SANTA CLARA UNIVERSITY

Date: June 12, 2023

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY

Atiyut Khemkhon
Dylan Lin

ENTITLED

Climate Cloud

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

COMPUTER SCIENCE AND ENGINEERING

Sean Choi
Sean Choi (Jun 14, 2023 13:33 PDT)

Thesis Advisor

Thesis Advisor

N. Ling
N. Ling (Jun 15, 2023 01:54 PDT)

Department Chair

Department Chair

Climate Cloud

by

Atiyut Khemkhon
Dylan Lin

Submitted in partial fulfillment of the requirements
for the degree of
Computer Science and Engineering
School of Engineering
Santa Clara University

Santa Clara, California
June 12, 2023

Climate Cloud

Atiyut Khemkhon
Dylan Lin

Santa Clara University
June 12, 2023

ABSTRACT

The increasing frequency and intensity of drastic climate change events have underscored the urgency for a comprehensive understanding of climate dynamics and the development of effective mitigation and adaptation strategies. In this context, our thesis addresses the critical issue of drought category prediction in Western America using machine learning techniques. We recognize the interconnectedness of efforts aimed at combating climate change and strive to make a meaningful contribution to the broader understanding of climate change, even if our impact may seem comparatively smaller. By leveraging the power of machine learning algorithms and utilizing a wide range of climatic and environmental variables, we aim to develop a robust predictive model that can accurately classify regions into distinct drought categories. The model's predictions have significant implications for water resource management, agriculture, ecosystems, and human livelihoods, enabling informed decision-making and proactive measures to mitigate the adverse consequences of droughts.

Through a review of existing methods, we highlight the need for advanced machine learning techniques in drought prediction. Our thesis focuses on the development of a machine learning model that utilizes climate variables, soil moisture levels, vegetation indices, and other relevant spatial and temporal features to predict drought categories at specific coordinates in Western America. By harnessing the vast amount of available data and leveraging the capabilities of machine learning algorithms, we aim to overcome the limitations of existing approaches and provide accurate and timely predictions. The outcomes of this research have the potential to inform policy decisions, enhance resource allocation, and contribute to effective drought mitigation strategies in Western America, ultimately fostering resilience in the face of a changing climate.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

In recent years, the world has witnessed an alarming increase in the frequency and intensity of drastic climate change events. These events, ranging from devastating hurricanes and wildfires to record-breaking heatwaves and extreme precipitation [21], have highlighted the urgent need for a comprehensive understanding of climate dynamics and the development of effective mitigation and adaptation strategies. The consequences of these events are far-reaching, affecting not only ecosystems and biodiversity but also posing significant threats to human lives, infrastructure, and economic stability. With each passing year, the severity and scale of these events serve as a stark reminder that urgent action is required to address climate change and its impacts before irreparable damage is done.

In light of the increasing gravity of these drastic climate change events, our group felt compelled to contribute to the broader understanding of climate change, even if our impact may seem comparatively smaller. Recognizing the interconnectedness of all efforts aimed at combating climate change, we embarked on a project that would leverage the power of machine learning to predict drought categories in Western America. By focusing on a specific region, we aimed to provide valuable insight into the complex dynamics of drought occurrences, contributing to the larger body of knowledge on climate change. We firmly believe that even incremental contributions, when combined with the collective efforts of scientists, policymakers, and communities worldwide, can drive meaningful change and pave the way for a more sustainable and resilient future.

Droughts are natural phenomena that have profound socio-economic and environmental impacts, particularly in arid and semi-arid regions such as Western America. The severity and duration of droughts pose significant challenges to water resource management, agriculture, ecosystems, and human livelihoods. Accurate and timely prediction of droughts plays a crucial role in developing effective mitigation strategies, optimizing resource allocation, and enhancing preparedness to minimize the adverse consequences of these extreme events. [15]

1.2 Background

Machine learning is a subfield of artificial intelligence that focuses on developing algorithms and models to enable computers to make predictions from data. It has gained prominence due to the exponential growth of data, advancements in computational power, and algorithmic breakthroughs. The core models of machine learning include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning[17]. Some popular machine learning algorithms include random forest, naive bayes, linear regression, and decision trees[14].

Supervised learning is a core concept of machine learning, where a model is trained on labeled data to recognize patterns and make accurate predictions on unseen data. This involves feeding the model with input data, allowing it to learn from examples, and adjusting its parameters for optimal performance. The goal is to develop models that generalize well and make accurate predictions on new, unseen data[19].

The process of supervised machine learning can be broken down into three components:

1. Decision process: A series of calculations are performed to determine the type of pattern to identify[20].
2. Error function: This function measures the accuracy of the model's predictions and determines how good its guesses are[20].
3. Optimization process: Based on the errors made, the model improves and adjusts its parameters to enhance performance[20].

1.2.1 Deep Learning

Deep learning is a subset of machine learning that focuses on training artificial neural networks with multiple layers to learn and extract complex representations from data. It has emerged as a powerful approach in the field of artificial intelligence and has revolutionized various domains such as computer vision, natural language processing, and speech recognition [18].

At its core, deep learning involves constructing artificial neural networks that mimic the structure and functioning of the human brain. These networks consist of interconnected layers of artificial neurons, also known as nodes or units. Each neuron receives input signals, performs computations, and passes the results to the next layer of neurons, ultimately producing an output[13].

Deep learning has achieved remarkable results in various applications. It has surpassed human-level performance in image classification, object detection, and segmentation tasks. In natural language processing, deep learning models have excelled in tasks like sentiment analysis, machine translation, and language generation. Moreover, deep learning has made significant contributions to various technological advancements such as autonomous driving, medical diagnostics, and drug discovery[13].

1.3 Solution

The existing methods for predicting drought categories in Western America often rely on historical climatic data and statistical models, which have inherent limitations due to their inability to capture complex non-linear relationships and dependencies within the data. Consequently, there is a growing need for more sophisticated and advanced techniques that leverage the power of machine learning algorithms to enhance drought prediction accuracy and efficiency.

Machine learning has emerged as a powerful tool in various scientific disciplines, including meteorology, hydrology, and climate science. By utilizing advanced computational techniques, machine learning models can analyze large volumes of multi-dimensional data, recognize hidden patterns, and generate accurate predictions. Consequently, applying machine learning to drought prediction has the potential to provide invaluable insights for effective water resource management and decision-making processes.[16]

The specific focus of this thesis is to develop a machine-learning model that predicts drought categories for a 1-year time frame in Western America. By harnessing the vast amount of available data, including climate variables, soil moisture levels, snowpack, and other relevant spatial and temporal features, we aim to construct a robust predictive model that can classify regions into distinct drought categories. These categories can include abnormally dry, moderate drought, severe drought, or no drought, providing decision-makers with actionable information for effective resource allocation and planning.

Chapter 2

Project Requirements

2.1 Technologies Used

Python is chosen as the primary language for this project due to its extensive range of data science and machine learning libraries. These libraries provide rich functionality for various tasks, enabling efficient development and implementation. Additionally, Python boasts a wide community support network, ensuring access to a wealth of resources, documentation, and assistance from experts.[3]

2.1.1 Python Libraries

The following libraries are used

- Pandas V1.44, used for reading and editing csv files.
- Xarray V2023.03.0, used to transform data into xarray's for splicing and exploration.
- netCDF4 V1.6.4, used to interact with the CDF file format and convert it into explorable data.
- XGBoost V1.7.5, used to train the data set with the XGboost model.
- Shapely V2.0.1, used when working with geopolygons, specifically when assigning a drought category to a coordinate.
- PyTorch V2.0, used to train the data set with the deep learning model.

2.1.2 Additional Technologies

- Google Colab will be used to run our Python notebooks in the background.
- Kepler will be used for data visualization.

2.2 Data

2.2.1 NCEP

The NCEP North American Regional Reanalysis (NARR) dataset, developed by the National Centers for Environmental Prediction (NCEP), is a comprehensive and high-resolution climate data set that offers a detailed perspective of atmospheric conditions across North America. Combining information from diverse sources including satellite observations, ground-based measurements, and numerical weather models, the NARR data set creates a cohesive and uninterrupted representation of the atmosphere over the region. With a spatial resolution of 32 kilometers, it encompasses a wide array of variables, including temperature, wind speed, precipitation, and atmospheric pressure, enabling a thorough analysis of the climate system. Spanning the period from 1979 to 2010, the NCEP data set provides an invaluable resource for researchers seeking to delve into the intricacies of North America's climate dynamics[12].From the NCEP datasets, we utilized data on the following features: average precipitation [4], surface air temperature [8], snowpack [9], total cloud coverage [11], plant canopy surface water [7], convective precipitation accumulation [5], moisture availability [6], and subsurface runoff (baseflow) [10].

2.2.2 USDM

The U.S. Drought Monitor is a weekly released map that provides an overview of drought conditions in different regions of the United States. The map employs five distinct classifications to assess the severity of drought: abnormally dry (D0), indicating areas that may be transitioning into or recovering from drought, as well as four progressive levels of drought: moderate (D1), severe (D2), extreme (D3), and exceptional (D4).

While the U.S. Drought Monitor does not rely on a statistical model, it incorporates numerical manual inputs from hundreds of scientists to inform its assessments. These inputs encompass various climatological factors, such as the Palmer Drought Severity Index, the Standardized Precipitation Index, and other relevant data sources. Additionally, the map incorporates data like the Keech-Byram Drought Index for fire risk, satellite-based evaluations of vegetation health, indicators of soil moisture, and hydrological data, particularly in the Western region, including the Surface Water Supply Index and snow pack measurements.[1]

2.3 Accuracy Requirements

We established the minimum threshold for success in this project to be an accuracy score exceeding 50 per cent. However, considering the timeline and project scope, our target was to achieve an F1 score and accuracy score of at least 70 per cent.

Chapter 3

Development Timeline and Risks

3.1 Timeline

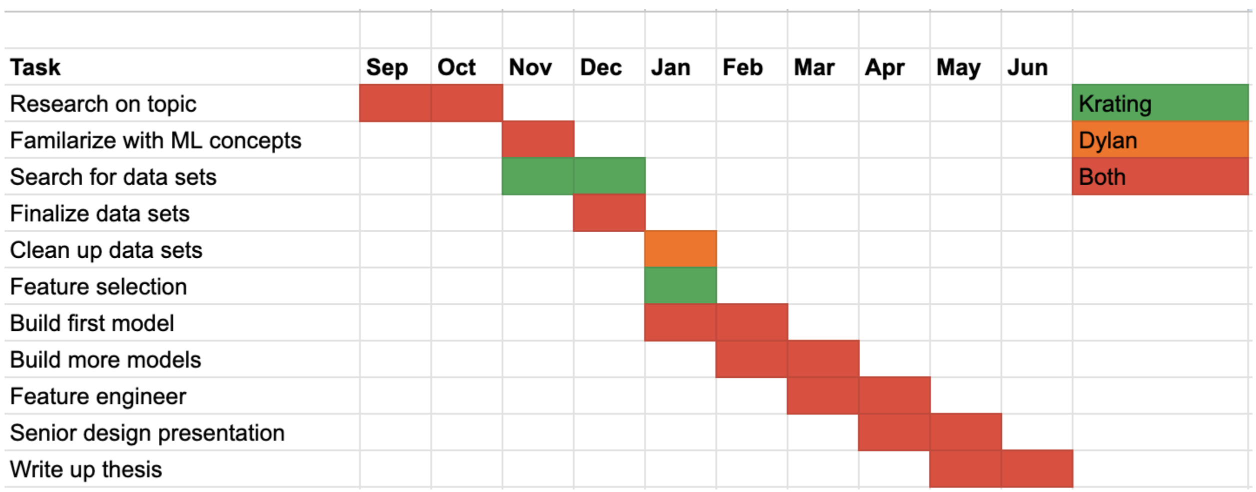

Figure 3.1: Gantt chart expected timeline with roles

Our timeline can be split up into the following four phases, Ideation, Exploration, Implementation and Optimization

1. **Phase 1: Ideation**

   - ***Time:*** Fall Quarter
   - ***Goals:*** Research climate issues and figure out which problem to approach. Learn machine learning and data science concepts.Explore previous works and existing data sets.

2. **Phase 2: Exploration**

   - ***Time:*** Fall + Winter Quarter
   - ***Goal:*** Finalize which data sets to use, explore and clean up data sets to generate a final usable data set.

3. **Phase 3: Implementation**
   - ***Time:*** Winter Quarter
   - ***Goal:*** Implement different machine learning models on our data set.
4. **Phase 4: Optimization**
   - ***Time:*** Spring Quarter
   - ***Goal:*** Optimize models through model modifications as well as feature engineering to generate the highest accuracy score within our scope of time.

3.2 Possible Risks Analysis

1. Missing Values in Data: Individual data sets may have missing values, resulting in an incomplete final data set where some rows may lack values. This can pose challenges during data analysis and modeling, requiring appropriate handling techniques such as data imputation or exclusion of incomplete sets.
2. Our team's limited experience with machine learning concepts may lead to underestimating the project's scope and time requirements. Tasks such as data exploration and model implementation could potentially take longer than anticipated, potentially hindering progress in subsequent project phases.
3. Time Constraints for Data Generation and Model Training: The process of generating the data set itself and training the machine learning model may be time-consuming. This extended time frame can limit the number of experiments or iterations that can be performed, potentially impacting the thoroughness of analysis and optimization efforts.
4. Potential Inaccuracy of the Model: There is a possibility that the final model may not achieve the desired level of accuracy. Machine learning involves iterative experimentation, and it may require multiple iterations and fine-tuning to reach the desired outcome.

3.3 Problems and Obstacles

Initially, we faced challenges in working with the netCDF data provided by NCEP. Understanding how to effectively utilize this format was a significant hurdle that required a lot more time than anticipated.

Another key issue we encountered was related to generating drought categories for our NCEP data set, particularly for the years 2020 and 2021. We encountered JSON encoding errors during this process, which would have necessitated manual intervention to add the drought categories for the affected values in order to ensure accurate analysis and predictions. We decided to forgo data for these 2 years.

Chapter 4

Implementation

4.1 Data Exploration

4.1.1 NCEP

To make the data from NCEP usable, a series of steps were undertaken. Firstly, the data for a specific atmospheric variable within a defined range of years (e.g., 2010-2019) was downloaded from NCEP. However, the data was provided in a CDF format, which posed challenges for exploration and analysis. In order to address this, the netcdf4 python library was utilized to convert the CDF data into a more accessible numpy array.

Once the data was in a numpy array, functions were developed to filter and limit the data set. For our initial scope, we decided to only include only coordinates within the state of California. Initially, a rectangular approach(see figure 4.1) was employed using four coordinate limits, but this included oceanic areas without any drought category information, thereby impacting the accuracy of the model. Consequently, a revised approach was adopted, involving the creation of a geopolygon representing the state of California(see figure 4.2). The shapely Python library was employed to select only the coordinates falling within the geopolygon, ensuring a more accurate representation of California's drought data.

Subsequently, Pandas was employed to eliminate unnecessary column values, such as Lambert conformal and time bounds, streamlining the data set for further analysis. These steps were then repeated for each NCEP variable of interest, allowing for the integration of multiple variables into the analysis.

By following this iterative process, the data from NCEP was transformed into a usable format, refined to include relevant coordinates within California, and consolidated with additional variables. These steps paved the way for subsequent analyses and the development of our data set.

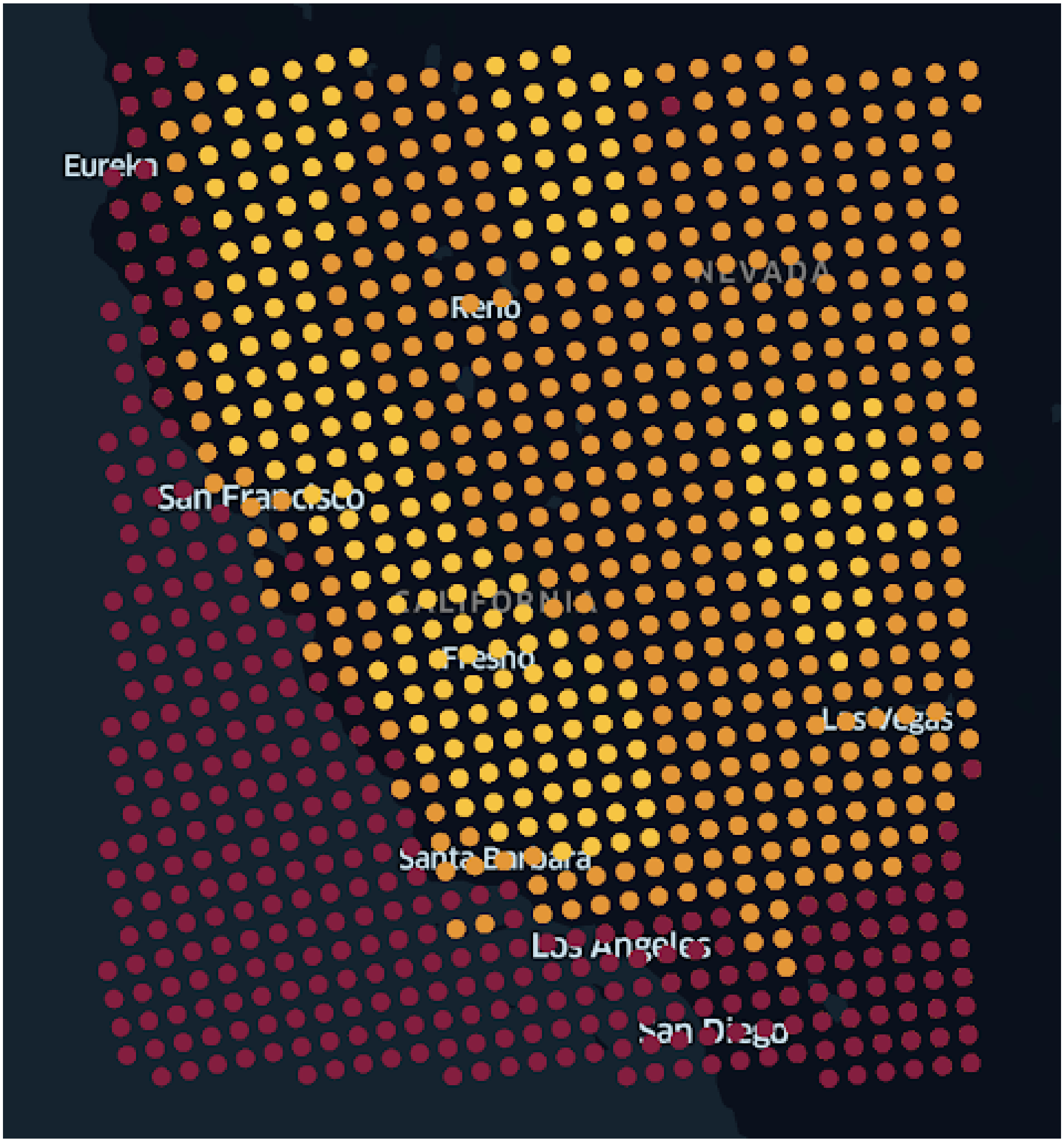

Figure 4.1: A visualization of data limited using rectangular coordinates

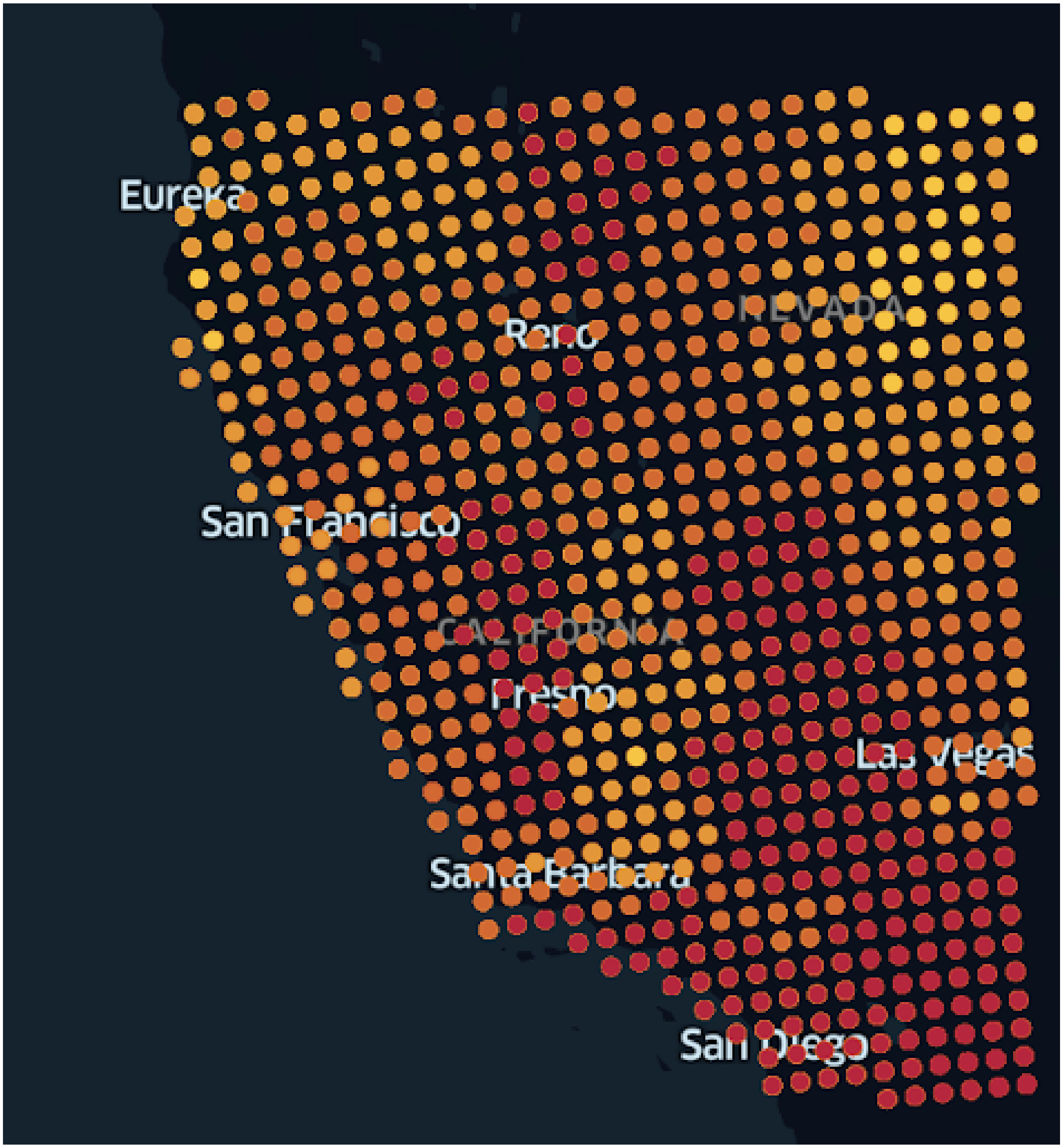

Figure 4.2: A visualization of data limited using a geopolygon

4.1.2 USDM

In the process of combining the NCEP data with the USDM data, several steps were taken to ensure accurate integration. Unlike the NCEP data, which has a data set for each year, the USDM data is updated weekly and required manual downloading of each file for the corresponding year, typically comprising 52-54 files. To streamline this process, a custom Python web scraper was developed to automatically download the USDM files for the desired year.

To align the NCEP data with the USDM data, a helper function was created to determine the starting and final dates for the NCEP dataset in a given year. This enabled assigning the nearest available USDM data set to the dates in the NCEP set that did not fall within a pair of USDM datasets. For instance, if the NCEP date was before the start date, it would use the drought category from the corresponding start date set, and vice versa if the date was after the end date. All other dates would use the closest corresponding date file that came after the given date. For example, the date range 01/05/2010-01/11/2010 would utilize data from the 01/12/2010 geojson file.

Matching a coordinate with a geojson file involved utilizing a helper function. Firstly, the appropriate geojson file was determined. Then, using the shapely library, the coordinate was matched with a geojson shape, returning the associated drought category for that specific geographic area. If the coordinate did not fall within any shape, a 0 value was assigned, indicating that the area was not experiencing drought conditions.

This process was repeated for each yearly data set to generate a comprehensive 10-year training data set and a 1-year data set, integrating the NCEP and USDM data effectively. These steps ensured the accurate alignment of data sources and the generation of a cohesive data set for subsequent analysis and machine-learning modelling.

4.2 Models

4.2.1 Deep learning model

The first approach to tackling our classification problem was to use deep neural networks. Deep Learning models have become a popular approach in machine learning as they are capable of tackling complex problems and are highly customizable for specific problems.

For our deep learning approach, the main architecture we chose was a simple multi layer perceptron neural network. This model consists of three parts which are the input layer, output layer and hidden layer. For the input layer, the amount of nodes is set to the corresponding amount of inputs, while the output layer has six nodes for each drought category. The hidden layer is where we made changes to the number of layers and nodes per layer to try to reach optimal results.

In terms of training methods, we incorporated different techniques. Our first approach was to use classical gradient descent but we improved it by using Mini-Batching. Mini-Batching is where the training data set is split into batches of a given size. The main advantages of using the Mini-Batching training method is computational efficiency and stable error gradient and convergence. Dropout layers were also used to reduce overfitting and it helped increase the test accuracy as well.

4.2.2 XGboost model

Due to the extended training time of our Deep Learning model, a decision was made to explore alternative approaches. Consequently, we turned our attention to XGBoost, an extremely efficient and scalable machine learning library, to address this challenge. XGBoost's reputation for its speed and scalability made it an ideal choice for our project, allowing us to overcome the training time limitations of the Deep Learning model.

XGBoost, short for Extreme Gradient Boosting, is a highly scalable and distributed gradient-boosted decision tree (GBDT) machine learning library. It excels in tackling regression, classification, and ranking problems, making it a versatile algorithm for supervised learning tasks.[2]

The foundation of XGBoost lies in gradient boosting, a technique that combines weak learners, typically decision trees, to create a strong learner. By iteratively adding trees to the ensemble, XGBoost gradually improves its predictive performance.[2]

One of the key strengths of XGBoost is its incorporation of advanced features that enhance its effectiveness. These include regularization techniques, which mitigate overfitting and improve generalization. Additionally, XGBoost leverages parallel processing capabilities, allowing for efficient computation across multiple cores or distributed systems. Another advantage is its ability to handle missing values, ensuring robustness in real-world data sets.[2]

Notably, XGBoost is renowned for its exceptional speed and scalability. It can efficiently handle large data sets which is what is needed to perform feature engineering with our data set.

Chapter 5

Model Results

5.1 Deep learning model results

The Multi layer Perceptron Model was experimented with mainly by testing the number of nodes in the hidden layers.

Experiment	Description	Accuracy
Multi layer Perceptron Model (MLP)	A simple multi layer perceptron model. This model had three layers in the hidden layer with the middle layer containing the most nodes.	0.487
Multi layer Perceptron Model More Complex Hidden Layer (MLP2)	To improve upon the basic model, we experimented with directly adding more pytorch linear layers in the hidden layer so that the model can fit the data better.	0.6205
Multi layer Perceptron Model with Dropout Layers	Instead of just having pytorch linear layers, we added dropout layers between each linear layer.The dropout rate was set to 0.2.	0.667

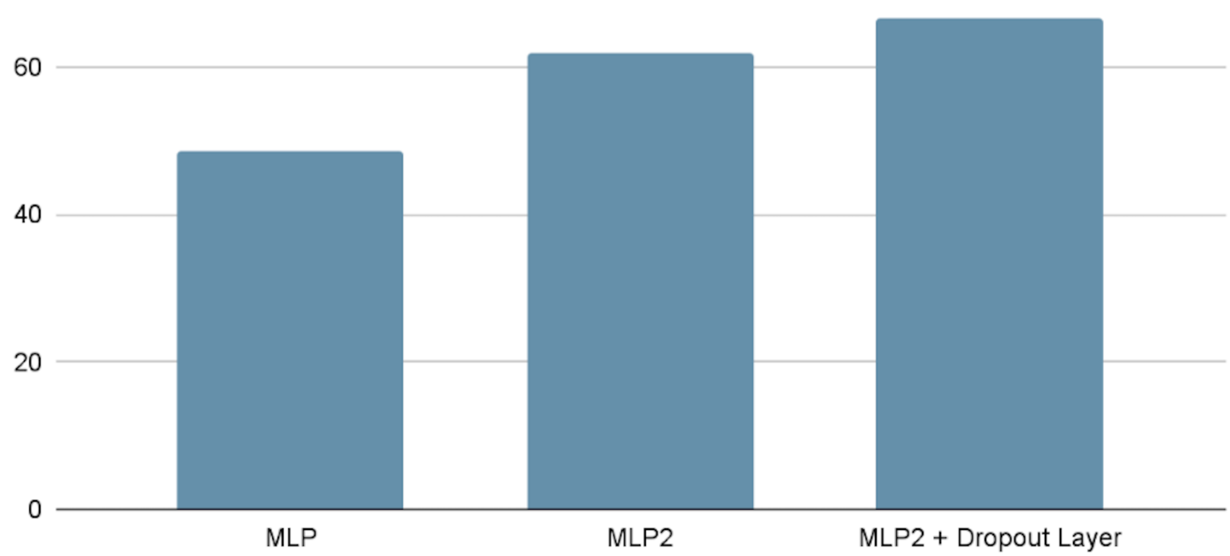

Figure 5.1: Deep learning model results

5.2 XGBoost results

We conducted four experiments using the XGBoost model, all of which were evaluated on our 2022 data set. The performance of each experiment was measured using Accuracy and F1 score as metrics.

Experiment	Description	F1 Score	Accuracy
Data set 1	This experiment utilized a 5-year data set ranging from 2015 to 2019, incorporating climate variables such as average precipitation, surface air temperature, snowpack, and total cloud coverage.	0.61	0.67
Data set 2	Data set 2 employed the same features as Data set 1 but expanded the data range to 10 years, covering the period from 2010 to 2019. With this extended time frame, the model demonstrated improved performance.	0.65	0.73
Data set 3	Dataset 3 encompassed a 10-year data range, similar to Data set 2, but introduced additional features of snow melt, plant canopy surface water and convective precipitation accumulation.	0.69	0.74
Data set 4	Data set 4 incorporated the same features as Data set 3, but added the features of moisture availability and subsurface runoff(baseflow).	0.71	0.74

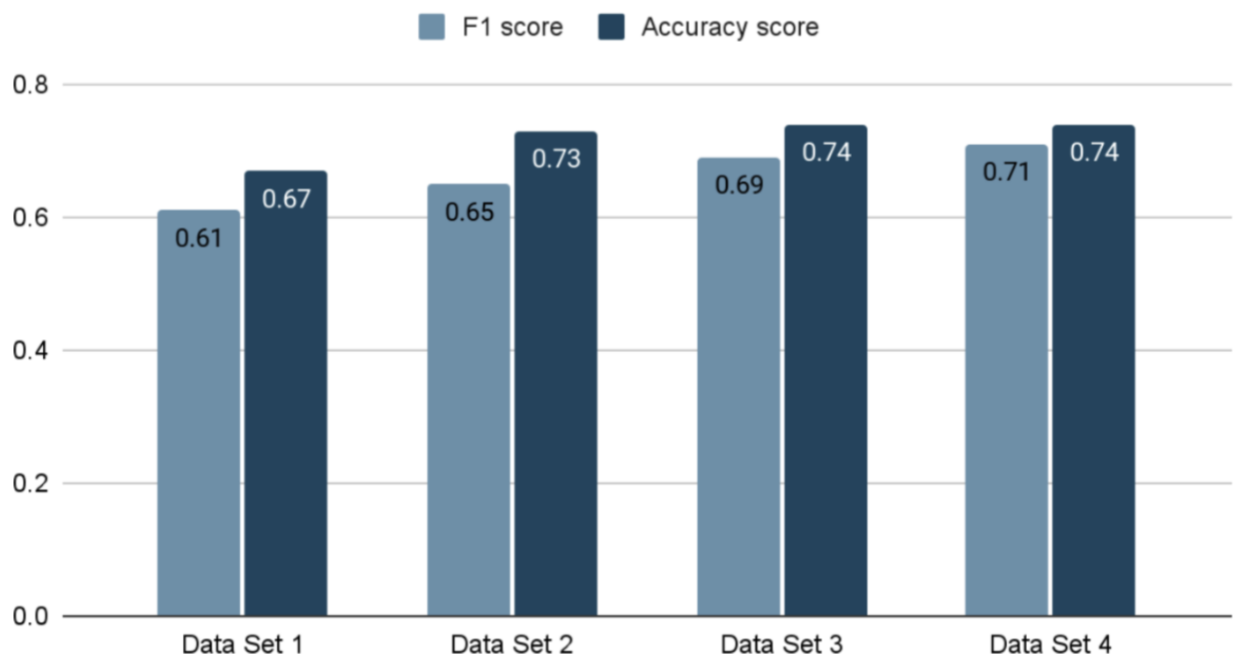

Figure 5.2: Example Image

Chapter 6

Social Ethics

6.1 Usability

Currently, our model relies on the manual downloading and processing of data. To enhance its usability and streamline the data preparation process for training, future improvements could involve the development of a user interface. Such an interface would automate and simplify the steps involved in data acquisition and preprocessing, ensuring a more seamless experience for users interested in training the model. By implementing a user-friendly interface, the overall accessibility and convenience of the model would be significantly enhanced.

6.2 Accessibility

Regarding the project's accessibility, the data sets utilized are accessible to the general public. However, processing these data sets can be time-consuming, particularly without a powerful computer. On the other hand, the training phase of the model can be completed relatively quickly on most modern personal computers. Although data set processing presents challenges, users with standard computing setups can still access and utilize the project efficiently by leveraging widely available hardware configurations for model training.

6.3 Economic

Implementing our project for professional use, particularly in the agricultural sector, introduces a critical consideration: even slight inaccuracies in the predictions could have severe repercussions for farmers during agricultural planning, resulting in wasted resources and increased costs. Therefore, it is crucial to strive for higher accuracy in our model. Achieving improved accuracy has the potential to not only enhance the effectiveness of our model but also disrupt traditional roles, such as those of climate scientists and observers responsible for manually creating the USDM. By developing a highly accurate model, we could potentially automate and streamline the process, reducing the need for manual intervention. While this presents opportunities for efficiency gains, it also raises ethical concerns regarding potential job displacement.

Chapter 7

Conclusion

7.1 Summary

In conclusion, our thesis successfully developed an XGBoost model for predicting drought categories in Western America. Achieving an accuracy score of 0.74 which was higher than the score we set out to achieve. This model provides valuable insights for informed decision-making in water resource management, agriculture, and land-use planning. Our research highlights the potential of machine learning in addressing the challenges of climate change and enhancing drought category prediction accuracy.

7.2 Next steps

To further improve our drought prediction model, several enhancements are planned. Firstly, we will integrate additional data sets to refine the model's performance. This will involve incorporating a broader range of climate variables and expanding the geographical coverage beyond California to encompass the entire Western America region.

Moreover, we aim to develop an interactive user interface that offers a seamless visualization of our model's results. Users will be able to input specific time frames and coordinates of interest, receiving predicted drought categories with corresponding accuracy percentages. This UI will provide a user-friendly platform to explore and interpret the model's outputs effectively.

Continuous model evaluation and improvement will be a priority, ensuring the model remains up-to-date and reliable over time. This process will involve periodic updates as new data becomes available and the incorporation of user feedback to enhance the model's predictive capabilities.

Bibliography

[1] What is the USDM? Webpage, 2023. URL https://droughtmonitor.unl.edu/About/WhatistheUSDM.aspx.

[2] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322. doi: 10.1145/2939672.2939785. URL https://doi.org/10.1145/2939672.2939785.

[3] K. Doshi. Leveraging Geolocation Data for Machine Learning: Essential Techniques. Webpage, 2021. URL https://towardsdatascience.com/leveraging-geolocationdata-for-machine-learning-essential techniques-92ce3a969bc.

[4] N. C. for Environmental Prediction (NCEP). Precipitation amount. Data set, 2023. URL https://psl.noaa.gov/thredds/catalog/Datasets/NARR/monolevel /catalog.html. Retrieved from National Centers for Environmental Prediction (NCEP) website.

[5] N. C. for Environmental Prediction (NCEP). Convective precipitation accumulation. Data set, 2023. URL https://psl.noaa.gov/thredds/catalog/Datasets /NARR/monolevel/catalog.html. Retrieved from National Centers for Environmental Prediction (NCEP) website.

[6] N. C. for Environmental Prediction (NCEP). Moisture availability. Data set, 2023. URL https://psl.noaa.gov/thredds/catalog/Datasets/NARR/monolevel /catalog.html. Retrieved from National Centers for Environmental Prediction (NCEP) website.

[7] N. C. for Environmental Prediction (NCEP). Plant canopy surface water. Data set, 2023. URL https://psl.noaa.gov/thredds/catalog/Datasets/NARR /monolevel/catalog.html. Retrieved from National Centers for Environmental Prediction (NCEP) website.

[8] N. C. for Environmental Prediction (NCEP). Air temperature. Data set, 2023. URL https://psl.noaa.gov/thredds/catalog/Datasets/NARR/monolevel /catalog.html. Retrieved from National Centers for Environmental Prediction (NCEP) website.

[9] N. C. for Environmental Prediction (NCEP). Snow pack. Data set, 2023. URL https://psl.noaa.gov/thredds/catalog/Datasets/NARR/monolevel /catalog.html. Retrieved from National Centers for Environmental Prediction (NCEP) website.

[10] N. C. for Environmental Prediction (NCEP). Subsurface runoff. Data set, 2023. URL https://psl.noaa.gov/thredds/catalog/Datasets/NARR/monolevel /catalog.html. Retrieved from National Centers for Environmental Prediction (NCEP) website.

[11] N. C. for Environmental Prediction (NCEP). Total cloud cover. Data set, 2023. URL https://psl.noaa.gov/thredds/catalog/Datasets/NARR/monolevel /catalog.html. Retrieved from National Centers for Environmental Prediction (NCEP) website.

[12] N. C. for Environmental Prediction (NCEP). NCEP North American Regional Reanalysis (NARR). Dataset, 2023. URL https://psl.noaa.gov/data/gridded/data.narr.html.

[13] IBM. Deep learning, 2021. URL https://www.ibm.com/topics/deep-learning. IBM Website.

[14] E. Kavlakoglu. Ai vs. machine learning vs. deep learning vs. neural networks: What's the difference? https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks, 2020. IBM Cloud blog.

[15] A. K. Mishra and V. P. Singh. A review of drought concepts. *Journal of Hydrology*, 391 (1):202–216, 2010. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2010.07.012. URL https://www.sciencedirect.com/science/article/pii/S0022169410004257.

[16] J. Passioura. Drought and drought tolerance. *Plant Growth Regul*, 20:79–83, 1996. doi: 10.1007/BF00024003.

[17] R. Pugliese, S. Regondi, and R. Marini. Machine learning-based approach: global trends, research directions, and regulatory standpoints. *Data Science and Management*, 4: 19–29, 2021. ISSN 2666-7649. doi: https://doi.org/10.1016/j.dsm.2021.12.002. URL https://www.sciencedirect.com/science/article/pii/S2666764921000485.

[18] N. Rusk. Deep learning. *Nature Methods*, 13(1):35–35, Jan 2016. ISSN 1548-7105. doi: 10.1038/nmeth.3707. URL https://doi.org/10.1038/nmeth.3707.

[19] I. SALIAN. Supervised and unsupervised learning explained, 2018. URL https://blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/. NVIDIA Blog.

[20] UC Berkeley School of Information. What is machine learning? https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/, 2022. UC Berkeley School of Information blog - iSchool Online.

[21] X. Wang, P. Luo, Y. Zheng, W. Duan, S. Wang, W. Zhu, Y. Zhang, and D. Nover. Drought disasters in china from 1991 to 2018: Analysis of spatiotemporal trends and characteristics. *Remote Sensing*, 15(6), 2023. ISSN 2072-4292. doi: 10.3390/rs15061708. URL https://www.mdpi.com/2072-4292/15/6/1708.