https://siddarthvarma2000k.wixsite.com/my-site-8 - website url
https://github.com/siddarth2959/Wildfire-Watch - github url

Individual Contribution Report - Moksha Sai

1. Introduction

This report summarizes the responsibilities carried out by Person 1 during Milestone 3 of the Wildfire Watch project. My primary task package centered on the implementation and documentation of the **Frequent Pattern Mining (FPM)** component, which included data discretization, Apriori algorithm execution, association rule analysis, interpretation of results, and preparation of the corresponding project website section. This work established the descriptive analytical foundation for understanding repeated climate conditions that precede wildfire events.

2. Data Cleaning and Discretization

My contributions began with validating the cleanliness of the dataset and preparing it for frequent pattern mining. Because the Apriori algorithm requires categorical rather than continuous features, one of my core responsibilities was to transform numerical climate variables—such as temperature, humidity, precipitation, and wind speed—into meaningful categorical bins. I designed discretization thresholds that aligned with meteorologically relevant cutoffs (e.g., high vs. low humidity, elevated vs. moderate temperature, dry vs. wet precipitation levels). This conversion ensured that environmental characteristics could function as discrete transaction-like items suitable for the Apriori model. Throughout this process, I inspected distributions, verified bin quality, and ensured that no critical climate information was lost during categorization.

3. Apriori Model Implementation

Once the data was successfully transformed, I developed and executed the Apriori algorithm to generate frequent itemsets based on minimum support thresholds. This involved tuning support values to ensure that only meaningful and non-trivial item combinations were considered. Following the identification of frequent itemsets, I applied association rule mining to generate metrics such as **support**, **confidence**, and **lift**, which helped quantify the strength and significance of climate patterns associated with wildfire activity. I ensured that both the itemset extraction and rule generation processes followed optimal methodology and maintained computational efficiency.

4. Analysis and Interpretation of Association Rules

After generating the association rules, I carried out a detailed analysis to interpret their implications. This analysis revealed several important climate combinations that consistently preceded wildfire events, such as high temperature paired with low humidity or low precipitation. I examined lift values to determine the relative strength of each rule and prioritized patterns that showed strong predictive relationships. These interpretations were essential for establishing the descriptive foundation of the project, allowing later modeling stages (classification, regression, and clustering) to be contextualized within real climate behaviors captured in the dataset.

5. Visualization and Output Preparation

To strengthen clarity and interpretability, I prepared visualizations that highlighted the relationships uncovered by the Apriori model. These included bar charts of the top frequent itemsets and scatter plots illustrating the relationship between confidence and lift. These figures were designed to be used in both the technical report and the project website. I ensured that all graphs clearly communicated the core insights while maintaining consistent formatting and academic quality.

6. Website Section Creation

I authored the complete **"Frequent Pattern Mining"** section for the project website. This included model overview, rationale for discretization choices, methodological explanations, interpretation of association rules, and all supporting visualizations. The section was structured to be accessible for general readers without compromising the academic rigor of the project

7. Summary of Contributions

I(Moksha Sai) contributed all work related to data discretization, Apriori frequent pattern mining, association rule evaluation, and interpretation. I also developed the FPM visualizations and prepared the website section for this model. These contributions established the descriptive analytical basis for understanding recurring climate conditions associated with wildfire emergence.

Individual Contribution Report - **Abhinaya Allu**

1.Introduction
This individual contribution report documents the complete set of responsibilities completed by Person 2 as part of Milestone 3 of the Wildfire Watch project. The focus of my assigned tasks was the full development, evaluation, interpretation, and documentation of the **K-Means clustering component**, including preprocessing validation, model selection, metric evaluation, visualization creation, and integration of this analysis into the project website. My work formed the unsupervised learning portion of the project and contributed to understanding naturally occurring climate–fire risk patterns in the dataset.

2.Data Preparation and Scaling Validation
Before implementing the K-Means clustering model, I conducted an assessment of the preprocessed dataset provided by the team. My responsibility was to validate that the features used for clustering were appropriately standardized. This step was crucial because K-Means relies on Euclidean distance, and features measured on different scales would distort cluster formation. I verified that numerical climate variables—such as temperature, humidity, wind speed, and precipitation—had been transformed using standard scaling. This ensured that each feature contributed proportionally to distance calculations and prevented model bias toward any attribute with a naturally larger range. Through summary statistics and visual inspection, I confirmed that the scaled dataset satisfied the assumptions required for unsupervised clustering.

3. K-Means Model Implementation and Evaluation
My primary contribution involved designing and executing the K-Means clustering experiments for a range of cluster counts, specifically **k = 2, 3, 4, and 5**. For each value of k, I trained the K-Means algorithm on the scaled climate variables and evaluated cluster quality using two internal validation metrics: the **Silhouette Score** and the **Davies–Bouldin Index**. These metrics measure the compactness and separation of clusters, ensuring that the groupings are meaningful rather than arbitrary.

The results showed notable variation across cluster counts. The Silhouette Score was highest for **k = 3**, indicating optimal balance between cohesion within clusters and separation between clusters. Similarly, the Davies–Bouldin Index reached its lowest value at **k = 3**, further reinforcing that this configuration produced the most distinct and interpretable cluster structure. Based on this evidence, I recommended selecting **k = 3** as the final clustering solution.

4. Cluster Interpretation and Profile Development
Once the optimal model was identified, I analyzed the resulting cluster assignments to characterize each group. By computing mean climate values and comparing relative fire frequencies, I derived meaningful profiles for all three clusters. The first cluster represented **cool and humid regions**, which displayed the lowest incidence of wildfires. The second cluster corresponded to **hot and dry conditions**, exhibiting significantly higher wildfire risk. The third cluster represented a **transitional climate zone**, showing moderate fire likelihood. These interpretations helped translate numerical clustering results into real-world, climate-based risk categories that strengthen the project's analytical narrative.

5. Visualization Development
To support the interpretability of the clustering results, I generated several key visualizations. These included a **silhouette score plot** across different k-values, illustrating why k = 3 was preferred, and a **Davies–Bouldin Index plot** demonstrating consistency between the two internal metrics. Additionally, I prepared a **PCA-based two-dimensional cluster visualization**, which allowed the team to observe cluster separation in reduced dimensionality. Each visualization was designed to improve the clarity and communicative strength of the K-Means analysis in both the report and final website.

6. Website Section Creation
I authored and uploaded the complete "Clustering Models" section for the project website. This included explanatory text, model descriptions, metric interpretations, and all relevant visualizations. I ensured that the website presentation aligned with the academic tone of the report while remaining accessible to general audiences. The section clearly explains why K-Means was chosen, how the optimal k value was determined, and what the resulting clusters suggest about wildfire-related climate patterns.

7. Summary of Contributions

I(Abhinaya Allu) contributed a comprehensive unsupervised learning component to Milestone 3, including scaling validation, K-Means model development, evaluation across multiple metrics, cluster interpretation, visualization construction, and website documentation. This work provided critical insights into how climate patterns naturally group into fire-risk categories and added significant analytical depth to the overall project.

Individual Contribution Report – Siddarth Varma

1. Introduction

This individual report outlines the work completed by Person 3 as part of the Milestone 3 submission. My primary responsibilities involved the **classification and regression components** of the project, including the implementation of the Support Vector Machine (SVM) classifier, development of the Ridge Regression model, hyperparameter tuning, performance evaluation, visualization of outputs, and preparation of the corresponding website section. These tasks provided the predictive modeling capability of the wildfire analysis pipeline.

2. SVM Classification Model Development

My first major task was constructing the Support Vector Machine (SVM) model to predict wildfire occurrence. I prepared the dataset for classification by confirming that all required features were scaled appropriately, given that SVMs are sensitive to feature magnitude. I implemented a grid search over the hyperparameters **C**, **gamma**, and **kernel type**, exploring both linear and RBF kernels. This extensive search allowed the model to identify the most effective decision boundary for separating fire and non-fire instances. Once the optimal hyperparameters were found, I trained the final SVM classifier and evaluated it using accuracy, precision, recall, F1-score, and ROC-AUC.

3. Classification Model Evaluation and Interpretation

After training the optimal SVM model, I analyzed its performance metrics in detail. The classifier demonstrated strong predictive power, with a high ROC-AUC indicating that it effectively distinguished between fire-prone and non-fire conditions across decision thresholds. I examined precision and recall to understand the trade-off between false positives and false negatives, concluding that the model was particularly effective at identifying fire events due to its strong recall. To aid interpretation, I produced a confusion matrix and ROC curve, and I prepared code for integrating both into the final report. These evaluations formed the core of the project's predictive classification component.

4. Ridge Regression Model Implementation

In addition to classification, I developed the Ridge Regression model to predict burned area during wildfire events. I validated the log-transformed target and ensured consistent feature scaling. Using a grid search over alpha values, I identified the optimal level of regularization that minimized error while preventing overfitting. After training, I evaluated the model using R^2, RMSE, and MSE, providing a quantitative assessment of how well climate features explained variability in fire magnitude. This regression model highlighted the linear relationships between climate variables and fire intensity, while also revealing limitations in predicting extreme fires.

5. Visualization and Model Output Preparation

To support model interpretation, I developed multiple visualizations for both classification and regression. For the SVM classifier, I generated a confusion matrix and ROC curve, which provided intuitive summaries of model performance. For Ridge Regression, I created an actual vs. predicted scatter plot and a residual distribution plot, both of which helped assess model validity. These visualizations were prepared for inclusion in the report and website to effectively communicate predictive insights.

6. Project Website Section

I authored the **Classification and Regression Models** section of the project website. This section included explanations of model rationale, tuning procedures, metric interpretations, and visual outputs. I ensured that the content remained consistent with the academic tone of the report while maintaining accessibility for broader audiences.

7. Summary of Contributions

I(Siddarth Varma) contributed all classification and regression modeling work for the milestone. This included implementing, tuning, and evaluating the SVM and Ridge Regression models; generating all related visualizations; and preparing the associated website documentation. These efforts completed the core predictive capabilities of the wildfire analysis system and connected the project's descriptive insights to actionable forecasting.