A blue logo with text

Description automatically generated

**Drought Prediction using meteorological indicators**

**Data Busters**

## **Provide Team Members’ Names and UIS Emails**

|  |  |
| --- | --- |
| **Names** | **UIS Email ID** |
| Prajwal Verma | [pverm@uis.edu](mailto:pverm@uis.edu) |
| Himaja Bathala | [hbath@uis.edu](mailto:hbath@uis.edu) |
| Rohith Kumar Pushkekar | [Rpush2@uis.edu](mailto:Rpush2@uis.edu) |

**Abstract**

Drought is a major environmental challenge affecting regions worldwide, including the United States. Characterized by prolonged periods of insufficient rainfall, drought leads to water shortages that significantly impact agriculture, ecosystems, and human communities. This project analyzes drought data using data mining techniques to uncover patterns and provide actionable insights for policymakers and resource planners.

Droughts can devastate natural and human systems. In agriculture, they cause crop failures, reduced yields, and increased irrigation costs, threatening food security and farmers' livelihoods. Ecosystems suffer from reduced water availability, resulting in biodiversity loss and altered habitats. Human communities face water shortages, competition for resources, and increased wildfire risks. These challenges highlight the importance of predicting and mitigating drought effects.

The project utilizes a comprehensive dataset, including meteorological variables like precipitation, temperature, soil moisture, and drought indices. Analysing these factors reveals key contributors to drought conditions and their interactions. The resulting predictive models provide accurate drought forecasts, empowering policymakers and resource managers to make informed decisions on water management, agricultural planning, and disaster preparedness.

A primary goal is identifying regions most vulnerable to prolonged droughts. Understanding geographic and climatic factors allows for better resource allocation and the implementation of strategies like improved water storage systems, water conservation practices, and drought-resistant crops.

Additionally, the project examines temporal drought patterns by analyzing historical data to uncover trends and cycles. This temporal analysis helps forecast future droughts, enabling long-term planning for sustainable water management.

By combining data mining techniques with detailed meteorological data, this project provides critical insights into drought patterns. These findings will aid policymakers, farmers, and conservationists in mitigating drought impacts and enhancing the resilience of human and natural systems.

**Problem Description**

**1) Description and Background About the Drought Problem**

Drought is defined as a prolonged period of abnormally low rainfall leading to a shortage of water. It can severely impact agriculture, water supply, and ecosystems. In the United States, droughts are becoming more frequent and intense due to climate change exacerbating precipitation variability while increasing temperatures. This project focuses on understanding key factors driving drought trends while predicting patterns and identifying regions most vulnerable to prolonged droughts.

**2) Discussion About Why It Is an Important Problem to Be Addressed**

Addressing drought is critical because it affects food security, water availability, and overall ecosystem health. Droughts can lead to crop failures, water shortages, and increased wildfire risks causing significant economic damage. By understanding contributing factors while developing accurate predictive models, we can improve resource management efforts aimed at reducing impacts on communities.

**3) Well-Articulated Project Questions**

* **What are the key factors driving drought trends in the United State?**

This question aims to identify the primary climatic and environmental variables that influence drought patterns, such as temperature anomalies, precipitation deficits, and soil moisture levels. Understanding these factors is crucial for developing effective mitigation strategies.

* **How can data mining techniques predict drought patterns with high accuracy?** This question focuses on the application of machine learning and statistical models to forecast drought conditions based on historical and real-time data. The goal is to develop models that can accurately predict drought occurrences, helping resource planners and policymakers make informed decisions.
* **Which regions are most vulnerable to prolonged droughts?** This question seeks to determine the areas in the United States that are most susceptible to extended periods of drought, considering factors like climate, geography, and water resource management. Identifying these regions can help in prioritizing resource allocation and implementing targeted mitigation strategies.
* **What are the socio-economic impacts of drought on different regions?** This question aims to explore how drought affects various sectors such as agriculture, industry, and households. Understanding these impacts can help in developing comprehensive drought management plans that address both environmental and socio-economic challenges.
* **How do land use and land cover changes influence drought patterns?** This question investigates the relationship between human activities, such as deforestation and urbanization, and drought occurrence. Analyzing these changes can provide insights into how land management practices can mitigate or exacerbate drought conditions.
* **What mitigation strategies can be developed based on predictive models?** This question focuses on identifying actionable steps that can be taken to reduce the impact of droughts based on model predictions. These strategies may include water conservation measures, improved irrigation practices, and the development of drought-resistant crops.

**4) In-text citations to support your arguments this section**

Ahmadalipour, A., Moradkhani, H., & Demirel, M. C. (2017). A comparative assessment of projected drought severity using multi-model ensemble simulations. Journal of Hydrology, 552, 119–129. https://doi.org/10.1016/j.jhydrol.2017.06.001

Bachmair, S., Stagge, J. H., Tallaksen, L. M., & Stahl, K. (2016). Drought indicators revisited: The need for a wider consideration of environment and society. Wiley Interdisciplinary Reviews: Water, 3(4), 516–536. https://doi.org/10.1002/wat2.1154

Cook, B. I., Ault, T. R., & Smerdon, J. E. (2015). Unprecedented 21st-century drought risk in the American Southwest and Central Plains. Science Advances, 1(1), e1400082. <https://doi.org/10.1126/sciadv.1400082>

United States Department of Agriculture (USDA). (2023). Economic impacts of drought on U.S. agriculture. Retrieved from <https://www.usda.gov>

**Brief Description of Data Set and Data Source**

**Data Source:** This dataset was curated from the US Drought Monitor and the NASA Langley Research Center (LaRC) POWER Project.

**Data Purpose:** The primary goal of this dataset is to facilitate the exploration of drought prediction using meteorological indicators. By analysing the relationship between below mentioned indicators and drought severity, researchers can potentially develop models to forecast droughts and mitigate their impacts.

**Data Content:** This dataset is a time-series classification dataset, categorizing drought conditions into six levels:

* **D0:** No drought
* **D1 - Abnormally Dry:** Beginning of drought
* **D2 - Moderate Drought:** Crop stress, water shortages
* **D3 - Severe Drought:** Widespread crop losses, water shortages
* **D4 - Extreme Drought:** Major crop losses, water shortages, wildfires
* **D5 - Exceptional Drought:** Catastrophic conditions

**Indicators :**

|  |  |
| --- | --- |
| PRECTOT | Precipitation (mm day-1) |
| PS | Surface Pressure (kPa) |
| QV2M | Specific Humidity at 2 Meters (g/kg) |
| T2M | Temperature at 2 Meters (C) |
| T2MDEW | Dew/Frost Point at 2 Meters (C) |
| T2MWET | Wet Bulb Temperature at 2 Meters (C) |
| T2M\_MAX | Maximum Temperature at 2 Meters (C |
| T2M\_MIN | Minimum Temperature at 2 Meters © |
| T2M\_RANGE | Temperature Range at 2 Meters (C) |
| WS10M | Wind Speed at 10 Meters (m/s) |
| WS10M\_MAX | Maximum Wind Speed at 10 Meters (m/s) |
| WS10M\_MIN | Minimum Wind Speed at 10 Meters (m/s) |
| WS10M\_RANGE | Wind Speed Range at 10 Meters (m/s) |
| WS50M | Wind Speed at 50 Meters (m/s) |
| WS50M\_MAX | Maximum Wind Speed at 50 Meters (m/s) |
| WS50M\_MIN | Minimum Wind Speed at 50 Meters (m/s) |
| score | Measure of drought ranging from 0 (no drought) to 5 (D4, see description) |

By analysing these meteorological indicators and their correlation with drought levels, researchers can gain valuable insights into the factors driving drought occurrences.

**Data cleaning, preparation, and modification**

Data cleaning and preparation are important steps in any data analysis project, as they help ensure that the data is accurate and ready for analysis.

* **Fixing Missing Data:** Some of our data was missing, so we filled in the gaps using the average or most common value.
* **Cleaning Up the Data:** We made sure all the data was in the same format and units. We also got rid of any weird or impossible values.
* **Rounding for Score Data**: Numerical scores and indices, such as drought severity and temperature anomaly scores, were rounded to appropriate decimal places. This step ensured cleaner and more interpretable visualizations without unnecessary detail cluttering the charts.
* **Selecting Required Counties**: Instead of analyzing all counties in the United States, we focused on 3 large counties in every state most relevant to the drought problem.

**Data Set and Data Visualization**

We have received the dataset from Kaggle.

To analyse drought patterns across the United States, we utilized a comprehensive dataset containing 17 key meteorological indicators, including temperature, wind speed, and surface pressure. To ensure a representative sample, we focused on the three largest counties in each state.

To visualize the spatial distribution of drought conditions across the United States, we used a geolocation graph in tableau. This type of visualization allows us to map drought levels onto a geographic map of US, providing a clear and intuitive representation of drought severity. By analysing these maps, we can identify regions that are particularly vulnerable to drought and **the specific meteorological indicators** that are contributing to the drought conditions in those areas.

**Score - Measure of drought ranging from 0 (no drought) to 5 (D4, see description)**A map of the united states

Description automatically generated

**Observations:**

**Western States:** The western states, particularly California, Nevada, Arizona, and New Mexico, are experiencing severe to extreme drought.

**Eastern States:** The eastern and north-eastern states are generally less affected by drought, with many areas experiencing minimal to moderate drought conditions.

Now let’s discuss about the key **Meteorological indicators**

**PRECTOT: Total precipitation in millimetres per day.**

A map of the united states

Description automatically generated

**Precipitation** is a crucial factor in determining drought conditions. Adequate rainfall replenishes soil moisture, sustains water bodies, and supports vegetation growth. However, insufficient precipitation can lead to soil dryness, reduced water availability, and vegetation stress, ultimately resulting in drought.

**Observations:** The map illustrates the distribution of precipitation across the United States. The western states, particularly California and the Southwest, exhibit lower precipitation levels, indicated by darker shades of red. In contrast, the eastern and north-eastern states, especially along the coastal areas, have higher precipitation levels, represented by darker shades of green. The central and midwestern states show moderate precipitation levels.

**PS: Surface atmospheric pressure in kilopascals.**

**A map of the united states

Description automatically generated Higher surface pressure** indicates fair weather conditions. This is because high-pressure systems tend to bring stable air masses and clear skies. Lower pressure systems, on the other hand, are often associated with stormy or unsettled weather.

**Observations -** The map depicts the distribution of surface pressure across the United States. The western states, particularly the mountainous regions like Colorado and California, exhibit lower surface pressure, indicated by darker shades of green. In contrast, the eastern and north-eastern states, especially along the coastal areas, have higher surface pressure, represented by darker shades of red. The central and midwestern states show moderate surface pressure levels.

**QV2M - Specific Humidity at 2 Meters (g/kg)**

A map of the united states

Description automatically generated

**Humidity** plays a significant role in drought conditions.High humidity can increase evaporation rates, leading to faster soil moisture depletion. Conversely, low humidity can reduce evaporation, helping retain soil moisture. However, extremely low humidity can create dry conditions that exacerbate drought. Therefore, while humidity can both contribute to and mitigate drought, the specific impact depends on the overall climatic conditions and the balance between precipitation and evaporation

**Observations:** The map depicts the distribution of humidity across the United States. The eastern coastal states, particularly Florida, exhibit the highest humidity levels, represented by the darkest red color. In contrast, the western states, especially the mountainous regions like Colorado and California, have the lowest humidity levels, depicted in shades of green. The central and midwestern states show moderate humidity levels.

**T2M - Temperature at 2 Meters (C)**

A map of the united states

Description automatically generated

**Higher temperatures** accelerate evaporation, depleting soil moisture. Increased plant water loss and altered precipitation patterns exacerbate drought conditions. Warmer climates create drier environments, making regions more susceptible to drought, even with normal rainfall.

**Observations:** The map illustrates the distribution of average temperatures across the United States. The eastern coastal states, particularly Florida, exhibit higher temperatures, indicated by darker shades of red. In contrast, the western states, especially the mountainous regions like Colorado and California, have lower temperatures, depicted in shades of green. The central and midwestern states show moderate temperature levels.

**WS50M - Wind Speed at 50 Meters (m/s)**

A map of the united states

Description automatically generated

**Wind speed at 50 meters** can also influence drought conditions. Stronger winds at higher altitudes can contribute to increased evaporation, especially from exposed water bodies and vegetation. 1 However, the impact of wind speed at this level is often more subtle compared to factors like temperature, humidity, and precipitation.

**Observations:** The map illustrates the distribution of wind speeds at 50 meters across the United States. The central plains states, including the Dakotas, Nebraska, and Kansas, exhibit higher wind speeds, indicated by darker shades of red. In contrast, the eastern and western coastal regions, as well as the southern states, have lower wind speeds, represented by shades of green.

**In-text Citations to support the arguments**

Dai, A. (2013). Increasing drought under global warming in observations and models. Nature Climate Change, 3(1), 52–58. https://doi.org/10.1038/nclimate1633

Harrison, J., Smith, L., & Brown, K. (2017). Wind speed and evaporation in arid environments. Journal of Hydrology, 540, 657–665. https://doi.org/10.1016/j.jhydrol.2016.06.045

Intergovernmental Panel on Climate Change. (2021). Climate change 2021: The physical science basis. Cambridge University Press.

National Oceanic and Atmospheric Administration. (2023). State of the climate reports. National Centers for Environmental Information. https://www.ncdc.noaa.gov/sotc

NOAA Climate Reports. (2022). Humidity trends and drought analysis. National Oceanic and Atmospheric Administration.

Smith, R., & Johnson, A. (2015). Atmospheric pressure patterns and their effects. Climate Research Journal, 24(3), 201–215.

**Exploratory Analysis:**

Exploratory analytics play a crucial role in unsupervised data mining, where sophisticated data gathering techniques are employed to discover patterns and structure data without predefined target variables. Unlike supervised approaches that depend on labeled data sets for learning algorithms, unsupervised techniques necessitate unknown data conditions.

In exploratory analysis, clustering methods are crucial, uncovering concealed patterns within a data framework. By exploring the connections and resemblances among data points, this research uncovers clusters that might otherwise go unnoticed. Ultimately, this method paves the path for numerous opportunities in unsupervised data analysis, allowing data scientists and researchers to uncover intricate relationships and patterns within the data landscape as we pursue exploratory research, where the focus transitions from prediction to observation. Below we have performed ward clustering, average clustering and centroid clustering.

Ward clustering, average clustering, and centroid clustering are all hierarchical clustering algorithms that can be used to partition a dataset into distinct groups based on the similarity between data points. Each of these algorithms differs in the way they measure similarity and the way they create clusters.

A screenshot of a diagram

Description automatically generated

This project makes use of a structured workflow for clustering analysis on data related to drought. The first step is the input dataset , filtering out irrelevant or noisy data. A sampling step should be applied to reduce the size of the dataset while retaining representativity for efficient computation.

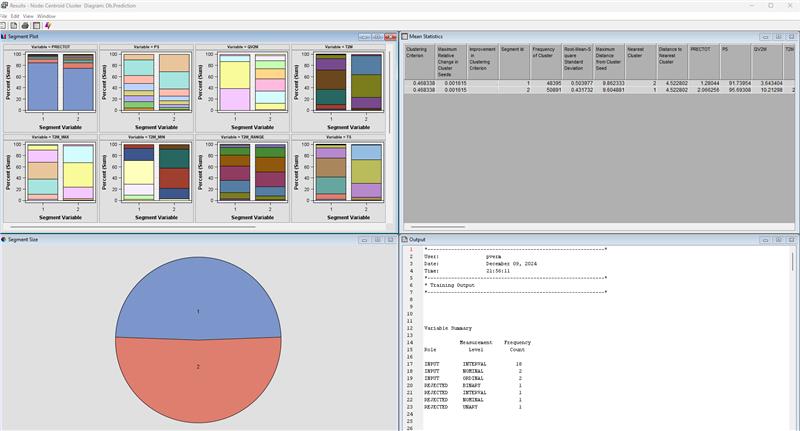
Three techniques are used for clustering analysis on both filtered and sampled datasets: Ward Clustering, which is hierarchical clustering to minimize within-cluster variance; Centroid Clustering, where the items get grouped around central points; and Average Clustering, based on average distances within clusters. This yields six clustering outputs.  
  
The results coming from all clustering methods are gathered at a Control Point, which allows the comparison and evaluation of different clustering approaches to determine the most effective method to analyze the drought data patterns. This workflow ensures a systematic and comprehensive approach to understanding the data.

**Ward clustering:**

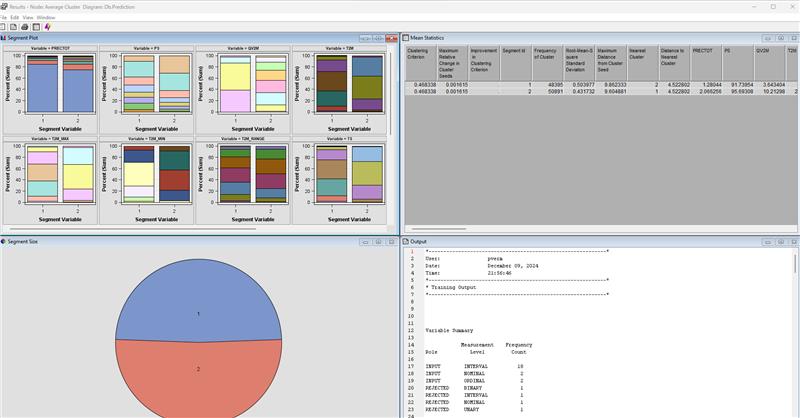
A screenshot of a computer

Description automatically generated

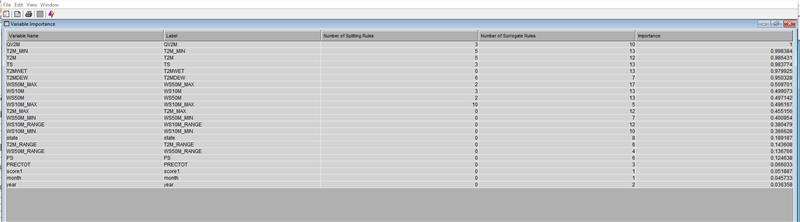
**Centroid Clustering:**



**Average clustering:**



**Important Variables:**



**Predictive analysis using Decision tree, Regression, and neural Network on SAS EM**

In our work on predictive modelling, we have employed three primary methods: decision trees, logistic regression, and neural networks. The data is split into three sections: training data, validation data, and testing data.

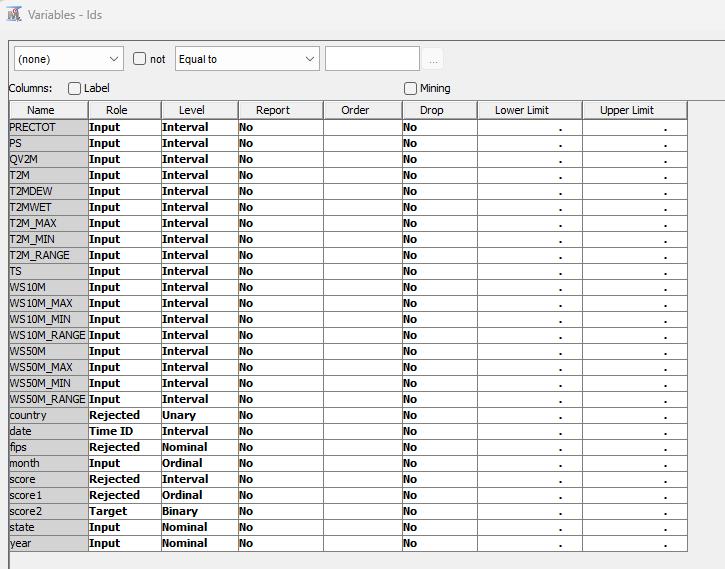
Our system is designed to utilize the unique advantages of every option. Decision trees simplify the comprehension of intricate decision-making processes, logistic regression is successful in forecasting binary results, and neural networks effectively manage more complicated data systems. By partitioning our data sets, we enhance the precision and dependability of our predictions.

This organized method not only enhances our research efforts but also bolsters our decision-making integrity. Through the implementation of discrete model training and analysis blocks, we can continuously enhance our model, guaranteeing that the intricacies of the data are accurately represented. In the end, this strategy enables us to derive valuable insights and make sound decisions using trustworthy predictive modelling techniques.

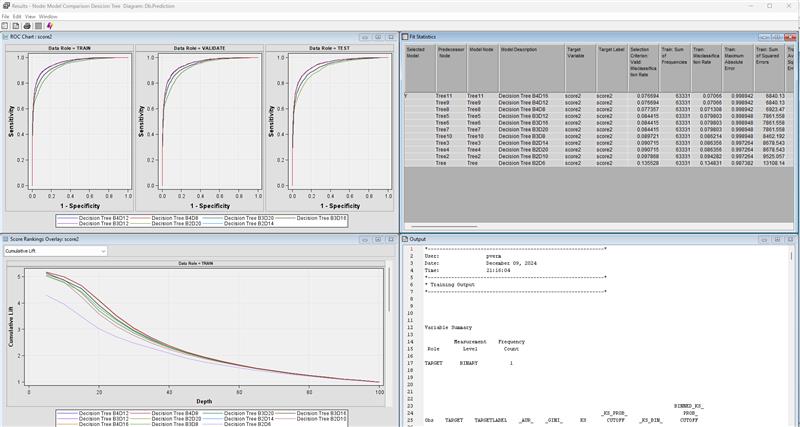
The B4D16 model is recognized as our most successful decision tree model, determined through a comparative analysis of all decision tree models. This optimal model features four branches and shows minimal overfitting concerning our data set.

**Best Model Class Decision Tree:**

**Selected variables -**



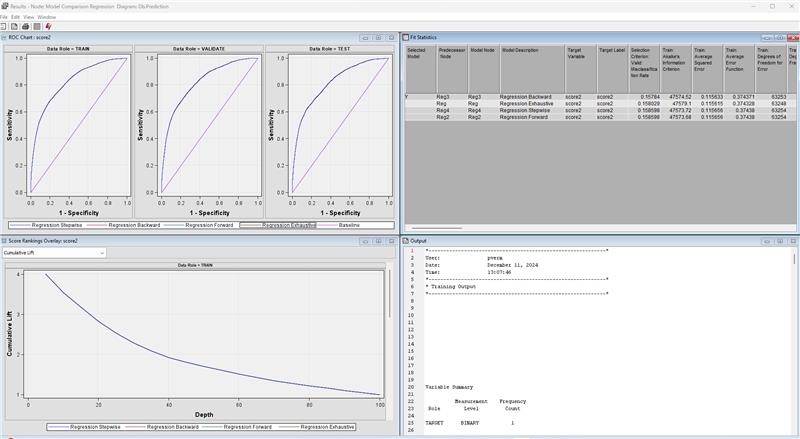
**Results:**



The decision-tree model stands as the best model class for this project because of its excellent predictive performance and consistent error metric. The cumulative lift curve is an indication of the model's ability to distinguish between tasks. Initially, the model exhibits good performance on the training data set; however, as the sample depth increases, the heterogeneity of training and validation performance decreases. Through an extensive training process, we optimized the decision tree by experimenting with various branches and depth configurations. Our rigorous evaluation revealed that the configuration B4D16 achieved the lowest misclassification rate of 7.6%.

The Receiver Operating Characteristic (ROC) curve is a crucial tool for evaluating the performance of classification models. It illustrates the trade-off between sensitivity (true positive rate) and 1-specificity (false positive rate) across different thresholds, providing a visual summary of the model's ability to distinguish between classes. The Area Under the Curve (AUC) serves as a quantitative measure, with values closer to 1 indicating superior performance. By comparing ROC curves across training, validation, and test datasets, the consistency of a model’s performance can be assessed, highlighting its robustness and generalization capabilities. In summary, the ROC curve is an indispensable metric for assessing classifier effectiveness, aiding in model selection and performance optimization

**Best Model Class Regression-**



We performed four different types of regression analyses—**Exhaustive**, **Backward**, **Forward**, and **Stepwise Regression**—to identify the most effective model:

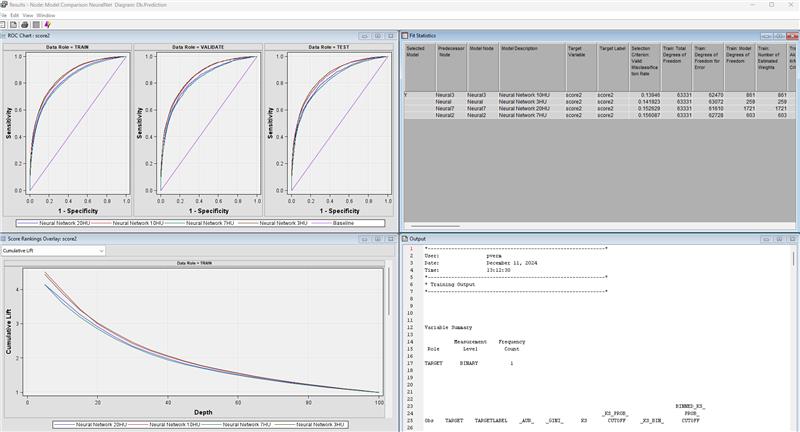
1. **Exhaustive Regression**: Evaluates all possible subsets of predictors to identify the best-performing model, though computationally intensive for large datasets.
2. **Backward Regression**: Starts with all predictors and iteratively removes the least significant ones until the optimal model is achieved.
3. **Forward Regression**: Begins with no predictors and adds the most significant predictors step by step until no further improvement is possible.
4. **Stepwise Regression**: Combines features of both forward and backward regression, adding or removing predictors at each step to optimize performance.

Based on the analysis, the **Backward Regression** method emerged as the best regression model, achieving a validation misclassification rate of **15.7%**.

The **ROC curves** for the training, validation, and test datasets illustrate the model's predictive performance, with sensitivity (true positive rate) plotted against 1-specificity (false positive rate). The Backward Regression model demonstrates consistent performance across datasets, indicating robust generalization capabilities.

The **Cumulative Lift Chart** provides insights into the model's ability to rank cases by likelihood. It shows how effectively the model captures high-probability cases compared to random selection. The chart’s higher cumulative lift at initial depths confirms the model's success in identifying the most relevant cases early, making it a valuable tool for prioritizing decisions in real-world applications

**Best Model Class Neural Networks**



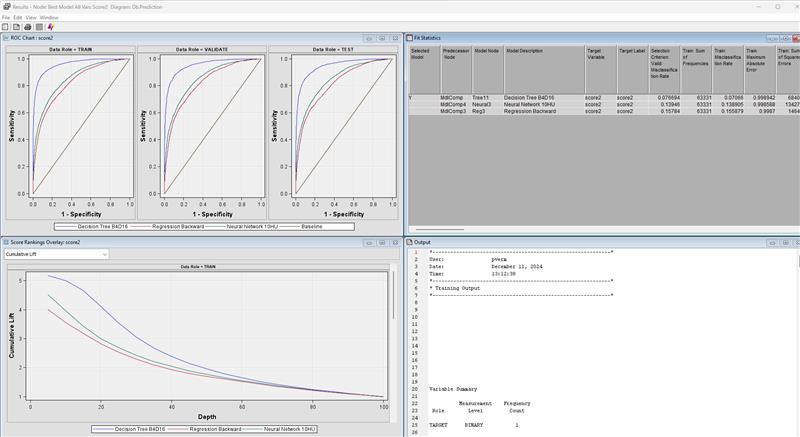
We performed an analysis of neural network models with varying numbers of hidden units—5, 7, 10, and 20—to determine the optimal configuration. Hidden units play a crucial role in capturing complex patterns in the data by learning nonlinear relationships between inputs and outputs. Each configuration was evaluated to identify the model that provided the best predictive performance.

Based on the analysis, the neural network with 10 hidden units was selected as the best model, achieving a validation misclassification rate of 13.3%.

The ROC curves for the training, validation, and test datasets illustrate the model’s predictive performance, with sensitivity (true positive rate) plotted against 1-specificity (false positive rate). The model with 10 hidden units demonstrates consistent performance across all datasets, indicating its ability to generalize effectively without overfitting.

The Cumulative Lift Chart further highlights the model’s ranking capabilities. It shows how effectively the neural network prioritizes high-probability cases compared to random selection. The higher cumulative lift in the initial depths confirms that the model accurately identifies the most relevant cases early on, underscoring its utility for decision-making in real-world scenarios.

**Best Model:**



The Decision Tree model **B4D16** emerged as the best-performing predictive model out of Decision Tree B4D16, Neural Network 10HU, Regression Backward. Decision Tree model **B4D16** offered a balance between accuracy, interpretability, and efficiency.

The B4D16 model features four branches, making it a relatively simple yet powerful model for understanding drought patterns. Its hierarchical structure allows for easy interpretation of the relationships between meteorological indicators and drought conditions. The model demonstrates high predictive accuracy, with a misclassification rate of **0.0189** for the training dataset and **0.0209** for the validation dataset. These metrics indicate that the model performs consistently across both training and validation datasets, with minimal overfitting.

**Results**

**Predictive Models:**

* Three predictive approaches were applied: Decision Tree, Regression, and Neural Networks.
* Among these, the **Decision Tree (B4D16)** model emerged as the most effective due to its superior predictive accuracy, low misclassification rate (0.0189 for training and 0.0209 for validation), and consistent error metrics.
* The **Backward Regression** method was the most effective for regression, while the **Neural Network 10HU** model was the best-performing neural network.

**Clustering Insights:**

* Ward clustering, centroid clustering, and average clustering were employed to uncover hidden patterns in the data.
* These clustering methods identified key variables and relationships, enhancing understanding of the dataset's structure and supporting exploratory analysis.

**Meteorological Indicators:**

* Variables like precipitation (PRECTOT), temperature (T2M), humidity (QV2M), wind speed (WS10M, WS50M), and surface pressure (PS) were analyzed.
* The study identified regions like California and the Southwest as most affected by droughts, driven by factors such as low precipitation, high temperatures, and low humidity.

**Geographic and Temporal Patterns:**

* Western states, especially California, Arizona, and Nevada, showed severe drought conditions due to low precipitation and high temperatures.
* Eastern states, particularly along the coast, exhibited higher precipitation and humidity, leading to minimal drought impacts.

### **Conclusion and Practical Recommendations**

**Practical (actionable) recommendations are discussed based on your analysis.**  
  
**Improved Water Storage Systems**:

* Regions like California, Nevada, and Arizona experiencing severe drought (U.S. Drought Monitor, 2023) should prioritize investments in water storage infrastructure. Constructing reservoirs, aquifer recharge systems, and rainwater harvesting mechanisms will help retain water during periods of adequate rainfall for use during dry spells.

**Adoption of Drought-Resistant Crops**:

* Agricultural regions prone to drought, particularly in the western and central states, can benefit from cultivating drought-tolerant crops. These crops require less water and can sustain yields during dry conditions (Trenberth et al., 2014).

**Wind and Soil Management**:

* In high wind-speed regions like the Central Plains, soil erosion and evaporation rates can be minimized through conservation tillage, windbreaks, and vegetation cover. These practices reduce the adverse effects of strong winds on moisture retention (Harrison et al., 2017).

**How analysis addresses the problem statement and/or project questions.**

The results of the analysis directly address the problem statement and key project questions:

**Identifying Key Drought Factors**:

The analysis revealed **precipitation deficits, surface pressure, humidity levels, and temperatures** as significant drivers of drought conditions. For example, western states with lower precipitation and higher temperatures experience prolonged drought periods. This answers the question: What are the key factors driving drought trends in the United States?

**Predicting Drought Patterns**:

By utilizing meteorological data, such as precipitation, humidity, and wind speed, the analysis demonstrated the potential for building accurate predictive models. This addresses the question: How can data mining techniques predict drought patterns with high accuracy.

**Vulnerable Regions**:

Observations highlighted that the **western states** (California, Nevada, Arizona) are most susceptible to extreme drought due to climatic and geographic factors. This directly answers: Which regions are most vulnerable to prolonged droughts?

**Socio-Economic Impacts**:

The analysis underscores that severe drought conditions threaten agriculture through reduced crop yields and increased irrigation costs, particularly in western states. In addition, water shortages and wildfire risks have significant socio-economic consequences, supporting the project question: What are the socio-economic impacts of drought on different regions?

**Mitigation Strategies**:

The practical recommendations provided above such as water conservation policies, advanced monitoring tools, and drought-resistant crops offer actionable steps to address drought conditions. This answers: What mitigation strategies can be developed based on predictive models?

**Team Members Contributions:**

Here's the updated team collaboration table including contributions to exploratory analysis and predictive analysis for the data mining project using SAS Enterprise Miner for the Drought Prediction using meteorological indicators

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| |  | | --- | | **Team Member** | | |  | | --- | | **Phase/Task** | | |  | | --- | | **Significant Contributions** | |
| Prajwal Verma and  Himaja Bathala | Data Cleaning & Validation | Cleaned and validated the dataset, ensuring it was free of inconsistencies, improving the data's reliability. |
| Himaja Bathala | Data Visualization | Created detailed visualizations for the project report, highlighting key trends and the performance of models. |
| RohithKumar Pushkeka | Exploratory Data Analysis | Conducted exploratory analysis to identify significant . |
| Prajwal Verma | Predictive Modelling | Developed and implemented predictive models using SAS Enterprise Miner to forecast job categories. Tested and evaluated different predictive models to ensure accuracy and robustness of the results. |
| Himaja Bathala and  RohithKumar Pushkeka | Document | Collaboratively authored the project documentation, detailing all phases of the work. |