Forecasting Drought Area Percentage in California using Machine Learning Algorithms

Rama Krishna Poluru, Dharma Teja Kolluri, Saranya Gondeli and Manish Kumar Sesetti

Department of Applied Data Science, San Jose State University

DATA 270 : Data Analyst Process

Dr. Eduardo Chan

April 28, 2023

Abstract

California is highly susceptible to frequent, severe droughts that significantly impact ecosystems, agriculture, and water supplies. To identify the adverse effects, it's crucial to forecast future drought scenarios based on past occurrences in California. Prior research has shown that using SPEI to predict drought is viable, with the majority of studies limited to predicting SPEI or SPI values but not the actual drought percentage. This research aimed to expand on the previous research by creating an accurate model for predicting the area percentage of drought-affected regions beyond just SPEI values for each county from 2000 to 2020 by utilizing climatic, drought percentage, and NDVI values. The calculation of SPEI values involves using Thornthwaite equation along with climatic and NDVI data. After preparing the dataset with climatic and SPEI values, Decision Tree Regressor with XGBoost, Random Forest, LSTM, and ANN models are used to forecast drought area percentage for California. The Random Forest model outperforms other models with MAE value of 6.81 and R^2 value of 0.74. ANN and decision tree models achieved good results, with MAE of 7.43 and 7.21, and an R^2 score of 0.70 and 0.72, respectively. These results suggest that decision tree-based models and traditional neural networks are effective in predicting the target variable. The findings indicate that RF model can assist in identifying drought-prone areas and help to implement effective mitigation measures.