JOURNAL OF
EMERGING INVESTIGATORS

Article

Monitoring drought using explainable statistical machine learning models

Clarisse Cheung[1], Sindhu Ghanta[2]
[1]Notre Dame San Jose High School, San Jose, California
[2]AIClub, Santa Clara, California

SUMMARY
Droughts have a wide range of effects, from ecosystems failing and crops dying, to increased illness and decreased water quality. Drought prediction is important because it can help communities, businesses, and governments plan and prepare for these detrimental effects. This study predicts drought conditions by using predictable weather patterns in machine learning models. The model utilized factors such as wind speed, temperature, and humidity to better understand the importance of certain features in drought prediction. We hypothesized that explainable machine learning models can effectively forecast future drought severity. Results of experiments with a public domain data set showed that explainable statistical machine learning models predict drought severity with average precision and recall values of 0.55 and 0.56, respectively. We identified a subset of meteorological indicators as major contributors to predicting the occurrence of drought. The experiments used K- nearest neighbors (KNN) and random forest algorithms to predict drought severity. The random forest algorithm achieved the highest accuracy at 200 trees and 30 depth, while KNN achieved the highest accuracy when K was set to one. Feature importance analysis showed that surface pressure, temperature range at two meters, and maximum temperature at two meters were the most important features. In conclusion, statistical machine learning models provide insights into factors that impact drought and can enable researchers to identify steps to prevent natural disasters.

INTRODUCTION
Drought is the lack of precipitation over a period of months causing a water shortage. The deficiency of water can cause ecosystems to fail, as critical functions dependent on precipitation are heavily impacted. Farms and ranches rely on precipitation for a large amount of water to grow crops, thus droughts lead to a decline in productivity as crops die (1). Regional effects can spread to entire countries. For example, California has experienced six droughts within the last century and currently is in drought (2). A majority of fruits and nuts that are supplied to the U.S. are grown in California, meaning drought in California severely affects the agricultural supply to the rest of the nation (3). Precipitation is also important to societal health, as a lack of precipitation leads to a decline in water amount and quality, which is associated with an increase in overall illness. Shipping transportation costs also increase, since droughts affect waterways and ports. The dry climate additionally leads to an increase in wildfires that can grow to large scales (1).

As global temperatures continue to rise, droughts continue to plague the U.S., they not only become more frequent, but last longer (4). At the beginning of the 21st century, several droughts were spread throughout the U.S., but by 2012, the geographic areas were combined, converting it into a nationwide event that had not been seen in decades (5). In 2022, 41% of the U.S. was in a drought, affecting 130 million people (6).

Drought prediction is important to help communities, businesses, and governments to plan and prepare for dry conditions that can have significant impacts on agriculture, water resources, and the economy (7, 8). By identifying factors that lead to drought, prediction not only facilitates loss prevention, but also drought prevention: predicting droughts in a timely manner prevents the intensification of droughts too, as it allows for preventative measures to be implemented. This leads to decreased impacts of water shortages in agriculture, ecosystems, and society (9).

While drought prediction is of utmost importance, it is an increasingly difficult task. Drought prediction is a complex and active area of scientific research, as it involves understanding and forecasting the complex interactions between atmospheric, hydrological, and land surface processes that influence the availability of water (10). Understanding and forecasting these complex interactions requires the use of a range of tools, data sources, and computer models. From monitoring these interactions, weather patterns can be predicted weeks to months in advance, and this information can be leveraged to predict drought (11).

In this study, we hypothesized that machine learning models trained on historical meteorological indicators such as wind speed, temperature, and humidity can classify weather conditions into the five categories of drought severity. We further hypothesized that explainable machine learning models can provide insights into the contributing factors behind drought occurrence. We evaluated these hypotheses using a publicly available dataset collected by NASA Power project and the authors of U.S. Drought monitor as well as machine learning algorithms, including K-Nearest Neighbors and random forest (12-15). Results showed that statistical machine learning models predicted drought severity across the five classes with an accuracy of 56.09%. In this study, we relied on the fact that weather patterns or meteorological indicators can be predicted weeks or months in advance, and these values can be used in turn to predict drought conditions at a future time. Analyzing feature importance, we identified that the most impactful indicators were surface pressure, temperature range at two meters, and maximum temperature

https://doi.org/10.59720/ 23-026

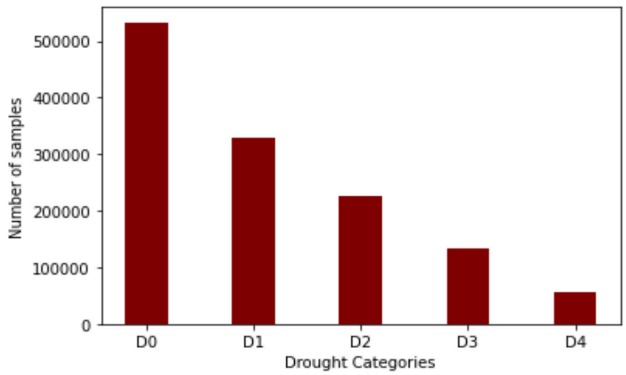

Figure 1: Frequency of the different categories prior to dataset cleaning and balancing. D0 contained 532,931 samples. D1 contained 329,007 samples. D2 contained 225,007 samples. D3 contained 132,089 samples. D4 contained 56935 samples. The dataset contained 2,756,796 inputs with categories.

at two meters.

RESULTS

Prior to cleaning and balancing, we graphed the distribution of each categorization in the dataset (**Figure 1**). The dataset includes 18 meteorological indicators with varying distributions of values in each indicator (**Figure 2**). To further understand these distributions, we projected each feature on a two-dimensional graph to identify the commonalities between indicators, resulting in large overlap and supporting the complexity of drought prediction (**Figure 3**).

We ran each combination of number of trees ranging from 10 to 300 and depth ranging from 1 to 40 through the random forest model (**Figure 4**). The highest validation accuracy from the validation subset of the data of 55.64% was achieved by the random forest model with 200 estimators/trees and a depth of 30 (**Figure 4F**). The test accuracy of this model was 56.09%. The highest validation accuracy achieved by the KNN algorithm was 52.21% for a K value of one (**Figure 5A**). Therefore, the random forest algorithm performed better than KNN in the validation accuracy metric. We determined PS (surface pressure measured in kilopascal), T2M_RANGE (temperature range at two meters, measured in Celsius) and T2M_MAX (maximum temperature at two meters, measured in Celsius) to be the most important features in predicting the drought condition by the random forest algorithm. To determine if the difference in ranges between the various features lowered the initial accuracy, we normalized the

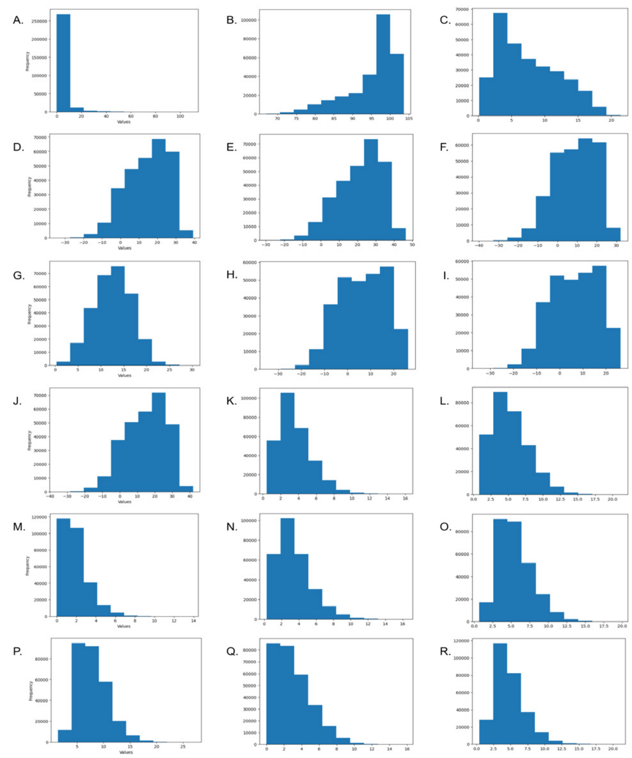

Figure 2: Histograms of the data distribution for each meteorological indicator in the dataset. Each graph has 10 bins. A) Precipitation of the previous day value distribution measured by millimeters. B) Surface pressure value distribution measured in kPa. C) Humidity at 2 meters value distribution measured in grams per kilogram. D-I) Temperature, maximum and minimum temperature, temperature range, dew point, and wet bulb temperature value distribution measured at 2 meters in Celsius. J) Earth skin temperature value distribution measured in Celsius. K-N) Wind speed, maximum and minimum wind speed, and wind speed range value distribution measured at 10 meters in meters per second. O-R) Wind speed, maximum and minimum wind speed, and wind speed range value distribution measured at 50 meters in meters per second.

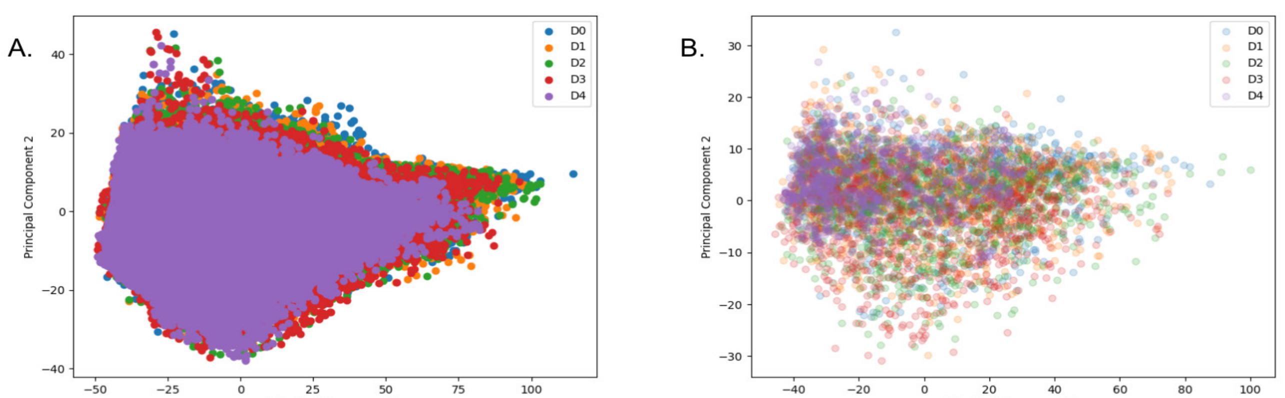

Figure 3: Principal component analysis visualization of the data. A) Graphed on the whole balanced and cleaned dataset. B) Graphed on the first 1000 samples to provide transparency.

JOURNAL OF
EMERGING INVESTIGATORS

https://doi.org/10.59720/ 23-026

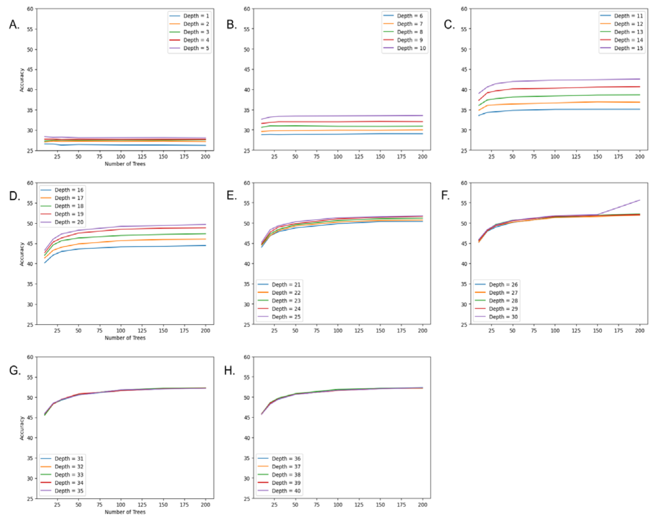

Figure 4: Random forest experiments with various trees and depths to determine the highest accuracy. Experiments run with number of trees ranging from 10 to 30 and depth ranging from 1 to 40. A) Depth 1 to 5 with accuracy under 30%. B) Depth 6 to 10 with accuracy under 35%. C) Depth 11 to 15 with accuracy under 45%. D) Depth 16 to 20 with accuracy under 50%. E) Depth 21 to 25 with accuracy under 52%. F) Depth 26 to 30 with the highest accuracy of 56.64%. F-H) Depth 31 to 40 with accuracy under 53%.

dataset to scale all the features to the same range. Then, we ran KNN on the test subset of the data, and the highest accuracy was 50.19% for a K value of one (**Figure 5B**).

In machine learning and data analysis, feature importance refers to the relative importance of each feature (also known as a predictor or input variable) in explaining the target variable. In other words, it is a measure of how much each feature contributes to predicting the target. Random forest, which achieved the highest accuracy, calculates feature importance across all features in the dataset. Using this aspect of random forest, we determined PS, T2M_RANGE and T2M_MAX to be the most important features (**Table 1**). To validate this conclusion, we ran a sensitivity analysis that resulted in different important features. The sensitivity analysis concluded that precipitation (PS), precipitation (PRECTOT), and temperature range in celsius at two meters (T2M_RANGE) were the most important features.

Random forest with 200 trees and depth 30 achieved the best validation performance. This model was used to make predictions on the test dataset. We obtained an accuracy of 56.09% on the test dataset. The confusion matrix of the model on the test dataset showed that the model has a good generalized performance meaning the model correctly predicted the majority of the classifications (**Figure 6C**). A confusion matrix is a powerful tool used to evaluate the performance of a classification model by organizing the number of instances in every combination of true or false categorizations.

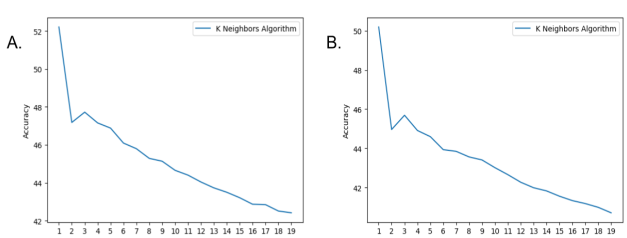

Figure 5: KNN accuracy graphs. Each experiment run with K values 1 to 19. A) Experiment run on the cleaned and balanced dataset achieved the highest accuracy of 52.21%. B) Experiment run on a cleaned, balanced, and normalized dataset achieved the highest accuracy of 50.19%.

Features	Description	Feature Importance
PS	Surface Pressure (kPa)	0.07450512
T2M_RANGE	Temperature Range at 2 Meters (C)	0.06140603
T2M_MAX	Maximum Temperature at 2 Meters (C)	0.06112899
TS	Earth Skin Temperature (C)	0.06087515
T2M_MIN	Minimum Temperature at 2 Meters (C)	0.06014872
T2M	Temperature at 2 Meters (C)	0.05725238
WS50M_RANGE	Wind Speed Range at 50 Meters (m/s)	0.05516957
WS50M_MAX	Maximum Wind Speed at 50 Meters (m/s)	0.0546286
WS10M_RANGE	Wind Speed Range at 10 Meters (m/s)	0.05393973
QV2M	Specific Humidity at 2 Meters (g/kg)	0.05389419
WS50M	Wind Speed at 50 Meters (m/s)	0.05376184
WS10M_MAX	Maximum Wind Speed at 10 Meters (m/s)	0.05167943
WS10M	Wind Speed at 10 Meters (m/s)	0.05155912
WS50M_MIN	Minimum Wind Speed at 50 Meters (m/s)	0.05134174
T2MDEW	Dew/Frost Point at 2 Meters (C)	0.0512547
T2MWET	Wet Bulb Temperature at 2 Meters (C)	0.05123049
WS10M_MIN	Minimum Wind Speed at 10 Meters (m/s)	0.05041334
PRECTOT	Precipitation (mm day-1)	0.04581085

Table 1. Meteorological indicators as represented in the dataset and descriptions sorted according to feature importance in the random forest algorithm with 200 trees and 30 depth.

DISCUSSION

In our study, we hypothesized that drought severity can be effectively forecasted through the use of explainable machine learning models. Using various meteorological features, we achieved a highest validation accuracy of 55.64% by the random forest model while the highest validation accuracy for KNN was 52.21%. The test accuracy of random forest was 56.09% and KNN resulted in 50.19%. With random forest, we determined the most important features to be PS, T2M_RANGE, and T2M_MAX.

We under-sampled the dataset for the majority class in order to overcome the issue of label skew. This may have resulted in loss of information about the majority class, which in turn impacts the performance of the machine learning models. More sophisticated machine learning algorithms based on neural networks could potentially improve performance. However, we did not explore these algorithms

https://doi.org/10.59720/ 23-026

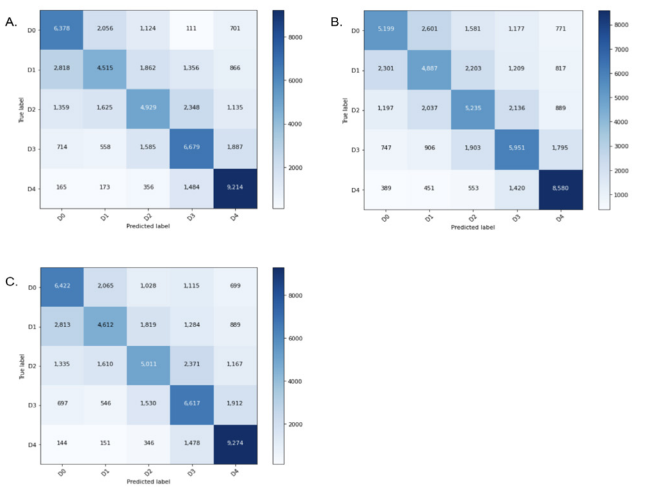

Figure 6. Confusion matrices for each algorithm and random forest on test data. A) Random forest with 200 trees and 30 depth run on validation data. Predicted labels match true labels for majority of instances. Most correctly predicted category was D4 with 9,214 instances. B) KNN with K=1 run on validation data. Predicted labels match true labels generally. D4 was the most correctly predicted category with 8,580 instances. C) Random forest with 200 trees and 30 depth run on test data. Predicted and true labels match for majority of instances. Category with most correct predictions was D4 with 9,274 instances.

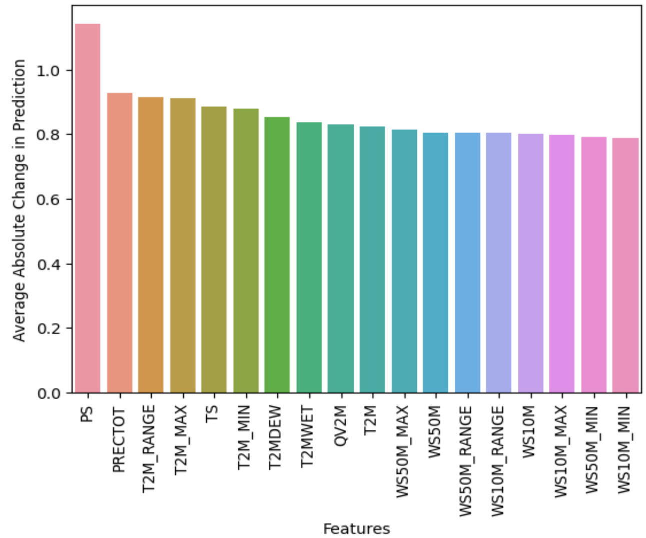

Figure 7. Feature importance according to sensitivity analysis. Run by removing one feature at a time and measuring the decrease in accuracy. PS was the most important feature and affected predictions by an average of 1.14% while the lowest achieved impacted predictions by an average of 0.79%.

as their predictions are not explainable and they did not easily provide any information on feature importance without performing additional analyses.

In classification tasks, ordinal data can be used as the target variable, which is what the model is predicting. Ordinal data is a type of categorical data that has a meaningful order or ranking. For example, if the true value of drought is D0 and the model predicts D1, it is much more acceptable than a prediction of D4. Enacted policies are similar to policies of another category, so it is acceptable to predict similar categories. On the other hand, opposing categories would call for dissimilar policies so distant predictions are not preferable to adjacent predictions. Given that the label categories are ordinal and there are five levels of drought values that the models are predicting, a validation accuracy of 55.64% can be acceptable if the mis-predictions are in the neighborhood of the true drought value. With this interpretation, the results from the machine learning algorithms showed confusion mostly with neighboring drought levels (**Figure 6**). It is important to note that the confusion matrix of the algorithm on test data had an accuracy of 56.09% and showed a similar trend of confusion shown by the validation data (**Figure 6C**), where the mis-predictions were in the neighborhood of the correct prediction.

Through random forest, we identified several important meteorological indicators in prediction. The most significant feature contributing to drought shown by random forest was high surface pressure. This is because high pressure in atmospheric circulation causes air to sink towards the ground and rush towards low pressure zones, causing less clouds. As the sinking air disperses clouds which eliminates rain, surface pressure is a direct contributor to droughts (16).

The next most important features were maximum temperature and temperature range, both measured in Celsius at two meters above sea level. Temperature plays an important role in drought prediction as higher temperatures lead to an increase in evapotranspiration, a combination of evaporation from bodies of water and the release of water vapor from plants (17). This causes less water to soils and vegetation to dry out as well as a reduction in surface water which furthers the severity of droughts during low-precipitation periods (18).

To affirm which features were most important, we ran a sensitivity analysis on random forest with 200 trees and 30 depth (**Figure 7**). The importance of each feature differed when comparing both methods. This is due to the fact that sensitivity analysis investigates the impact of varying input variables on a model's predictions by removing one feature at a time, allowing for an understanding of the model's sensitivity to individual variables (19). On the other hand, random forest feature importance assesses the significance of features by measuring the reduction in model performance when a particular feature is altered or removed. This collective behavior analysis, known as "mean decrease impurity" or "Gini importance," provides insights into the relative importance of features in the random forest model. The results of the sensitivity analysis (**Figure 7**) indicated that precipitation does indeed have a significant influence on the occurrence of drought. However, it is interesting to note that the random forest algorithm itself did not assign a high rank to precipitation based on its feature importance metrics. This finding underscores the importance of considering multiple factors and adopting diverse approaches when evaluating the significance of predictors in complex phenomena like drought.

Additionally, the dataset does not include non-drought conditions, meaning we only trained the model on instances where there was drought. The lowest category of D0 states that there is minimal drought in the region, not none. This means that inputs of drought-free conditions will result in an incorrect categorization of likely D0. This limits the inputs to regions where drought is present and drought-free areas

JOURNAL OF
EMERGING INVESTIGATORS

https://doi.org/10.59720/ 23-026

Category	Description	Possible Impacts
D0	Abnormally Dry	**Going into drought:** • Short-term dryness slowing planting, growth of crops or pastures **Coming out of drought:** • Some lingering water deficits • Pastures or crops not fully recovered
D1	Moderate Drought	• Some damage to crops, pastures • Streams, reservoirs, or wells low, some water shortages developing or imminent • Voluntary water-use restrictions requested
D2	Severe Drought	• Crop or pasture losses likely • Water shortages common • Water restrictions imposed
D3	Extreme Drought	• Major crop/pasture losses • Widespread water shortages or restrictions
D4	Exceptional Dry	• Exceptional and widespread crop/pasture losses • Shortages of water in reservoirs, streams, and wells creating water emergencies

Table 2. Drought category description and possible impacts.

will not be able to be categorized correctly. We can combat this by integrating non-drought conditions in future training to allow the model to recognize drought-free data. This would expand the categories available for the model and permits more versatile inputs in the future.

The dataset has time stamped records of meteorological factors, which can be used to do a time series analysis of the data. In the future, the time series information that is a part of this dataset can be utilized to build forecasting models that consider not just the meteorological factors at a given point in time, but also their variation over a period of time. Other explainable algorithms such as Support Vector Machine or Naive Bayes can be tested with the dataset to compare their performance to KNN and random forest (20, 21).

In addition to the variables considered in this study, it is important to acknowledge that other factors, including soil moisture, land use, and vegetation cover, can significantly impact drought conditions. However, these factors were not incorporated into the current model. This limitation highlights the importance of an interdisciplinary approach to allow comprehensive drought prediction. By integrating knowledge from various fields and considering a broader range of factors, we can improve the accuracy and reliability of drought prediction models in the future.

MATERIALS AND METHODS

Dataset Description

The dataset used in this paper contains 19.3 million samples and 20 features and was collected by NASA Power project and the authors of U.S. Drought Monitor (12, 13). The U.S. Drought Monitor is a weekly map that shows the status and intensity of drought conditions across the U.S. It is produced by the National Drought Mitigation Center (NDMC) (22) at the University of Nebraska-Lincoln in partnership with the U.S. Department of Agriculture (USDA) (23), the National Oceanic and Atmospheric Administration (NOAA) (24), and other agencies. The Drought Monitor map is based on a combination of data sources, including satellite imagery, ground-based observations, and model outputs. It uses a standardized classification system to categorize drought conditions into five categories, D0 (abnormally dry) to D4 (exceptional drought) (**Table 2**) (25). Level D0 describes a region that is only abnormally dry, meaning regions in this category are typically going into or coming out of drought. Level D1 is moderate drought where there is minimal damage to crops and pastures, and water shortages are developing. This causes voluntary water-use restrictions to be requested of residents within the affected region. Level D2 is severe drought, where crop and pasture losses are likely and water shortages prompt restrictions to take effect. Level D3 is extreme drought that causes major crop and pasture losses which leads to widespread water shortages and/or mandatory restrictions. Level D4 is exceptional drought characterized by exceptional and widespread crop and pasture losses, as well as water shortages in reservoirs, streams, and wells, and water emergencies.

We visualized the distribution of each feature in the dataset to gain insights into their characteristics (**Figure 2**). Furthermore, to better understand the distribution of the five label classes, we applied principal component analysis (PCA) to project the features into a two-dimensional space (**Figure 3**). The visualization obtained through PCA highlights the considerable overlap between different classes, indicating the complexity of predicting drought. This overlap underscores the need for a powerful machine learning algorithm to tackle the challenging task of drought prediction.

There are 18 meteorological indicators in the data (**Table 1**). These indicators were used to predict drought severity by running the dataset through different algorithms. Each indicator is numerical and was analyzed to understand its range (**Figure 2**). Many indicators had a large range of data spread across the various categories, while others had a more confined range with high frequency.

Dataset Cleaning

Rows with missing values were removed. The raw dataset had a heavy skew in the label column, where label skew refers to the imbalance in the distribution of classes or labels in a dataset (Figure 1). Label skew can have a number of negative effects on the performance of machine learning models. For example, if a model is trained on a dataset with label skew, it may be more prone to making predictions for the dominant class and less accurate for the minority classes. This can lead to poor performance on unseen data or even bias in the model's predictions. To address this concern, the dataset was cleaned and balanced. The number of samples was set to 56,935 rows per drought category because this was the maximum number of samples we could use if we wanted the data to be perfectly balanced.

Algorithms

Two algorithms, K- nearest neighbors and random forest, were assessed for their predictive performance.

Random forest is a powerful machine learning algorithm that makes predictions based on the collective vote of multiple decision trees. A random forest algorithm consists of many decision trees inside it, where the number of trees is a user defined hyper-parameter. By considering the prediction made by each decision tree, random forest can make more accurate predictions and handle complex patterns in data (15).

In KNN, when a prediction is made for a new data point, the model finds the K nearest data points in the training set and uses these points to make a prediction. The prediction is based on the majority class among the K nearest neighbors. KNN classifier predictions can be explained by showing the data points in the training set that were used to make the prediction. This can provide insight into how the model arrived at its prediction and may help to build trust in the model's

https://doi.org/10.59720/ 23-026

output (14).

Another way to explain the prediction made by a KNN classifier is to look at the distances between the new data point and the K nearest neighbors. For example, if the K nearest neighbors are all very similar to the new data point, then this may be a strong indication that the model's prediction is reliable. On the other hand, if the K nearest neighbors are very different from the new data point, then this may indicate that the model's prediction may not be as reliable.

Overall, the KNN classifier is considered to be a relatively simple and explainable model, as it relies on a straightforward distance-based approach for making predictions. However, it does not display feature importance and solely relies on the values of surrounding points, which in this case may be unrelated, as one point could be depicting a location across the country from another point.

The value of K was varied from 1 to 19 to analyze the impact of the hyper-parameter K on the prediction of drought. The K value of one achieved the highest validation accuracy. This model was used to predict on the test dataset, which resulted in an accuracy of 52.21% (**Figure 5A**). When the dataset was normalized, the highest accuracy was 50.19% (**Figure 5B**).

The hyper-parameter values of the random forest, number of trees/estimators, varied between 10 and 300 in increments ranging from 10 to 50. The maximum tree depths were varied between 1 and 40 (**Figure 4**). The model with the highest accuracy used 200 trees with a depth of 30 (**Figure 4F**). This model was used on the test dataset which resulted in an accuracy of 56.09%.

The confusion matrix obtained by the best model for K-nearest neighbors and random forest shows that the model predicts the right drought level more than 50% of times (**Figure 6**). When the random forest model is incorrect, it typically predicts a drought level that is one level away from the true drought condition (**Figure 6A**). A confusion matrix is a simple but powerful tool used to evaluate the performance of a classification model. It provides a clear picture of how well the model is able to predict different classes or categories. The matrix is called a "confusion" matrix because it helps us understand where the model might be getting confused. The confusion matrix is usually presented as a table with rows and columns. The rows represent the true or actual classes, while the columns represent the predicted classes made by the model. Each cell in the matrix represents the number of instances that fall into a specific combination of true and predicted classes.

According to the random forest model, the most important factor in predicting the drought condition was PS (surface pressure), followed by T2M_RANGE (temperature range at two meters measured in Celsius) and T2M_MAX (maximum temperature at two meters measured in Celsius) (**Table 1**). Feature importance is calculated based on how much each feature reduces the impurity in the splits of the trees. Impurity represents how well the trees split the data. The impurity of a node is then used to determine the best split at each node of the decision tree, with the goal of creating nodes that are as pure as possible.

Code

Code used to perform the experiments is available at: https://github.com/clarissecheung/JEI_Drought_Prediction

Received: February 6, 2023
Accepted: July 6, 2023
Published: October 28, 2024

REFERENCES

1. “Drought Basics.” *Drought.gov*, drought.gov/what-is-drought/drought-basics. Accessed 30 Nov. 2023.
2. Mann, Michael E., and Peter H. Gleick. "Climate change and California drought in the 21st century." *Proceedings of the National Academy of Sciences* 112.13 (2015): 3858-3859. https://doi.org/10.1073/pnas.1503667112
3. Nuccitelli, Dana. “California, 'America's Garden,' Is Drying out " Yale Climate Connections.” *Yale Climate Connections*, 20 Oct. 2022, yaleclimateconnections.org/2021/06/california-americas-garden-is-drying-out/. Accessed 30 Nov. 2023.
4. Mosley, Luke M. “Drought Impacts on the Water Quality of Freshwater Systems; Review and Integration.” *Earth-Science Reviews*, vol. 140, Jan. 2015, pp. 203–214. https://doi.org/10.1016/j.earscirev.2014.11.010
5. “Historical Drought.” *Drought.gov*, drought.gov/what-is-drought/historical-drought. Accessed 30 Nov. 2023.
6. “National Current Conditions.” *Drought.gov*, drought.gov/current-conditions. Accessed 30 Nov. 2023.
7. Hao, Zengchao, et al. "Seasonal drought prediction: advances, challenges, and future prospects." *Reviews of Geophysics* 56.1 (2018): 108-141. https://doi.org/10.1002/2016RG000549
8. Wood, Eric F., et al. "Prospects for advancing drought understanding, monitoring, and prediction." *Journal of Hydrometeorology* 16.4 (2015): 1636-1657. https://doi.org/10.1175/JHM-D-14-0164.1
9. Chen, L. Gwen, et al. "Flash Drought Characteristics and Prediction." Climate Prediction S&T Digest (2019): 126.
10. “Monitoring Drought.” *Drought.gov*, drought.gov/what-is-drought/monitoring-drought. Accessed 30 Nov. 2023.
11. “Drought.” *Department of Water Resources*, 20 Dec. 2022, water.ca.gov/water-basics/drought. Accessed 30 Nov. 2023.
12. “NASA Power.” *NASA*, NASA, power.larc.nasa.gov/. Accessed 30 Nov. 2023.
13. “Current Map: U.S. Drought Monitor.” *Current Map | U.S. Drought Monitor*, droughtmonitor.unl.edu. Accessed 30 Nov. 2023.
14. Zhang, Zhongheng. "Introduction to machine learning: k-nearest neighbors." *Annals of translational medicine* 4.11 (2016). https://doi.org/10.21037/atm.2016.03.37
15. Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32. https://doi.org/10.1023/A:1010933404324
16. Le Houérou, Henry N. "Climate change, drought and desertification." *Journal of arid Environments* 34.2 (1996): 133-185. https://doi.org/10.1006/jare.1996.0099
17. “Temperature & Precipitation.” *Drought.gov*, drought.gov/topics/temperature-precipitation. Accessed 30 Nov. 2023.
18. Hanson, Paul J., and Jake F. Weltzin. "Drought disturbance from climate change: response of United States forests." *Science of the total environment* 262.3 (2000): 205-220. https://doi.org/10.1016/S0048-9697(00)00523-4
19. Christopher Frey, H., and Sumeet R. Patil. "Identification and review of sensitivity analysis methods." *Risk analysis* 22.3 (2002): 553-578. https://doi.org/10.1111/0272-

https://doi.org/10.59720/ 23-026

4332.00039
20. Pisner, Derek A., and David M. Schnyer. “Support Vector Machine.” *Machine Learning* (2020): 101-121. https://doi.org/10.1016/B978-0-12-815739-8.00006-7
21. Frank, E., Trigg, L., Holmes, G. et al. Technical Note: Naive Bayes for Regression. *Machine Learning* 41, 5–25 (2000). https://doi.org/10.1023/A:1007670802811
22. “About Us.” *About Us | National Drought Mitigation Center*, drought.unl.edu/AboutUs.aspx. Accessed 30 Nov. 2023.
23. “About the U.S. Department of Agriculture.” *USDA*, usda.gov/our-agency/about-usda. Accessed 30 Nov. 2023.
24. “About Our Agency.” *National Oceanic and Atmospheric Administration*, noaa.gov/about-our-agency. Accessed 30 Nov. 2023.
25. “Explaining Drought Category Maps.” *Drought.gov*, drought.gov/explaining-drought-category-maps. Accessed 30 Nov. 2023.