U.S. Drought Projection

Result (200 characters)

We predict the U.S. Palmer Drought Severity Index (PDSI) from 2024 to 2050 using the scikit-learn library. Our model indicates that the southern-western regions are at an increased risk of drought.

Context and Parameters

Key	Value
Years – Training Data Period	2012 - 2023
Years – Prediction Period	2024 - 2050
Prediction Model	Support Vector Machines (sklearn)
Country	U.S.A.
Number of Climate Regions	355

Figures (5 figures max, separate files, URLs only, one plot or diagram per figure, 4:3 aspect ratio)

1. 2022 Actual PDSI
   https://drive.google.com/file/d/1I1Bgs2YvpA7ty1bXaDx26bn1Yhkty_E4/view?usp=drive_link
2. 2022 Forecasted PDSI
   https://drive.google.com/file/d/1GC2aKnFq3Hif3cwIwiiHkfwqA0wWQVyq/view?usp=drive_link
3. 2023 Actual PDSI
   https://drive.google.com/file/d/1MgDo5cilfz7R_Ku-73oEdN5YUssXqVRN/view?usp=drive_link
4. 2023 Forecasted PDSI
   https://drive.google.com/file/d/149QbqdVsFwtr7Do4d3NYcT19dUJd-ePz/view?usp=drive_link
5. 2024-2050 Forecasted PDSI
   https://drive.google.com/file/d/1ZA9NCYYLzKIJOBGS83rn_b69mMhlgwUC/view?usp=drive_link

Description (3,000 characters)

This study introduces a machine learning-driven approach for predicting drought severity across the United States, leveraging the scikit-learn library with a particular focus on the support vector machine (SVM) model. We selected the SVM model after cross-validation testing, where it outperformed other machine learning models. Data for this analysis was acquired through the Google Earth Engine, drawing from five distinct datasets, which were then processed and transformed to mitigate skewness and ensure optimal predictive performance. The climate data aggregated monthly and regionally (355 climate regions in total) spans from January 2010 to December 2023, with additional climate projections extending from January 2024 to December 2050.

Key predictive features included minimum and maximum temperature, precipitation, minimum and maximum humidity, and wind speed. To enhance context, 3-, 6-, and 12-month average values for these variables were calculated across each geographic region. Data from 2011 to 2021 was designated for training, while 2022 and 2023 were reserved for model testing, ensuring a robust validation process.

Model performance was primarily assessed using Root Mean Square Error (RMSE), with SVM achieving the lowest RMSE of approximately 1.82. Note that the range of the mean values is 17.1. Testing across the entire test dataset yielded an RMSE of around 1.13, which decreased to approximately 1.06 following hyperparameter optimization (regularization increased from 10 to 20). The SVM model reached an RMSE as low as 0.71 on the test dataset, underscoring its ability to provide good fits.

Through feature importance analysis, adjustments to feature selection were made, improving the model's performance. These included the removal of less informative features (mean temperature and rainfall) and removal of 3-month average and the addition of 18- and 24-month average climate data, which together reduced RMSE to 0.97. Tests with principal component analysis (PCA) were conducted to assess potential benefits of dimensionality reduction, though PCA was not found to enhance model accuracy.

A final visual assessment was performed by mapping actual and predicted Palmer Drought Severity Index (PDSI) values across the United States for 2022 and 2023 (Figures 1 to 4). Additionally, the predicted PDSI values from 2024 to 2050 were animated (Figure 5) to illustrate the evolution of future drought trends.

These results underscore the potential of machine learning in environmental monitoring, offering an accessible and scalable tool for drought forecasting. This approach has practical implications for proactive water resource management and policy-making, aiding decision makers in making informed decisions to mitigate future drought impacts.

References

1. Géron, Aurélien. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media, 2019.
2. https://cartographyvectors.com/map/1290-us-climate-regions
3. https://developers.google.com/earth-engine/tutorials/community/time-series-visualization-with-altair
4. https://www.kaggle.com/code/akshayasrinivasan2/drought-prediction-using-ml-algorithms

Protocols & Code

1. Regression analysis with Support Vector Machines
2. Visualization on US map
3. Source code available under the GitHub repository
   https://github.com/mertalpaydin/USA-Drought-Projection-with-scikit-learn

Data

1. https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_TERRACLIMATE#bands
2. https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_GRIDMET
3. https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD13A2#bands
4. https://developers.google.com/earth-engine/datasets/catalog/GRIDMET_DROUGHT#bands
5. https://developers.google.com/earth-engine/datasets/catalog/NASA_GDDP-CMIP6#bands

Extracted Data is also available under the GitHub repository