Generation Drift Detection

Svanik Mareddy, Luke Troutman, Jada Reymann, Ria Nimbkar, Chakshu Grover,
Abha Gupta, Shahrzad Ghadiri, Eric Tang

Introduction

Interconnection delays increase costs, discourage investment, and weaken MISO's reliability. This project aims to accelerate energy deployment, reduce costs, improve grid planning, and target needed transmission investments by identifying delay drivers, analyzing patterns with clustering, and predicting COD changes using machine learning.

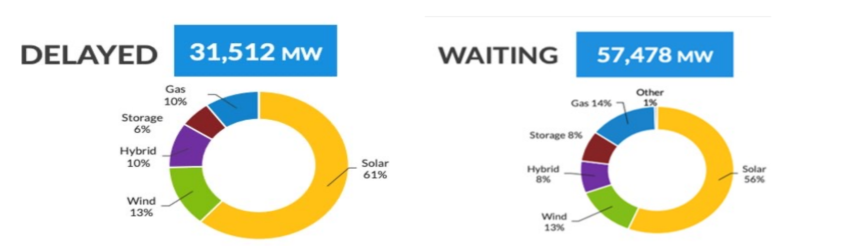
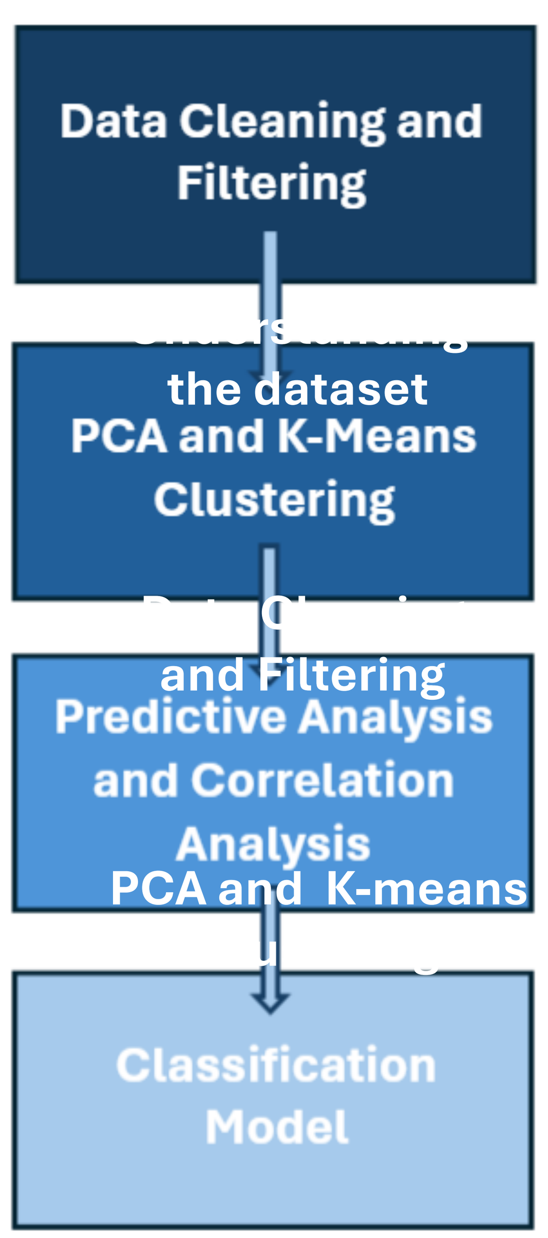


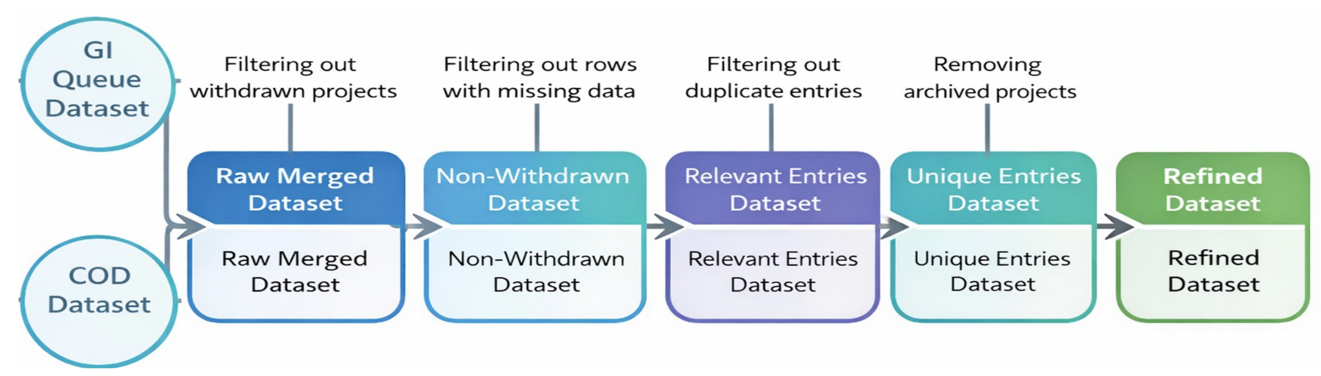
Figure 1: September 2025 – MISO Public Data Report

Methodology



To address these challenges, we applied a structured machine learning approach to organize our project data, using PCA and K-means clustering to group projects into four clusters. We then conducted predictive and correlation analysis and built a classification model to identify patterns and support future insights.

1 - Data Cleaning



Here we have a step by step break down of how we cleaned and filtered our dataset in order to provide accurate and effective information.

2 - Silhouette & PCA Graphs

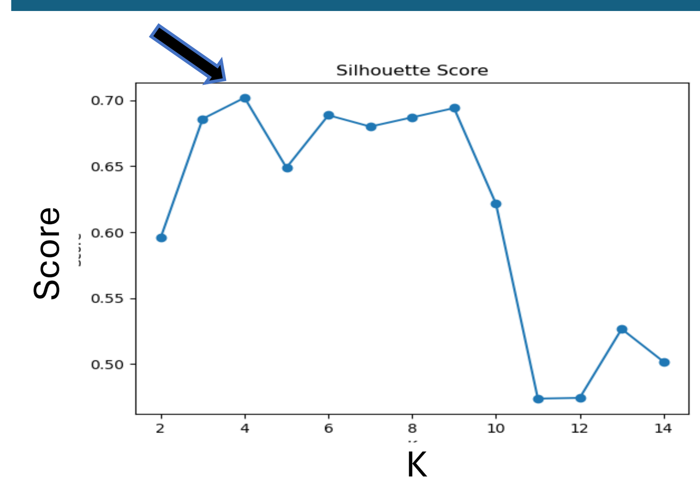


Figure 2: K Means Clustering Silhouette Graph

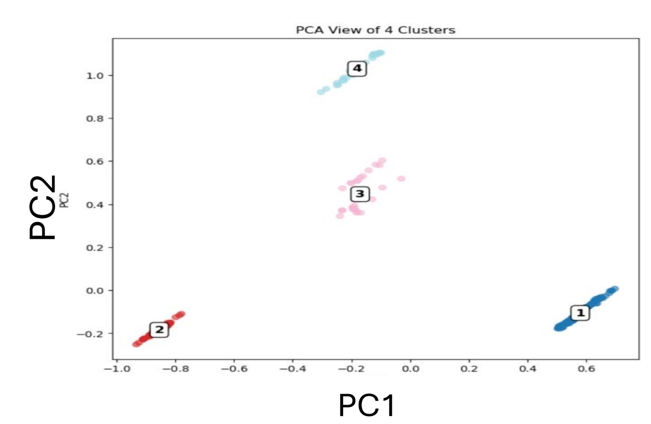


Figure 3: Principal Component Analysis of Clusters

Using our silhouette graph, we determined the optimal number of clusters to be four.

3 - Dominant Traits

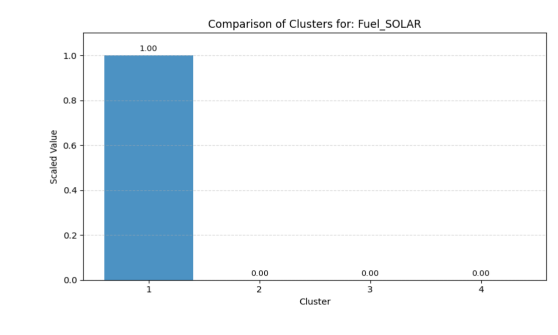


Figure 4: Cluster 1 Feature Sample

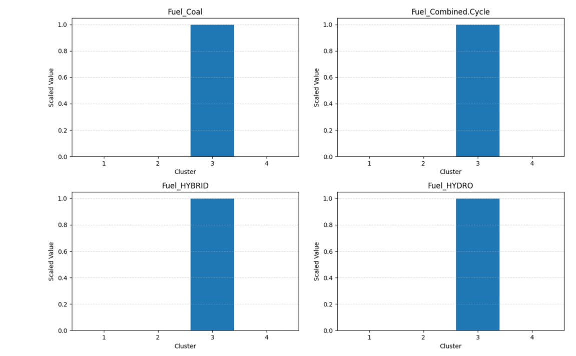


Figure 6: Cluster 3 Feature Sample

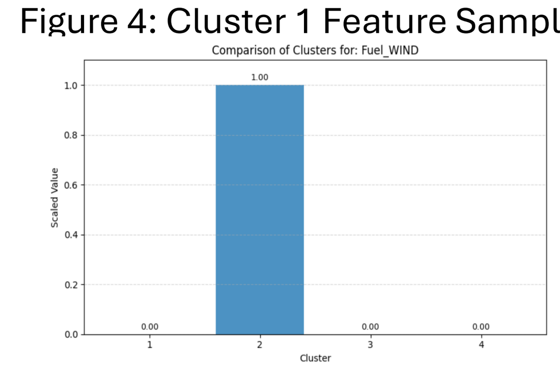
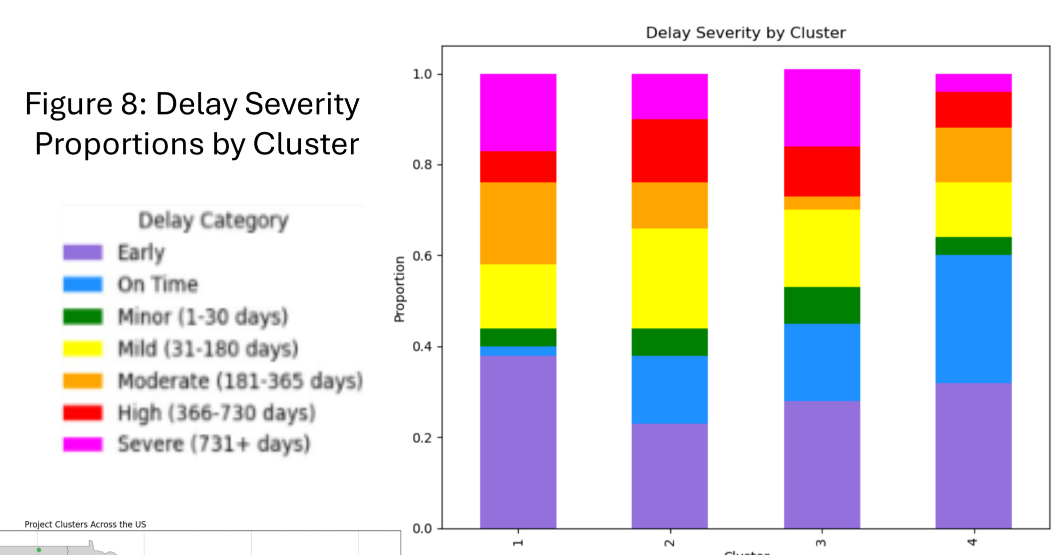


Figure 5: Cluster 2 Feature Sample

Figure 7: Cluster 4 Feature Sample

4 – Delay Type Frequency

Figure 8: Delay Severity Proportions by Cluster



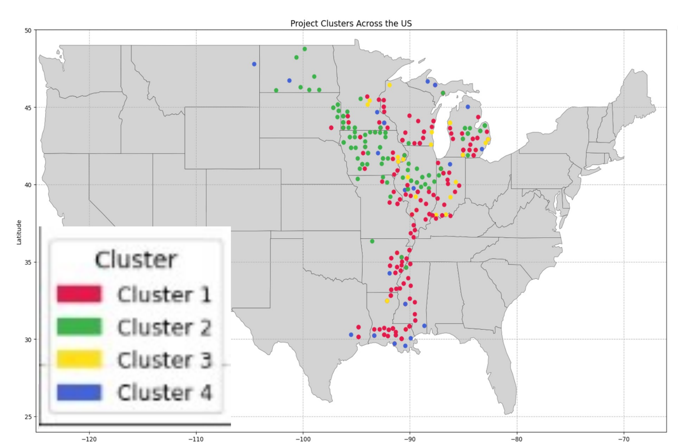


Figure 9: Cluster Geographic Distribution

5 – Classification Model

Here you can see a comparison between project cluster placement and the project placement of a cluster prediction model. There is an 88.6% overall agreement between the two methods. They use a combination of project features such as fuel type, location, and energy output to achieve this goal. The clustering of these projects allows us to better classify and predict the likelihood of each delay type that a project could experience.

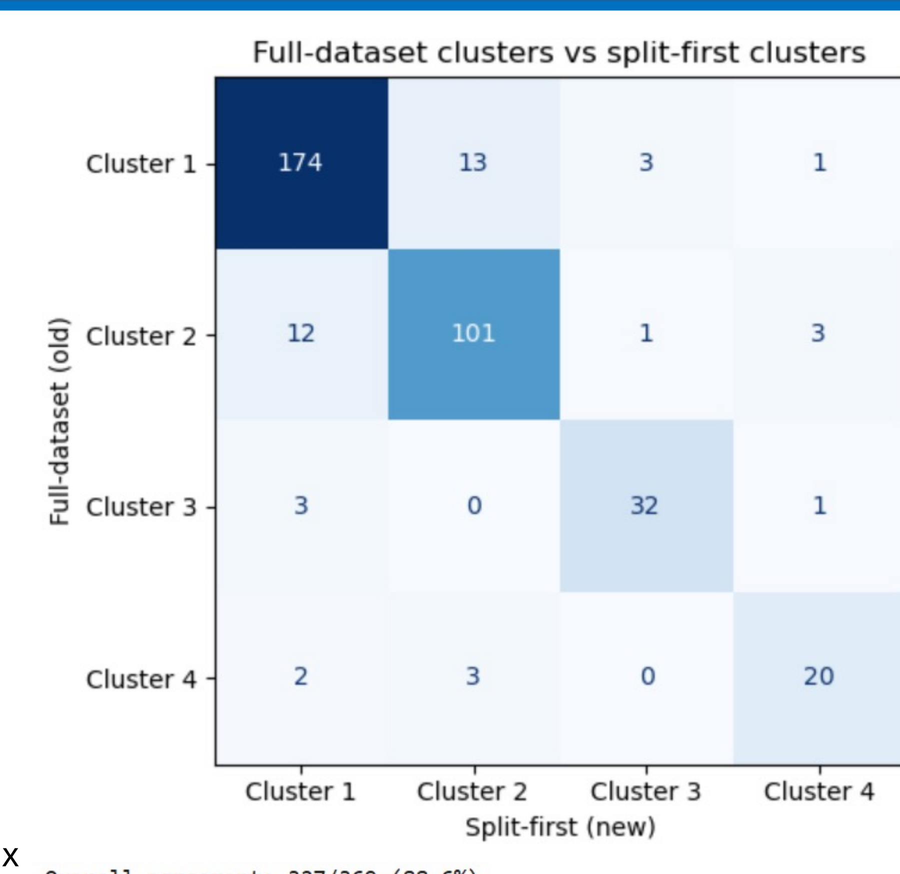


Figure 10: Confusion Matrix of Classification Model

Conclusion

In conclusion, we were able to build a classification model that is able to predict which of the 4 project clusters that project would fall into. These clusters represent varying amounts of delay time, from early completion to severe delay. We hope this work helps increase MISO's efficiency by determining which projects are at highest delay risk from the interconnection queue.

Future Work

In the future we hope to work on training a more specific predictive model that will be able to predict when a project is expected to be done and identify specific indicators that MISO can adjust in order to maximize efficiency and cost effectiveness.

Acknowledgements

Corporate Partner Mentor - Abigale Wynn
Advising Professor - Dr. Kwonsik Song
TA- Abha Gupta
Data Mine Staff - Maggie Betz and Bryce Castle
Fall Contributors - Paria Movahed and Xena Nguyen