sizes or units. This method is frequently used in machine learning algorithms where the scale of the data affects the distance between data points or the convergence of gradient descent optimization techniques, such as principal component analysis, support vector machines, and k-nearest neighbors. For example, in our datasets, we applied standardization on the data using StandardScaler function.

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

3.3 Data splitting

One essential step in getting a dataset ready for machine learning is data splitting. In order to assess the efficacy and generalization of prediction models, the data is split into distinct training and testing sets. Usually, the data is divided into two parts: one part is utilized to train the model and the other part is kept for testing. The dataset was split in this instance into 70 percent for training and 30 percent for testing.

The machine learning model is trained on the training set, which enables it to identify patterns and connections in the data. Contrarily, the test set is used to examine the model's performance on untested data, offering an objective appraisal of the model's propensity to perform well on fresh, real-world data. By using this method, it is possible to prevent the model from overfitting to the training set and to improve its generalization to different datasets. Creating robust and dependable machine learning models requires splitting the data in this manner.

3.4 Feature Extraction

A crucial step in machine learning is feature extraction, which is turning unprocessed data into a collection of insightful and useful features that a model may use. Reducing the dimensionality of the data while keeping the most important information is the aim. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two methods frequently used in this procedure that find and use the data's underlying structure to generate new features. By concentrating on the most crucial elements of the data, feature extraction can boost the model's interpretability, lower computing costs, and increase model performance.

- **Linear Discriminant Analysis (LDA):** A machine learning technique called linear discriminant analysis (LDA) is utilized for dimensionality reduction and feature extraction. It works especially well for jobs involving classification. Finding the linear combinations of the initial attributes that most effectively divide two or more classes is how LDA operates. It achieves this by making sure that the predicted classes are as distinct as feasible by optimizing the ratio of within-class variation to between-class variance. New characteristics that are linear discriminants—the best kind for differentiating between classes—are produced by this process. LDA improves class separability in addition to lowering the dimensionality of the data, which facilitates more accurate performance of classification methods. In our datasets, we transformed the scaled data compatible with the LDA into dense arrays. To minimize the datasets' dimensionality while maintaining their ability to discriminate between classes, LDA was used. This reduction in dimensionality was very helpful in raising the effectiveness and performance of model training.

3.5 Optimizers

In machine learning and deep learning, an optimizer is an essential algorithm that modifies a model's weights to reduce its loss function and improve the model's performance. Optimizers are essential to deep learning because they help identify the ideal set of weights by quickly and effectively exploring the intricate, high-dimensional parameter space. One of the most widely used optimizers is the Adam (Adaptive Moment Estimation) optimizer, which maintains distinct learning rates for each parameter and modifies them according to their first and second moments of the gradients. It combines the advantages of two other extensions of stochastic gradient descent, namely AdaGrad and RMSprop. Because of this, Adam works especially well on challenges involving large amounts of sparse data. Also, Grid Search is a significant methodology for optimizing machine learning models. It is a methodical approach that investigates many combinations of hyperparameters to identify the optimal configuration for a model. Grid Search assists in determining which hyperparameters produce the best accuracy or the least amount of error by using cross-validation to assess each combination's performance.

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

- **Adam Optimizer:** A popular optimization approach in deep learning, the Adam optimizer (short for Adaptive Moment Estimation) works especially well when training multilayer perceptrons (MLPs). Adam combines the benefits of RMSProp and AdaGrad, two additional well-liked versions of stochastic gradient descent (SGD). By keeping two moving averages, one of the gradients (which helps smooth out oscillations) and the other of the squared gradients (which helps scale the learning rate), it calculates adaptive learning rates for each parameter. Adam's adaptive nature makes it ideal for sophisticated neural network topologies such as MLPs, as it can perform well on noisy situations and sparse gradients. Adam facilitates the effective navigation of the high-dimensional parameter space when combined with MLPs, resulting in faster convergence and frequently greater performance than traditional SGD. Adam's efficiency and resilience make it the go-to option for many deep learning applications.
- **Grid Search:** In machine learning, Grid Search is a hyperparameter optimization method that is used to methodically examine a predetermined collection of hyperparameter values. Grid Search maximizes the model's predicted accuracy or minimizes its error by building a grid of possible values for each hyperparameter and assessing model performance for each combination. In order to achieve strong and dependable findings, this thorough search procedure entails training and testing the model numerous times across various combinations, frequently utilizing cross-validation. Grid Search is useful for fine-tuning algorithms such as support vector machines, neural networks, and ensemble approaches, even though it requires a lot of processing power.

3.6 MACHINE LEARNING MODELS

As machine learning algorithms can examine intricate patterns and relationships within large datasets, they are essential for weather prediction. Several models are frequently used for this purpose. A non-linear model called the Decision Tree Regressor divides data into subsets according to feature values and then builds a tree-like structure to provide predictions. Fitting a linear equation to represent the relationship between dependent and independent variables is a simple method known as linear regression. An ensemble technique called Random Forest constructs several decision trees and averages them to improve accuracy and decrease overfitting. By penalizing high coefficients, the regularization techniques Ridge and Lasso are applied to linear regression to address multicollinearity and enhance model robustness. Regression can be carried out by Support Vector Machines (SVM) by identifying a hyperplane that maximizes the margin and best fits the data. Because of its ease of use and effectiveness, the probabilistic classifier Naive Bayes, which is based on Bayes' theorem, is useful for classification problems in weather prediction. Last but not least, K-Nearest Neighbors (KNN) is a simple and straightforward non-parametric algorithm that forecasts the value of a new observation based on the values of its closest neighbors in the feature space. These models are useful resources in the complex field of weather prediction since they each have specific advantages.

- **Machine Learning Models**
  - **RBF SVM**
    Regression and classification tasks in weather forecasting can both benefit from the highly effective machine learning method known as the Radial Basis Function Support Vector Machine (RBF SVM). This non-parameterized model works effectively with the multidimensional weather data that is often non-linear. The mechanism by which the RBF SVM operates is by projecting the weather input data onto a higher-dimensional feature space, from which the various weather classes may be divided using a hyperplane. The approach calculates the similarity between pairs of weather data points in this feature space using a kernel function, such the Radial Basis Function. In RBF SVM, the Radial Basis Function is a frequently used kernel function. It has the following definition:
    $K(x, x')$ =$(exp(-gamma||x - x'||)$
    The weather input data points are denoted by x and x', the Euclidean distance between the points is represented by ——.——, and gamma is a hyperparameter that regulates the kernel's width. Using their distance in the feature space as a basis, the similarity between pairs of weather data points is determined by this kernel function. The gamma parameter in the RBF SVM intuitively defines the reach or influence of a single training weather example. Greater gamma values denote a more concentrated impact, whereas lower gamma values signify a wider influence. gamma determines

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

the extent to which each support vector influences the weather forecasting decision boundary. The balance between accurately categorizing the training weather data and maximizing the decision function's margin is represented by the RBF SVM's C parameter. If a larger value of C leads to a better categorization of the training weather instances, then a smaller margin will be accepted. On the other hand, a smaller C value will promote a wider margin, which might result in a simpler weather forecasting decision function at the expense of training accuracy. The C parameter serves as the SVM's regularization parameter in this manner. Finding the ideal hyperparameter values for a particular model is accomplished through hyperparameter adjustments. The optimal hyperparameter values cannot be determined in advance, therefore making manual adjustments would require a lot of time and resources. To speed up this procedure, GridSearchCV is employed. To evaluate each possible combination of the hyperparameter values listed in a dictionary, GridSearchCV uses the Cross-Validation approach. After that, it evaluates the model's performance for every combination, letting you choose the combination of hyperparameters that produces the highest accuracy.

– **K-Nearest Neighbors (KNN)**
A machine learning approach known as the k-nearest neighbor (KNN) method uses learning outcomes to classify an item according to its k nearest neighbor .This is a straightforward classification approach that is considered simple to apply in data analysis with high-dimensional data. The training data is mapped into a multi-dimensional space, where each dimension represents a data feature. This space is then divided into regions based on the training data labels. A new data point is classified as belonging to the class that is most common among its k nearest neighbors in the feature space. Despite being straightforward, the KNN approach has a benefit versus other methods since it can efficiently generalize from a small training sample. Typically, a distance metric, such the Euclidean distance, is used to determine how close or far apart neighbors are. KNN often applies the idea of "majority voting" when dealing with issue statements of the "classification" category. The class with the highest number of votes is selected within the specified range of K values. As for the "regression" category, KNN predicts the value of fresh data using the mean/average technique. All the closest neighbors would be taken into account based on the value of K. Until it has located every closest neighbor within a specific range of the K value, the algorithm tries to get the mean for each of the numbers of the nearest neighbors. After gathering related weather-related information, the dataset is split into training and testing sets to make model assessment easier. Initializing the KNN model and setting its hyperparameters. Cross-validation aids in this optimization process, which finds the model configurations that optimize predicted performance. The effectiveness of the KNN technique's capacity to apply generalization on limited training datasets and its simple, intuitive behavior of identifying weather conditions based on the properties of surrounding data points are its main benefits for weather prediction. Through adherence to this all-inclusive sequence of steps, which include data preprocessing, model optimization, and performance assessment, the KNN approach can yield significant insights and enhance weather forecasting systems.

– **Random forest**
Due to its ease of use and adaptability, the Random Forest method has become quite popular in weather forecasting. It can successfully handle issues involving both classification and regression. The algorithm's strength is its capability to manage intricate datasets while reducing the chance of overfitting, which makes it an invaluable instrument for a range of machine learning prediction tasks. The versatility of the Random Forest technique to handle datasets with both continuous variables—as in regression—and categorical variables—as in classification—is one of its primary characteristics. This flexibility allows the algorithm to perform well on a variety of machine learning tasks, including weather forecasting. In an ensemble, such as the trees within a Random Forest, decision trees are typically trained using a technique known as "bagging." This method, also called Bootstrap Aggregation, involves combining predictions from multiple machine learning algorithms to enhance predictive accuracy beyond what a single model can achieve. Random Forest, being an ensemble method itself, operates on this principle. Bootstrap randomly selects rows and features from the dataset to create sample datasets for each model. The process of aggregation then combines and summarizes the findings to create a more cohesive sample. The Bootstrap Aggregation, sometimes known as "bagging," is especially useful for lowering the variation of highly variable algorithms, such as decision trees, in weather forecasting scenarios. Errors brought on by sensitivity to minor variations in the training dataset are referred to as variance. Excessive variance can cause an algorithm to represent noise—irrelevant data—instead of signal, or the desired results. The phenomenon, also known as overfitting, arises when a model works well in training but has

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

trouble telling the difference between signal and noise in practical testing situations. The bootstrap approach is used to a high variance machine learning algorithm through a process called bagging, particularly in weather forecasting applications.

– **Gaussian Naive bayes**
A probabilistic machine learning classification method called Gaussian Naive Bayes makes the assumption that every class has a normal distribution. It functions on the presumption that every input parameter has the capacity to independently predict the output variable. In the process of weather forecasting, one may forecast the chances that a dependent variable will fall into each category using Gaussian Naive Bayes. In order to achieve this, it adds up all of the input parameter predictions, which provides a chances for the dependent variable to fall into each potential group. The group with the highest anticipated probability is then given the final classification. The normal distribution is another name for the Gaussian distribution. The normal distribution, which is typified by a curve with a bell shape, is a statistical model that depicts the distributions of continuous variables that are random in nature. The mean μ and standard deviation σ are the two key components of the normal distribution. A distribution's "width" around the mean is known as the standard deviation, while the mean is the distribution's average value. The overall area under the model curve is one for a normally distributed variable X, which is distributed continuously (continuous variable) from $-\infty < X < \infty$.

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

– **Decision Tree**
The Train Using AutoML software employs decision trees, a supervised machine learning technique, to classify or regress data based on specific queries' positive or negative responses. This tree-like structure comprises three types of nodes: root, internal, and leaf. It starts at the root node and branches out to the internal and leaf nodes, where leaf nodes represent the final classification categories or actual values. Decision trees are known for their clarity and interpretability. The process begins by selecting a feature to serve as the root node. Impurities occur because a single feature often can't perfectly predict the final classes. Methods such as Gini, entropy, and information gain are used to measure this impurity and assess how effectively a feature classifies the data. At each step, the feature with the lowest impurity is chosen as the node. For numerical features, the Gini impurity is calculated by first sorting the data and finding the mean of adjacent values. Data points are then ordered based on whether their feature values are higher or lower than a specified threshold, determining if this separation accurately classifies the data. The Gini impurity is calculated using the formula: Gini(p) =

$$1 - \sum_{i=1}^{n} p_i^2$$

where p is the proportion of each classification category and K is the total number of categories. The weighted average of the Gini impurity for the leaves at each level is then computed, selecting the value with the lowest impurity for that feature. This procedure is repeated for each feature to form the nodes, iterating through each depth level until all the data is categorized. After the tree is constructed, predictions for new data points are made by following the tree's path and applying the criteria at each node to determine the final classification or value. Alternatively, decision trees can use the sum of squared residuals or variance to assess impurity, but the other steps in the process remain the same.

– **Regression**
Whereas classification algorithms organize data into categories, regression models forecast continuous numbers. Regression tasks are handled by linear regression methods, whereas logistic regression is applied on classification. A linear connection between independent and dependent variables can be found using linear regression. It identifies the line that fits the best using a linear formula, making it easier to forecast and visualize the outcome of the dependent variable.Simple and multiple linear regression are two further subtypes of linear regression in which the outcome is predicted using one or more independent variables, respectfully.A supervised machine learning technique called logistic regression is used for applications involving binary classification. It

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

forecasts the probabilities of a result, an occurrence, or an observation, yielding a binary or dichotomous result limited to the possibilities of either true or false, 0/1, or positive or negative.By examining the correlation between one or more independent variables, logical regression divides data into distinct groups. It is often applied in forecasting, in which the algorithm calculates the mathematical probability of an event falling into a particular category or not. The linear association between variables, specifically predictors and responses, in linear regression models is expressed by the following equation:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_i x_i$$

• y represents the response variable. • xi represents the i-th predictor variable. • ai represents the average effect on y when xi increases by one unit, assuming all other predictors remain constant. Likewise, logistic regression applies the equation as follows to determine the possibility of a particular observation or occurrence:

$$y(x) = \frac{e^{(a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_i x_i)}}{1 + e^{(a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_i x_i)}}$$

One method of statistical regularization is ridge regression. It compensates up for machine learning models' over-fitting on data used for training.
One of the multiple regularization techniques for linear regression models is ridge regression, sometimes referred to as L2 regularization. A statistical technique called regularization is used to lessen mistakes in training data that result from overfitting. Regression analysis with ridge regression explicitly accounts for multicollinearity. This is helpful for creating large-parameter machine learning models, especially when those parameters have significant weights. The algorithm imposes a penalty proportional to the square of the coefficients' magnitude on the minimization objective.The Minimization objective is represented in

$$= \text{LS Obj} + \alpha \times \left(\sum \text{ of square of coefficients} \right)$$

Lasso regression is one regularization technique. Regression procedures are not chosen for a more accurate forecast. This model makes advantage of shrinkage. Shrinkage is the term used to describe when data values move closer to a central point, or the mean. The lasso technique promotes simple, sparse models, or models with fewer parameters. This particular type of regression is useful when you want to expedite certain steps in the model selection process, like variable selection and parameter elimination, or when your model shows notable levels of multicollinearity.Lasso Regression uses the L1 regularization approach. It is used when there are more features since it chooses features automatically.

$$\text{minimize} \left(\sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 \right)$$

$$\text{subject to} \sum_{j=1}^{p} |\beta_j| \leq t$$

• yi represents the observed response for the ith observation.
• xij denotes the value of the jth predictor variable for the ith observation.
• n indicates the total number of observations.
• p stands for the number of predictor variables.
• β_0 signifies the intercept, while β_j are the coefficients linked to the predictor variables.
• The term t serves as a tuning parameter regulating the extent of shrinkage applied to the coefficients.

3.7 ENSEMBLE LEARNING AND DEEP LEARNING MODELS

Ensemble learning and deep learning are two potent models with distinct advantages. By utilizing the diversity of different models, ensemble learning aims to increase prediction accuracy and robustness by integrating

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

numerous models. Common ensemble methods include Bagging, which builds multiple instances of a model on different subsets of data and averages their outputs; Boosting, which incrementally builds models with the goal of correcting previous errors; and Voting, which aggregates the predictions of various models to make a final decision. Deep learning, on the other hand, uses multiple-layered, complicated neural networks to automatically identify complex patterns and representations from unprocessed data. While deep learning is well known for its capacity to handle big, high-dimensional datasets and extract complex features straight from the data, ensemble learning excels in utilizing a variety of models and reducing overfitting.

- **Ensemble Learning Models**
  - **Bagging Regressor** Bootstrap Aggregating, or Bagging, is a well-liked ensemble learning method that tries to increase the precision and stability of machine learning models. Several base learners, such decision trees, are trained separately on arbitrary portions of the training set, sampled with replacement, in the bagging process. Bagging, which averages these base learners' predictions, efficiently minimizes variation and helps avoid overfitting. A particular application of bagging for regression tasks is the Bagging Regressor, which combines the predictions of several regression models trained on various subsets of the training data. The Bagging Regressor lessens the effect of noise and outliers by aggregating the predictions from various models, producing predictions that are more reliable and accurate. When the base learners show low bias but large variation, this strategy works especially well since it reduces prediction fluctuations and produces more consistent outcomes.
  - **AdaBoost Regressor** Adaptive Boosting is a well-liked ensemble learning technique that may be applied to regression and classification problems. When it comes to regression, AdaBoost Regressor fits a sequence of weak learners to the data in a stepwise manner, with each learner after the other emphasizing more on cases that the preceding ones mispredicted. This is how AdaBoost refines the prediction power of the model by giving more weight to data points that are hard to forecast. The final ensemble prediction is formed by summing the weighted predictions of each weak learner, which is usually a straightforward decision tree regressor. AdaBoost attempts to build a robust ensemble model that reliably identifies the underlying patterns in the data by iteratively changing the weights of misclassified examples. This iterative procedure keeps going until the target accuracy level is attained or a certain number of weak learners are reached.
  - **Voting Classifiers** Voting classifiers are a subset of ensemble learning techniques that aggregate predictions from several distinct models to arrive at a conclusion. These individual models, also known as weak learners or basic learners, can be trained on distinct subsets of the training data or on models of different sorts. Every base learner in a voting classifier makes an independent prediction of the result, and the final prediction is decided using a predetermined voting strategy. Hard voting and soft voting are the two primary categories of voting tactics. In soft voting, the final forecast is based on the average probability of each class across all models, but in hard voting, the final prediction is the majority class predicted by each individual model. Three different regression models are used in this case: LinearRegression, KNeighborsRegressor, and DecisionTreeRegressor. Predictions are made on the test dataset once each model has been trained on the training dataset. The voting classifier then uses a voting strategy—which can be ”hard” or ”soft”—to aggregate these individual predictions. The majority class projected by each individual model is the predicted class label in the ”hard” voting strategy, but the class label with the highest average probability is selected in the ”soft” voting approach.

- **Deep Learning Models**
  - **Multi-layer Perceptron (MLP)** A type of feedforward artificial neural network known as a multilayer perceptron (MLP) is made up of several layers of nodes that are feedforward coupled to one another. MLPs are frequently utilized in deep learning for supervised learning tasks like regression and classification. An input layer, one or more hidden layers, and an output layer are the standard components of an MLP's design. The model becomes non-linear as a result of each node in the hidden layers applying an activation function to the weighted sum of its inputs. As a result, MLPs can discover intricate patterns and connections within the data. MLPs modify the weights of the connections between nodes during training by using gradient descent and back-propagation algorithms, which minimize a loss function and increase prediction accuracy. A neural

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

network design with two hidden layers of 64 and 32 units, respectively, is used to instantiate the MLP. Cross-validation is then used to train and assess this classifier. Throughout the training process, the 'adam' optimizer—a well-liked option for building deep neural networks—is implicitly employed. Following training, metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-Squared score are used to evaluate the model's performance. In addition, a unique deep learning model with the MLP's architecture is built utilizing TensorFlow and Keras. The 'adam' optimizer, which is explicitly stated in the model, is used to train it.the compile() method. The same set of measures are used to assess this deep learning model's performance through K-fold cross-validation. To make comparing results across folds easier, the results are combined into a DataFrame.

3.8 Performance metrics

Several criteria are frequently used to evaluate the performance and dependability of trained models for weather prediction. When performing regression tasks, two often used metrics are Mean Squared Error (MSE) and Mean Absolute Error (MAE). Large errors are given more weight in the MSE calculation, which calculates the average squared difference between the actual and predicted values. A more intuitive view of the prediction errors is offered by MAE, which computes the average absolute difference between the actual and anticipated values. The R-squared (R2) score, which measures the percentage of the target variable's variance that the model can account for, is another crucial parameter. Higher values of the R2 score indicate better model fit. The score goes from 0 to 1. Accuracy is a basic indicator for classification jobs; it is the percentage of correctly categorized examples out of all instances.

- **Mean Squared Error (MSE)** Mean Squared Error (MSE) is a commonly used statistic in regression tasks that calculates the average squared difference between real and predicted values. It is calculated by averaging the squared discrepancies between each expected and actual value. The formula for the MSE is:
$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1}$$
where n is the number of predictions, y_i represents the actual rating, and $\hat{y}_i$ is the predicted rating. A lower MSE indicates better accuracy, with zero representing a perfect prediction.

- **Mean Absolute Error (MAE)** Mean Absolute Error (MAE) is another regression statistic that computes the average absolute difference between actual and expected data. It gives a more intuitive comprehension of the prediction mistakes than MSE. The MAE formula is as follows:
$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$
where n is the number of predictions, y_i represents the actual rating, and $\hat{y}_i$ is the predicted rating. Similar to MSE, a lower MAE indicates better accuracy, with zero representing a perfect prediction.

- **R-squared (Coefficient of Determination)** The R-squared (R2) score is a statistic for assessing the goodness of fit of a regression model. It quantifies the fraction of the dependent variable's variance that can be explained by the independent variables. The R2 value goes from 0 to 1, with a score closer to 1 indicating greater match. The formula for calculating R2 score is:
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$
where $\hat{y}$ is the mean of the actual ratings, and n is the number of predictions. A higher R-squared value signifies a better model fit.

- **Accuracy** Accuracy is a classification metric that counts the percentage of cases that are properly classified out of all the instances. It is computed by dividing the total number of occurrences by the number of correctly classified instances. The performance of a classifier can be easily and intuitively measured by its accuracy, particularly when working with balanced datasets. It might not be appropriate, therefore, for datasets that are unbalanced and have unequal class distributions.

DOI: https://doi.org/10.54216/JCIM.150220
Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

4 Experimental Results and Discussion

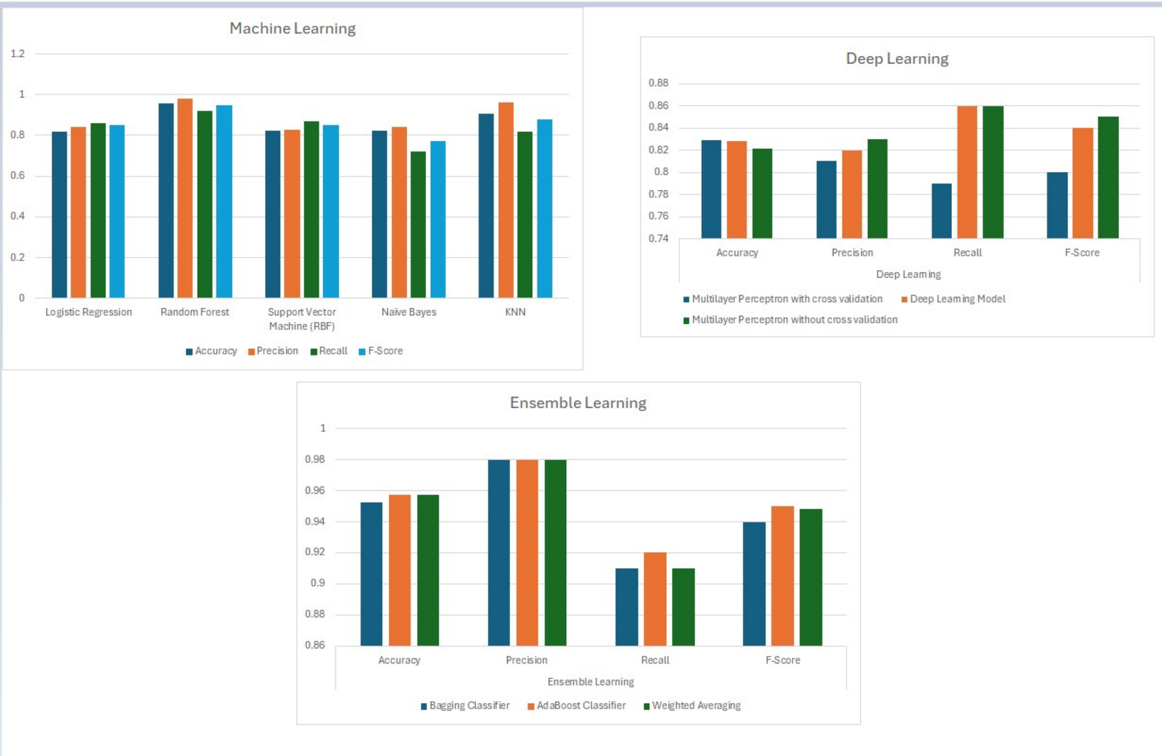

Figure 2: Comparison between Models Results in Case Study I

The conducted experiments on 4 different datasets for weather forecasting. Two of the datasets were classification datasets, referred to as Dataset 1 and Dataset 2, where the task was to classify the weather as either ”True” or ”False”. The other two datasets were regression datasets, called the ”Weather Dataset” and ”ManuaLoaDailyTemps”, where the goal was to predict the average temperature. To tackle these weather forecasting tasks, we applied a wide range of machine learning approaches, including linear regression, decision trees, random forest, ridge regression, support vector machines, k-nearest neighbors, and lasso regression. In addition to applying ensemble techniques, such as AdaBoost Regressor, weighted average, voting classifier, and bagging regressor.

Furthermore, we experimented with deep learning models, building them with k-fold cross-validation and early stopping to prevent overfitting. The use of cross-validation was an important technique to ensure the robustness of the models. In the context of regression tasks, common performance metrics utilized include R-squared, mean squared error (MSE), and mean absolute error (MAE). These metrics help evaluate how well the regression model fits the data and how accurate its predictions are.

On the other hand, for binary classification tasks, the key performance metrics are accuracy, precision, recall, and F1-score. These metrics provide insights into how well the classification model is able to correctly identify classes in the dataset. Accuracy measures the overall correctness of the model’s predictions. Precision indicates the model’s ability to avoid false positives, while recall measures the model’s ability to identify all positive instances. The F1-score is the harmonic mean of precision and recall, offering a balanced evaluation of the model’s performance. The breadth of modeling approaches evaluated demonstrates the value in pursuing further explorations within this field to uncover new advancements and innovations for improving weather prediction capabilities.

4.1 Case Study I (Classification Datasets)

To analyze the performance of different models on these two classification datasets, we are using key evaluation metrics such as recall, precision, accuracy, and F-score. These metrics provide a comprehensive assessment

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

of the models' capabilities in terms of correctly identifying positive and negative instances, as well as their overall predictive power. By comparing the values of these metrics across the different models applied to the two datasets, we can determine the best performing model for each dataset and gain insights into their relative strengths and weaknesses. The in-depth analysis of these evaluation metrics allows us to make an informed decision about which model is most suitable for the given classification tasks represented by these two datasets.

Figure 2 depicts a comparative analysis of various models performance in Case Study I,focusing on classification analysis of our datasets (Dataset "1" and Dataset "2")

In terms of the algorithm performance on the two datasets ,for Dataset 1, the Support Vector Machine (RBF) achieved the highest accuracy at 73.12%, along with the best precision, recall, and F-score. Naïve Bayes had showed outstanding performance as it is the second-highest F-score at 0.76, matching the performance of the Support Vector Machine in this metric. Random Forest came in second for accuracy with a rate of 71.22%, while KNN recorded the lowest accuracy at 67.43%. In contrast, for Dataset 2, Random Forest excelled with the highest accuracy at 95.78%, as well as the best precision and F-score. Logistic Regression, Support Vector Machine (RBF), and Naïve Bayes had similar performance, with accuracy, precision, and F-scores ranging from 81.93% to 82.22%. KNN showed a outstanding performance with the third-highest accuracy at 90.82%. Overall, the performance of the algorithms was generally better on Dataset 2 compared to Dataset 1, with higher accuracy, precision, recall, and F-scores across the board.

For the first dataset, the ensemble learning techniques -Bagging Classifier outperforms the AdaBoost Classifier and Weighted Averaging models. The Bagging Classifier has the highest values for Accuracy, Precision, Recall, and F-Score compared to the other two methods. In contrast, for the second dataset, all three ensemble learning techniques - Bagging Classifier, AdaBoost Classifier, and Weighted Averaging - demonstrate excellent and comparable performance. Among these, the AdaBoost Classifier has a slight edge, with the highest Accuracy, Precision, Recall, and F-Score

For Dataset 1, the Multilayer Perceptron model with cross-validation emerges as the top performer, demonstrating the highest Accuracy, Recall, and F-Score compared to the other models. In contrast, for Dataset 2, the Multilayer Perceptron model without cross-validation shows the best overall performance across the evaluation metrics, including Accuracy, Precision, and F-Score.

Dataset 1 is quite small, with only 1461 data points, which is insufficient to effectively train and model. Additionally, there is a significant class imbalance in the dataset, with some classes like drizzle, snow, and fog having very few samples (53, 26, and 101 occurrences respectively), while classes like rain and sun have much larger samples (640 occurrences). This class imbalance can lead to poor model performance, with the model being biased towards the majority classes and failing to accurately predict the minority classes. This is evidenced by the warnings of the precision, recall, and F1-score being set to 0 for the underrepresented classes. Further analysis of the dataset's specific characteristics and the user's preferences would be beneficial in identifying the most suitable algorithm for weather forecasting.

4.2 Case Study II (Regression DATASET)

Figure 3 depicts a comparative analysis of various models performance in Case Study II,focusing on regression analysis of our dataset (ManuaLoaDailyTemps Results) . Evaluation metrics such as mean square error, mean absolute error and R-squared score highlight the effectiveness of each model. To mitigate over-fitting, cross-validation was applied in both machine learning and deep learning segments. The ensemble model yielded an impressive R-squared score ranging from 0.95 to 1.

The performance of the machine learning models is generally consistent across the two datasets, with the Random Forest Regression model exhibiting the highest R-squared value in both cases. The exception is the Support Vector Machine model, which shows some notable differences in MSE and MAE between the two datasets. Overall, the results for the other models are quite similar across the two datasets. The consistency

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

of the Random Forest Regression model's performance further reinforces the conclusion that it is the best-performing model among the ones provided.

Regarding the ensemble learning methods, Dataset 2 generally exhibits superior performance in terms of MAE and R-squared compared to Dataset 1. The Bagging Regressor appears to be the best-performing ensemble model, with the lowest MSE and MAE, and the highest R-squared value across the two datasets. The AdaBoost Regressor and Weighted Average also deliver commendable performances, with the former showcasing the lowest MSE among the ensemble models. The deep learning models show more variation in performance between the two datasets. For Dataset 1, the MultiLayer Perceptron with Cross-Validation (CV) generally performs better in terms of R-squared, while for Dataset 2, the MultiLayer Perceptron without CV performs better in terms of MSE, MAE, and R-squared. Further analysis, considering dataset specifics and user preferences, would be beneficial in identifying the most suitable algorithm for weather forecasting.

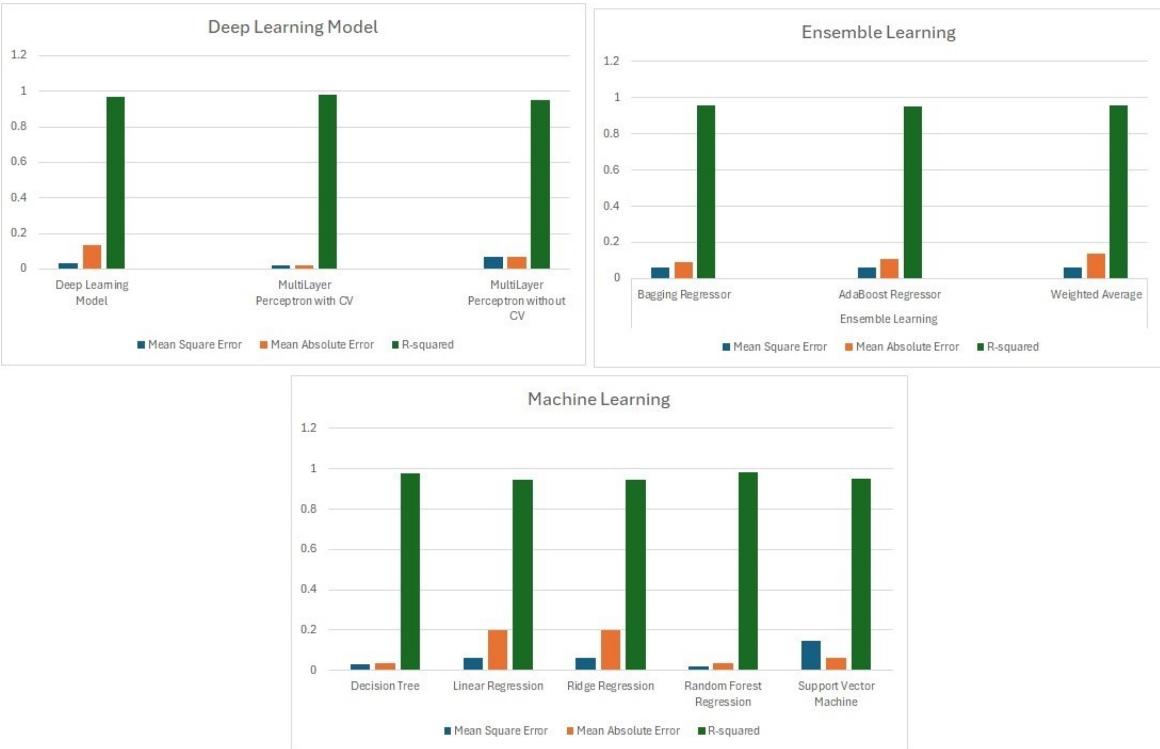

Figure 3: Comparison between Models Results in Case Study II

5 Conclusion and future work

For conclusion, the examination of four weather forecasting datasets—two classification datasets and two regression datasets—showed the usefulness of several machine learning, deep learning, and ensemble learning models. Overall, this thorough analysis using a variety of datasets and modeling approaches emphasizes how crucial it is to use cutting-edge techniques for weather prediction tasks, like ensemble learning and deep learning. These techniques often showed improved prediction ability, which makes them useful for producing weather forecasts that are more precise and trustworthy. A thorough evaluation was ensured by the use of several evaluation criteria, which also helped to identify the advantages and disadvantages of each modeling strategy and direct the choice of the best model for particular prediction tasks.

5.1 Case Study I (Classification Dataset)

The comparison of different models on two classification weather forecasting datasets reveals different patterns in deep learning, ensemble learning, and machine learning techniques. Machine learning methods that forecast

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

weather patterns well include Logistic Regression, Random Forest, SVM, Naive Bayes, and KNN. These models typically show high accuracy and precision, with considerable variation in recall and F1 score. While deep learning models—which include various Multilayer Perceptron (MLP) configurations—achieve balanced performance in precision, recall, and F1 score, indicating their strength in handling complex data structures—they also exhibit slightly lower accuracy when compared to certain machine learning models. When it comes to precision, ensemble learning techniques such as Bagging Classifier, AdaBoost Classifier, and Weighted Averaging perform better than individual models. Notably, Bagging and AdaBoost exhibit exceptionally high accuracy and F1 scores.

There are various intriguing paths that future work on weather forecasting using classification datasets can pursue. Firstly, by better capturing the spatial and temporal relationships in meteorological data, the use of more sophisticated deep learning architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), could enhance model performance. Prediction timeliness and accuracy can also be improved by utilizing big data technology and combining real-time data. By utilizing the advantages of each strategy, hybrid models that include machine learning, deep learning, and ensemble learning techniques may produce even greater advances. Moreover, model performance may be enhanced by fine-tuning hyperparameters using methods like evolutionary algorithms or Bayesian optimization. Finally, increasing the datasets to cover more diverse weather conditions and geographic locations helps improve the model's robustness and generalization, providing accurate forecasts across areas and climate scenarios.

5.2 Case Study II (Regression Dataset)

On the other hand, the comparison of regression models for weather forecasting reveals distinct performance characteristics among different techniques. Machine learning models such as Decision Tree, Linear Regression, Ridge Regression, Random Forest, and Support Vector Machine (SVM) produce consistent Mean Squared Error (MSE) and Mean Absolute Error (MAE) metrics, with R-squared values indicating varying degrees of explanatory power. In deep learning, Multilayer Perceptron (MLP) models, both with and without cross-validation (CV), produce promising results, albeit their performance varies, particularly in terms of R-squared, which measures the models' ability to match the data. The deep learning model with CV performs better, indicating that cross-validation improves generalization. The value of mixing many models to increase prediction dependability is highlighted by the strong performance of ensemble learning techniques, such as Weighted Average, AdaBoost Regressor, and Bagging Regressor, which show appreciable gains in accuracy and error metrics.

To increase the accuracy of weather forecasting, future research could concentrate on enhancing ensemble learning strategies and looking at unique model combinations. Investigating cutting-edge ensemble techniques like Gradient Boosting and Stacked Generalization, which have demonstrated promise in other fields and might provide extra improvements in forecasting performance, could be one line of inquiry. Furthermore, investigating hybrid approaches that combine the advantages of ensemble methods, deep learning, and machine learning may provide synergistic effects and result in more reliable forecasting systems. Meteorologists, data scientists, and machine learning specialists working together could make it easier to create customized models that accurately represent the intricate dynamics of weather systems and offer useful information to the public and decision-makers.

Author contributions

All authors contributed equally to this paper, where Elissa Ayman , Enjy Yasser , Mohamed Osama and Adham Omar participated in sorting the experiments, discussed and analyzed the results, wrote the paper and revised the paper. All authors have read and approved the content of this paper.

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

References

[1] M. Bhawsar, V. Tewari, P. Khare, "A survey of weather forecasting based on machine learning and deep learning techniques", International Journal of Emerging Trends in Engineering Research, vol. 9, no. 7, 2021.

[2] S. Singh, M. Kaushik, A. Gupta, A. K. Malviya, "Weather forecasting using machine learning techniques", Proc. of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE), 2019.

[3] B. Bochenek, Z. Ustrnul, "Machine learning in weather prediction and climate analyses—applications and perspectives", Atmosphere, vol. 13, no. 2, p. 180, 2022.

[4] M. G. Schultz, C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari, S. Stadtler, "Can deep learning beat numerical weather prediction?", Philosophical Transactions of the Royal Society A, vol. 379, no. 2194, p. 20200097, 2021.

[5] I. Maqsood, A. Abraham, "Weather analysis using ensemble of connectionist learning paradigms", Applied Soft Computing, vol. 7, no. 3, pp. 995–1004, 2007.

[6] P. Kankanala, S. Das, A. Pahwa, "AdaBoost^{+}: An ensemble learning approach for estimating weather-related outages in distribution systems", IEEE Transactions on Power Systems, vol. 29, no. 1, pp. 359–367, 2013.

[7] D. Cho, C. Yoo, B. Son, J. Im, D. Yoon, D. H. Cha, "A novel ensemble learning for post-processing of NWP model's next-day maximum air temperature forecast in summer using deep learning and statistical approaches", Weather and Climate Extremes, vol. 35, p. 100410, 2022.

[8] A. H. M. Jakaria, M. M. Hossain, M. A. Rahman, "Smart weather forecasting using machine learning: A case study in Tennessee", arXiv preprint arXiv:2008.10789, 2020.

[9] P. Grönquist, C. Yao, T. Ben-Nun, N. Dryden, P. Dueben, S. Li, T. Hoefler, "Deep learning for post-processing ensemble weather forecasts", Philosophical Transactions of the Royal Society A, vol. 379, no. 2194, p. 20200092, 2021.

[10] S. Murugan Bhagavathi, A. Thavasimuthu, A. Murugesan, C. P. L. George Rajendran, L. Raja, R. Thavasimuthu, "Retracted: Weather forecasting and prediction using hybrid C5.0 machine learning algorithm", International Journal of Communication Systems, vol. 34, no. 10, p. e4805, 2021.

[11] J. A. Weyn, D. R. Durran, R. Caruana, N. Cresswell-Clay, "Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models", Journal of Advances in Modeling Earth Systems, vol. 13, no. 7, p. e2021MS002502, 2021.

[12] J. A. Weyn, D. Kumar, J. Berman, N. Kazmi, S. Klocek, P. Luferenko, K. Thambiratnam, "An ensemble of data-driven weather prediction models for operational sub-seasonal forecasting", arXiv preprint arXiv:2403.15598, 2024.

[13] K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, Q. Tian, "Pangu-weather: A 3D high-resolution model for fast and accurate global weather forecast", arXiv preprint arXiv:2211.02556, 2022.

[14] S. Rasp, P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, N. Thuerey, "WeatherBench: A benchmark data set for data-driven weather forecasting", Journal of Advances in Modeling Earth Systems, vol. 12, no. 11, p. e2020MS002203, 2020.

[15] S. Athiyarath, M. Paul, S. Krishnaswamy, "A comparative study and analysis of time series forecasting techniques", SN Computer Science, vol. 1, no. 3, p. 175, 2020.

[16] A. C. Bart, R. Whitcomb, "CORGIS Datasets Project", available: `https://corgis-edu.github.io/corgis/csv/weather/`, 2017. [Accessed: 22-May-2024].

[17] Machine-Learning-Datasets/MaunaLoaDailyTemps.csv, available: `https://github.com/gupta24789/Machine-Learning-Datasets/blob/master/MaunaLoaDailyTemps.csv`, 2020. [Accessed: 24-May-2024].

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024

[18] ”WEATHER PREDICTION”, available: https://www.kaggle.com/datasets/ananthr1/weather-prediction, 2019. [Accessed: 22-May-2024].

[19] ”Seattle Weather Dataset”, available: https://www.kaggle.com/datasets/malihachaity/heuristic-model, 2021. [Accessed: 24-May-2024].

[20] N. Sani, A. H. Abd Rahman, A. Adam, I. Shlash, M. Aliff, ”Ensemble learning for rainfall prediction”, International Journal of Advanced Computer Science and Applications, vol. 11, no. 11, 2020.

[21] C. Yu, H. C. Li, J. J. Xia, H. Q. Z. Wen, P. W. Zhang, ”A data-driven random subfeature ensemble learning algorithm for weather forecasting”, Communications in Computational Physics, vol. 28, no. 4, pp. 1305–1320, 2020.

[22] S. Scher, G. Messori, ”Ensemble methods for neural network-based weather forecasts”, Journal of Advances in Modeling Earth Systems, vol. 13, no. 2, 2021.

DOI: https://doi.org/10.54216/JCIM.150220

Received: May 22, 2024 Revised: July 19, 2024 Accepted: November 11, 2024