Evaluating the Accuracy of Current Weather Data Models Using Different Machine Learning Models

Jesus Rojas
Computer Science
University of Illinois at Chicago
Chicago, United States of America
jroja9@uic.edu

Zach Etzkorn
Computer Science
University of Illinois at Chicago,
Chicago, United States of America
zetzk2@uic.edu

Bryce Sadelski
Computer Science
University of Illinois at Chicago
Chicago, United States of America
bsade2@uic.edu

Chang Won Choi
Computer Science
University of Illinois at Chicago
Chicago, United States of America
cchoi50@uic.edu

Jose Bolanos
Computer Science
University of Illinois at Chicago
Chicago, United States of America
jbola6@uic.edu

I. Introduction

Weather prediction is a multifaceted and inherently challenging task that involves the integration and modeling of numerous atmospheric variables, each of which evolves dynamically over time. Accurate and reliable forecasts of critical metrics such as average temperature (tavg) and daily precipitation (prcp) are essential for a wide range of applications. These forecasts can profoundly impact sectors such as agriculture, where crop yields depend on temperature and rainfall patterns, infrastructure planning to mitigate weather-related disruptions, water resource management to ensure adequate supply and flood control, public health initiatives to address weather-related illnesses, and disaster preparedness efforts to minimize the impact of extreme weather events.

The advent of machine learning has significantly advanced the field of weather forecasting, offering tools to analyze historical data and uncover patterns that were previously difficult to identify. These advancements have the potential to enhance the accuracy of day-to-day predictions, providing more actionable insights for various stakeholders [1], [2].

In this report, we evaluate the performance of several machine learning regression models in predicting next-day average temperature and precipitation based on historical weather observations. The models explored include Linear Regression (LR), Random Forest (RF), Decision Tree (DT), Ridge Regression (RR), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Each of these models offers unique advantages: linear methods such as LR and RR are valued for their simplicity and interpretability, making them suitable for straightforward relationships. On the other hand, non-linear methods like RF, DT, KNN, and SVM excel at capturing complex and intricate patterns within the data, which are often necessary for modeling weather systems.

However, it is essential to note that no single model provides a universally superior solution, as the effectiveness of each approach depends on the specific characteristics of the data and the nature of the prediction task. The challenge becomes even more pronounced when attempting to predict intermittent and extreme events, such as heavy rainfall or sudden temperature fluctuations, which exhibit high variability and are difficult to model accurately [3], [4]. This analysis aims to shed light on the strengths and limitations of these models in addressing such challenges, providing a comparative perspective on their utility for weather forecasting.

II. Data and Methodology

A. *Data Collection:*

We obtained our dataset from Meteostat, an open-access platform that aggregates historical meteorological and climate data from a wide range of global weather stations. The dataset spans from January 1, 2015, to October 16, 2024, encompassing 3,577 daily observations. Each of these daily entries contains 11 meteorological features: date(N/A), average temperature(°C), minimum temperature(°C), maximum temperature(°C), precipitation(mm), snowfall(mm), wind direction(degrees), wind speed(km/h), wind peak gust(km/h, not used), atmospheric pressure(hPa), and recorded sunshine duration(N/A). Date fields were converted into a proper datetime format and not directly used as features in the models. This extensive coverage and diverse set of features are a great foundation for training a model for analyzing weather patterns and developing forecasting models.

For our testing suite we chose to compare our models on the last month of data (October 17 - November 17), as a small sample that is presentable to a class, but still informative. This data was also from Meteostat and had the same parameters as the training data. It also requires the same testing and cleaning as the training data.

B. *Data Cleansing and Processing:*

Prior to modeling, we undertook a systematic cleaning and preprocessing of the dataset. First, we removed irrelevant columns, specifically wind peak gust and recorded sunshine duration. These columns were left blank from Meteostat and would not be useful for our models. We addressed the missing values in the data set using a few different imputation strategies. First, our decision tree model used forward fill. With forward fill, any missing entries in the data set are replaced with the most recent valid value seen above that row.

To capture temporal dependencies and trends, lagged features (tavg_lag1, prcp_lag1) were generated to include information from the previous day's observations. Additionally, rolling statistics such as a 7-day rolling mean for temperature (tavg_roll_mean) and a 3-day rolling sum for precipitation (prcp_roll_sum) were calculated. These features help incorporate historical patterns into the dataset, enhancing the KNN model's ability to predict next-day weather conditions.

While not a data cleaning method, we chose to incorporate the Random Forest model because it is effective at preventing overfitting. Since Random Forest trains each tree on a different random subset of the data, it reduces the risk of overfitting and helps the model generalize better to unseen data.

C. *Training and Validation Setup*

The dataset was split into training and testing sets, typically using an 80/20 ratio. In some cases, a separate hold-out period (e.g., the last month's data) was used for final evaluation. Training proceeded by fitting each model to historical data, then testing on unseen recent observations, mirroring operational forecasting scenarios [6].This last month's data was chosen because it is an operationally relevant period of time, reflecting recent trends and patterns that are more likely to be applicable to current forecasting needs. Additionally, the choice of the last month makes the results more intuitive and relatable, as recent data is often easier for people to recall compared to data from several years ago.

D. *Metrics and Evaluation*

We evaluated model performance using standard regression metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). These metrics are standard in weather prediction verification [7]. MSE can be decomposed into components that quantify specific aspects of model performance, such as bias and variance [12]. We used MAE in addition because the error for individual data points can be either positive or negative, potentially offsetting one another. To determine the overall forecast accuracy of the model, the absolute values of these errors are calculated to account for their magnitudes, and then averaged to provide a comprehensive measure of performance [13].

Temperature predictions tend to be less variable, enabling models to achieve moderate-to-high R^2. By contrast, precipitation is challenging to predict due to its non-linear and sporadic nature [8].

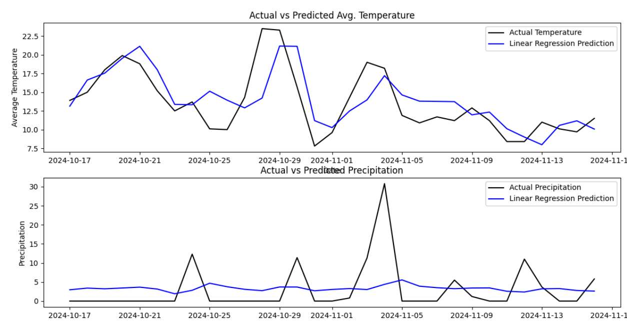
III. Model Descriptions and Their Individual Performance

A. *Linear Regression (LR)*

Linear Regression finds a linear relationship between predictors and the target. While simple and interpretable, LR's assumption of linearity does not always hold in complex atmospheric processes [1]. On a lower level, linear regression is like drawing a line through data points to see how well they relate to each other. It tries to predict an outcome by determining how much each factor contributes.This model is good for predicting data that follows a straight pattern, which will be tested in our analysis to see if weather data can be accurately predicted using a linear relationship between the factors.

Performance:

- Temperature: Linear Regression achieved moderate accuracy for temperature prediction, with an R^2 score of 0.92 indicating that it explained 92% of the variance in the data. While it captured basic trends effectively, the model occasionally struggled with extreme values or rapid fluctuations in temperature, as its linear nature restricts flexibility in modeling non-linear patterns.
- Precipitation: For precipitation prediction, the performance of Linear Regression was suboptimal. With a negative R^2 score (-0.04), the model failed to capture the sporadic and highly variable nature of rainfall. Its predictions were often clustered near-zero values, reflecting a bias toward the mean and an inability to model sudden precipitation spikes



accurately.

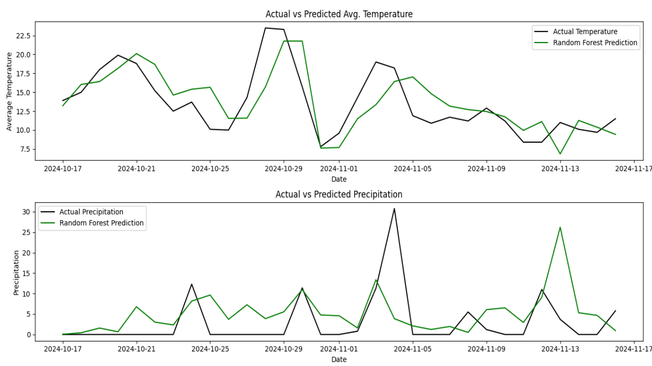
B. *Random Forest (RF)*

Random Forest uses ensemble averaging of multiple decision trees, reducing variance and capturing non-linear interactions [9]. It is well-regarded in climate and weather applications for its robustness. Random Forest is a machine learning model that combines many decision trees to make more accurate predictions. Each tree in the forest makes its own prediction, and then the final result is determined by averaging the predictions. It is called random due to each tree being trained on random subsets of the data, which helps overfitting.

Performance:

- Temperature: Random Forest demonstrated a slight improvement in temperature prediction over Linear Regression. Its R^2 score of 0.92 was comparable to Linear Regression, but its ability to model non-linear relationships provided better handling of temperature fluctuations. By leveraging its ensemble nature, Random Forest captured more complex patterns, particularly during days with significant deviations from average temperatures. However, this improvement came at the cost of higher computational complexity.

- Precipitation: For precipitation, Random Forest exhibited better differentiation between rainy and non-rainy days compared to Linear Regression. While it achieved slightly more accurate predictions for smaller precipitation events, it struggled with larger, less frequent spikes. The negative R^2 score (-0.37) indicates that the model still failed to generalize effectively for extreme rainfall events. This limitation could stem from an imbalance in the dataset or the inherent unpredictability of precipitation, which requires finer-grained features or additional temporal context to predict accurately.

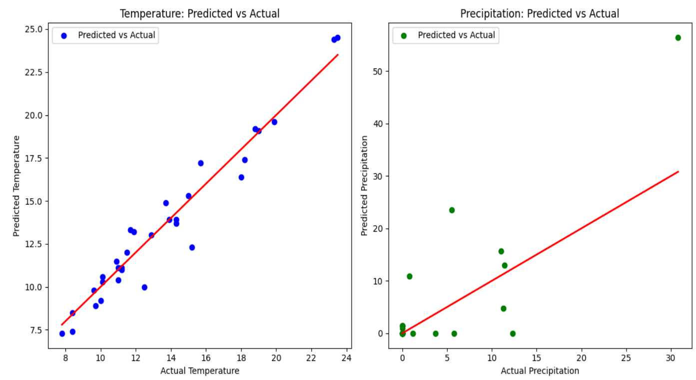


C. *Decision Tree (DT)*

Decision Tree modeling is a widely used modeling technique for creating frameworks driven by many input variables or for constructing prediction models for a given target variable. The decision tree approach segments a population into branch-like structures that form an inverted tree with a single root node, several internal anodes and final leaf nodes. Due to the non-parametric modeling strategy, it can process large and complex datasets efficiently relying on predefined parametric assumptions [10]. On a lower level, decision tree modeling focuses on making predictions by asking a series of questions at each branch. In our case, the questions were based on the features of our data, trained by the model to split the data into meaningful segments. For this model, we utilized forecasting data, which consisted of numerical values, allowing the decision tree to effectively analyze and predict outcomes based on these quantitative inputs. This model was chosen because it is easy to interpret, handles both numerical and categorical data effectively, and can capture complex relationships in weather data without requiring extensive preprocessing. This makes it a good baseline model for evaluating the accuracy of current weather data models.

Performance:

- Temperature: Improved accuracy over linear regression, often achieving a higher R^2. The model has a test MSE of 1.02 and an R^2 of 0.94, with a validation MSE of 1.29 and a validation R^2 of 0.99.
- Precipitation: Showed some capacity to detect rainy vs. non-rainy days, but large precipitation events were not consistently predicted. The model has a test MSE of 42.47 and an R^2 of -0.06, with a validation MSE of 107.99 and a validation R^2 of -1.18. The model performs well on days with no rain.

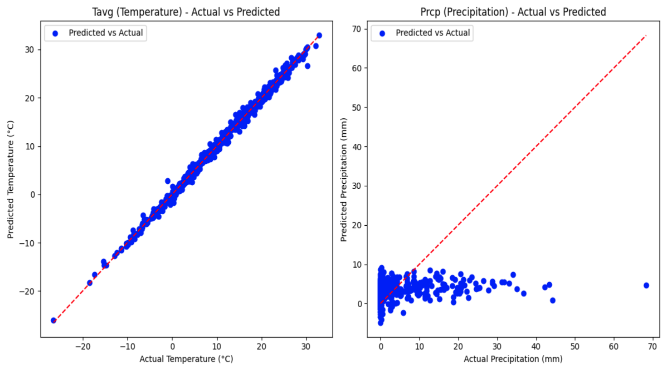


D. *Ridge Regression (RR)*

Ridge Regression, a linear model regularized with an L2 penalty, is well-suited for addressing multicollinearity and mitigating overfitting by constraining the magnitude of model coefficients. This characteristic is particularly beneficial in weather modeling, where predictors like temperature, precipitation, and other atmospheric variables often exhibit high degrees of intercorrelation. Ridge Regression's simplicity and interpretability make it a commonly used method in climate modeling, as it provides insights into the relationships between variables while maintaining robustness in predictions.

Performance:

- Temperature: Ridge Regression excelled in predicting average temperatures, demonstrating remarkable consistency and accuracy. The model effectively captured the linear trends present in temperature data, producing predictions that closely aligned with observed values. For example, when observed average temperatures ranged from 11.0°C to 19.9°C, the predicted temperatures fell within a narrow margin of error, rarely deviating more than 1°C. This strong performance highlights Ridge Regression's ability to generalize well for stable, continuous variables like temperature, where linear relationships dominate.
- Precipitation: In contrast, the model struggled to predict daily precipitation accurately, particularly for medium to high precipitation amounts. For instance, while precipitation values ranged from 0.0 mm to 30.8 mm, Ridge Regression frequently underestimated higher values, predicting significantly lower precipitation amounts even in cases of heavy rainfall. This limitation can be attributed to the inherent variability and non-linearity in precipitation patterns, which Ridge Regression, as a linear model, is less equipped to handle. Predicted precipitation values often clustered near zero or failed to reflect the sporadic spikes in rainfall accurately, resulting in substantial errors.

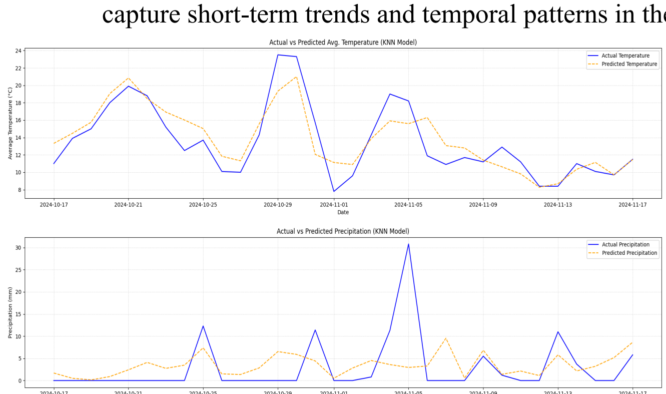


E. K-Nearest Neighbor (KNN)

KNN is a simple, but effective machine learning algorithm that classifies or predicts the value of a data point based on the average of its nearest neighbors. This algorithm is well suited for temperature prediction because of its ability to capture local trends and patterns.

Performance:

- Temperature: The model was able to garner respectable accuracy scores. With an MSE of 10.36 and an MAE of 2.38, the model was able to predict temperature to a good extent and was able to follow the pattern most of the time. The addition of rolling and lag features were able to further enhance the accuracy and R^2 score of the model. The model also used distance weights and Manhattan metrics. By no means was this model the best, but it performed as it should considering the limitations of KNN.
- Precipitation: The model was unable to accurately predict precipitation consistently and struggled with capturing the patterns of precipitation with an MSE of 51.1 and an MAE of 4.2. The model used uniform weights and Chebyshev metrics, and also utilized rolling and lag features. Despite these additions, the model performed far worse in predicting precipitation compared to temperature prediction which is likely the result of the inherent issues with KNN that is KNN tends to favor majority classes which leads to it performing suboptimally on minority classes (days where precipitation does not occur).
- The integration of lagged features and rolling statistics contributed to improved model performance, particularly for temperature prediction. These features enabled the KNN model to better capture short-term trends and temporal patterns in the data, as reflected in their lower MSE and higher R^2 scores

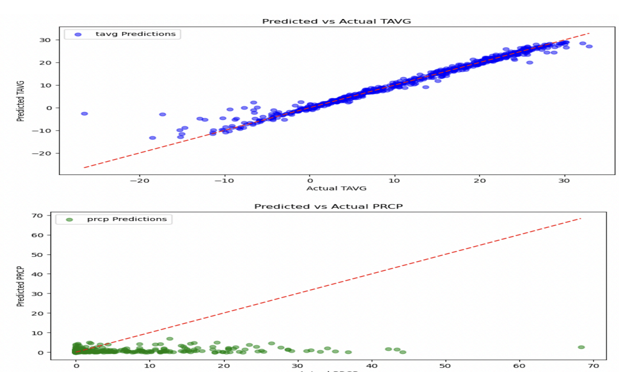


F. Support Vector Machine (SVM)

A Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. It works by finding the optimal hyperplane, a boundary that separates data points from different classes in a multidimensional feature space. The key idea behind SVM is to maximize the margin, which is the distance between the hyperplane and the nearest data points from each class, known as support vectors. This maximized margin helps the model generalize better to new, unseen data. If the data is not linearly separable, SVM uses a technique called the kernel trick to transform the data into a higher-dimensional space where a linear separation becomes possible. Common kernel functions include linear, polynomial, and radial basis function (RBF) kernels. SVMs are particularly effective in high-dimensional spaces and work well in cases where there is a clear margin of separation between classes, making them a robust choice for a weather predicting machine learning problem.

Performance:

- Temperature: This model achieved accurate scoring prediction for temperature based on the data we used. The Mean Average Error (MAE) was 0.79 meaning the average error for a predicted point was approximately 0.79. The model also achieved an impressive R^2 of about 0.976 which signifies that the model fits very well on the new data.
- Precipitation: Similarly to other models, the SVM performed poorly when predicting precipitation. The MAE was higher than temperatures, with a metric of 2.83. This means that our model was on average more inaccurate when predicting precipitation. Additionally, the R^2 was roughly -0.05 which indicates that the model did not fit well on the new data.

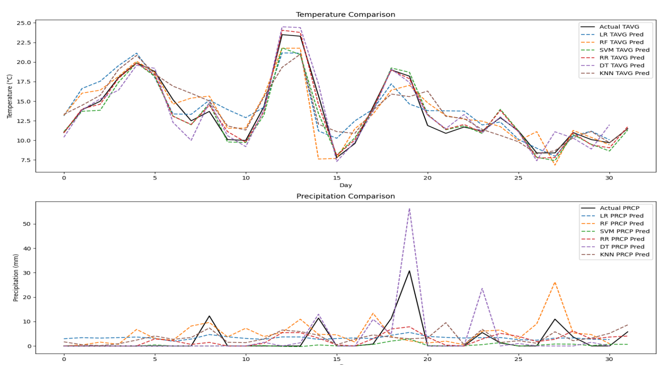


IV. Novelty Statement

This study presents a comparative analysis of multiple machine-learning regression models for next-day weather prediction, focusing on average temperature and daily precipitation. While prior research has explored individual model applications, this work systematically evaluates both linear models (Linear Regression, Ridge Regression) and non-linear approaches (Random Forest, Decision Tree, K-Nearest Neighbors, Support Vector Machine) under identical conditions using historical weather data. Notably, the integration of lagged features and rolling statistics enhances temporal dependency capture, particularly for challenging precipitation forecasts. The study identifies Ridge Regression's superior performance in temperature prediction due to its regularization, and it highlights the limitations of existing methods in handling precipitation's high variability, offering valuable insights into model selection and performance trade-offs for operational forecasting scenarios.

V. Discussion

The results of this study highlight the inherent complexities of weather prediction, particularly the challenges of modeling precipitation compared to average temperature. This section explores the performance of the machine learning models tested, emphasizing their strengths and limitations in the context of the study's objectives. We were limited with the data we could have used. Meteostat provided a very large and useful dataset, being the last nine years of very accurate data from our area. On the other hand, having only 9 used features proved to make our models not very accurate in predicting precipitation.



A. Predicting Average Temperature

The models generally performed well in predicting average temperature, with Ridge Regression (RR) emerging as the most accurate. Its regularization technique (L2 penalty) effectively reduced the impact of large coefficients, allowing the model to capture linear trends in temperature data accurately. This aligns with the model's reputation for mitigating intercorrelation and producing robust predictions in climate modeling applications. Support Vector Machine (SVM) also excelled in temperature prediction, achieving a high R^2 of approximately 0.976 and a low Mean Absolute Error (MAE) of 0.79. This performance demonstrates SVM's ability to handle non-linear relationships and generalize well to unseen data.

Linear Regression (LR) and Random Forest (RF) also exhibited strong predictive capabilities for temperature, with both achieving R^2 scores of approximately 0.92. LR, while simple and interpretable, effectively captured basic linear trends in the data, making it suitable for scenarios with limited complexity. However, RF provided additional accuracy and flexibility, leveraging its ability to model non-linear interactions and variability in the dataset. RF performed particularly well in capturing fluctuations during days with significant deviations, thanks to its ensemble averaging. Despite these strengths, both models fell short of the precision achieved by RR and SVM, highlighting the advantages of regularization and hyperplane optimization in this context.

Other models, including Decision Tree (DT) and K-Nearest Neighbors (KNN), demonstrated moderate success in temperature prediction. DT captured nonlinear interactions well but was more prone to overfitting, particularly with smaller datasets. KNN effectively leveraged local trends in the data but struggled with global patterns, leading to slightly lower accuracy compared to RF and RR.

B. Predicting Precipitation

Unlike temperature, precipitation proved significantly more challenging to predict, primarily due to its high variability and sporadic nature. All models struggled to achieve acceptable accuracy, highlighting the limitations of traditional regression approaches in this domain. Linear Regression (LR) had the lowest Mean Squared Error (MSE) but failed to capture medium to large precipitation events, often defaulting to near-zero predictions on rainy days. Decision Tree (DT) was the most effective at identifying general trends in precipitation but was sensitive to outliers, leading to high MSE(42.48).

Ridge Regression (RR), despite its success with temperature, underperformed in precipitation prediction, likely because its regularization method is less suited for modeling sporadic and extreme events. Similarly, SVM's high MAE of 2.83 and negative R^2 of -0.05 reflect its inability to fit the precipitation data effectively.

Other models including Random Forest (RF) and KNN also struggled, with no significant improvements over other models in terms of capturing the complexity of precipitation patterns.

C. Insights and Implications

The results of our comparative analysis reveal several critical insights into the modeling of weather variables, particularly average temperature and precipitation. While machine learning models demonstrated moderate to excellent performance for temperature prediction, they struggled to accurately forecast precipitation. This discrepancy underscores the fundamentally different statistical properties and climatological characteristics of these two targets.

The poor precipitation forecasting performance suggests that future efforts should focus on specialized modeling strategies better suited to handle rare and extreme events. For instance, probabilistic models or quantile regression techniques could provide not just point estimates but also uncertainty bounds, which are essential when predicting events like sudden downpours. Likewise, deep learning architectures (e.g., Long Short-Term Memory networks, Convolutional Neural Networks), which have shown promise in capturing temporal and spatial patterns in meteorological data, may be beneficial.

Our findings also highlight the pivotal role of data quality, diversity, and preprocessing. While adding lagged features and rolling statistics introduced some temporal context, these engineering steps did not fully capture the drivers of precipitation variability. Precipitation is not only a function of recent weather but also influenced by regional moisture availability, topography, and large-scale weather patterns, which are factors not adequately represented by the basic meteorological variables used in this analysis.

Ultimately, the disparity in performance between temperature and precipitation predictions highlights the need to tailor machine learning strategies to the unique characteristics of the target variable. Whereas stable, continuous variables like temperature respond well to regularization and kernel methods, sporadic phenomena like rainfall may require entirely different modeling approaches such as focusing on the probability of occurrence rather than exact amounts, or leveraging specialized loss functions that emphasize correct detection of rare but impactful events.

IV. Conclusion

The comparative performance of multiple machine learning models offered valuable insights into the challenges and best practices of weather prediction.

For average temperature, the analysis found that Ridge Regression (Model 5) achieved the lowest Mean Squared Error (MSE). Its success likely stems from its regularization technique, which reduced the impact of large, influential coefficients. Such coefficients are common in weather datasets, where certain predictors may have disproportionate effects on the outcome. By preventing these coefficients from growing excessively large, Ridge Regression maintained more stable and accurate predictions. Overall, all models performed reasonably well for temperature, indicating that this variable's smoother temporal dynamics are relatively straightforward to learn.

In contrast, precipitation forecasting remained problematic. While Linear Regression (Model 1) ended up with the lowest MSE in absolute terms, this does not necessarily mean it captured the reality of precipitation's erratic nature. Decision Tree Regression (Model 4), for example, sometimes aligned better with observed trends and patterns, but its predictions were strongly affected by outliers and extremes, thus inflating its MSE. Across the board, none of the tested models excelled in predicting precipitation amounts. Including additional columns such as humidity and sunshine duration may have improved the predictions, as these features are more strongly correlated with precipitation [14]. Regardless of the possible improvements, higher error rate reflects an inherent limitation that daily rainfall is difficult to forecast accurately without richer input features and more complex modeling frameworks.

These findings support a key takeaway: machine learning models must be carefully chosen and tuned based on the target variable's characteristics. While Ridge Regression and SVM are advantageous for temperature forecasts, precipitation might require more specialized approaches. Advanced ensemble methods, integration with physical-based models, or deep learning architectures that capture non-linear and non-stationary patterns of rainfall may be necessary to achieve meaningful improvements.

In summary, this study highlights both the promise and the current limitations of standard machine learning approaches in weather prediction. As the field advances and more nuanced strategies are employed—combining richer datasets, novel architectures, and domain-specific insights—forecasters may move closer to accurate and reliable predictions for challenging phenomena such as precipitation.

Group member contributions

- Jose Bolanos: Assisted in sourcing and preparing the data used to train the models for our research. Additionally, developed the Ridge Regression model and trained it with the provided data to predict both precipitation and temperature.
- Bryce Sadelski: Helped clean the data, found the last month testing suite, designed the decision tree model and trained it with the provided data to predict both precipitation and temperature.
- Jesus Rojas: Developed the K-Nearest Neighbor model and trained the model with the provided data to predict precipitation and temperature. Additionally, used log and rolling features to provide insights to possible enhancements to our models.
- Zach Etzkorn: Designed and implemented the support of the vector machine model on the data provided.
- Chang Won Choi: Assisted in sourcing the data used to train the models in our research. Additionally, developed the linear regression and random forests models. Lastly, combined the models to create an application and graphs to compare our models.

References

1. R. Wilks, *Statistical Methods in the Atmospheric Sciences*, 3rd ed. Academic Press, 2011.
2. G. Mariani et al., “Short-term forecasting of weather conditions using machine learning techniques,” *J. of Appl. Meteorol. Clim.*, vol. 57, no. 8, pp. 1793-1805, 2018.
3. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794.

4. M. Grover and S. Chawla, “Precipitation forecasting using machine learning techniques: A review,” *Int. J. of Eng. Tech. Sci. & Res.*, vol. 5, no. 3, pp. 212-219, 2018.
5. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, 2009.
6. E. Wilcox and J. Ding, “Seasonal forecasting of local surface temperatures using regression-based machine learning,” *Atmos. Sci. Lett.*, vol. 20, no. 10, pp. 1-10, 2019.
7. J. Murphy, “What is a good forecast? An essay on the nature of goodness in weather forecasting,” *Weather and Forecasting*, vol. 8, pp. 281–293, 1993.
8. R. V. Rohli and A. J. Vega, *Climatology*, Jones & Bartlett Learning, 2017.
9. L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
10. Song, Yan-Yan, and Ying Lu, “Decision tree methods: applications for classification and prediction.” Shanghai archives of psychiatry vol. 27,2 (2015): 130-5.
11. Mohita Narang, “Top 4 techniques for handling missing values in machine learning,” Paperspace Blog, 2
12. Hodson, T. O., Over, T. M., & Foks, S. S. (2021). Mean squared error, deconstructed. Journal of Advances in Modeling Earth Systems, 13, e2021MS002681.
13. Vijay Kotu, Bala Deshpande, Chapter 12 - Time Series Forecasting, Editor(s): Vijay Kotu, Bala Deshpande, Data Science (Second Edition), Morgan Kaufmann, 2019, Pages 395-445, ISBN 9780128147610,
14. Sarmad Dashti Latif, Nur Alyaa Binti Hazrin, Chai Hoon Koo, Jing Lin Ng, Barkha Chaplot, Yuk Feng Huang, Ahmed El-Shafie, Ali Najah Ahmed, Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches, Alexandria Engineering Journal,Volume 82, 2023, Pages 16-25