

Clustering With EMNIST

Identifying Handwritten Letters With Basic Computers

Cameron Wood, 03/05/2024

[Github Repo](#)

Agenda

- Problem Statement
- Machine Learning Approaches
- Results
- Summary

Problem Statement

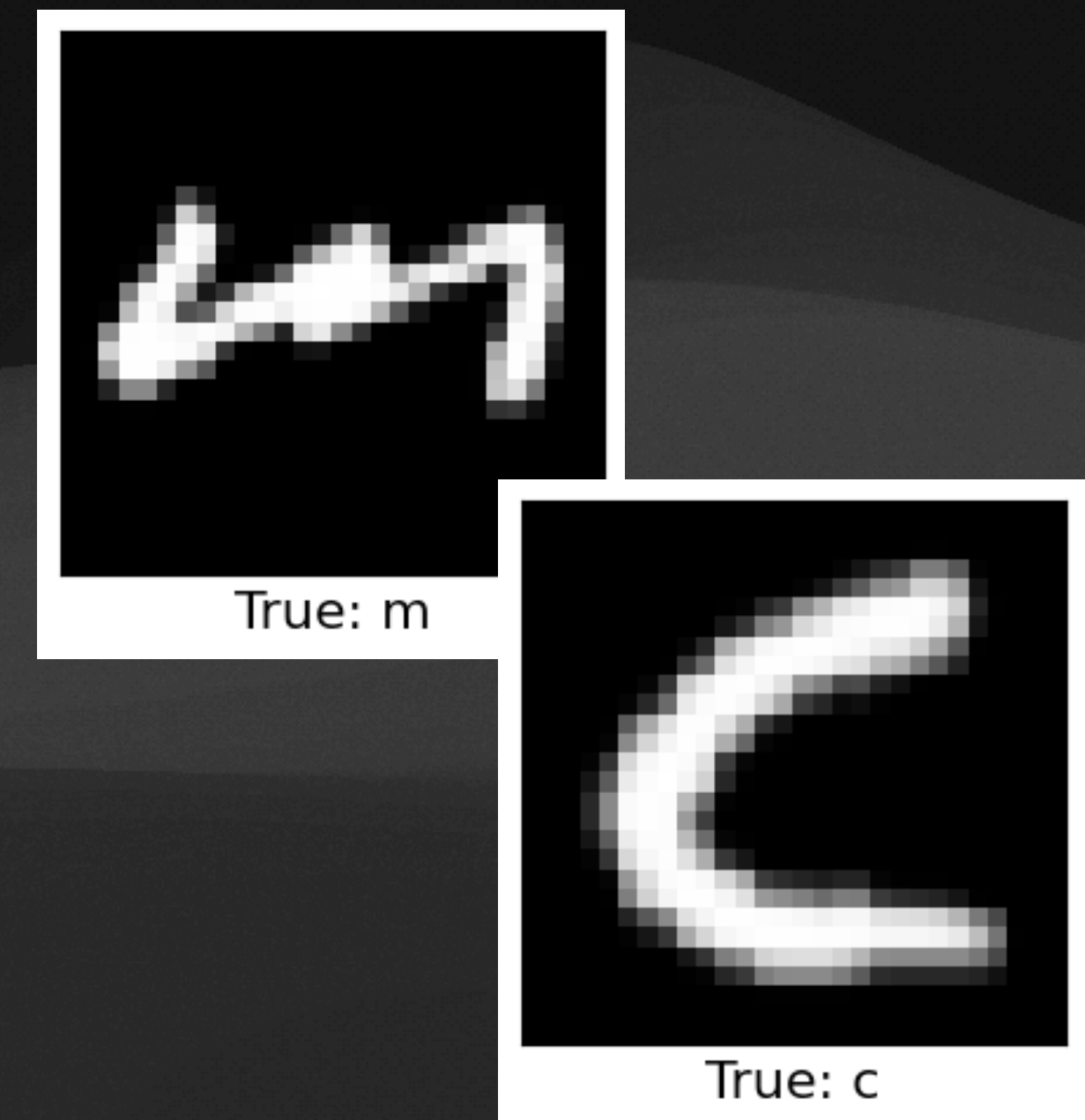
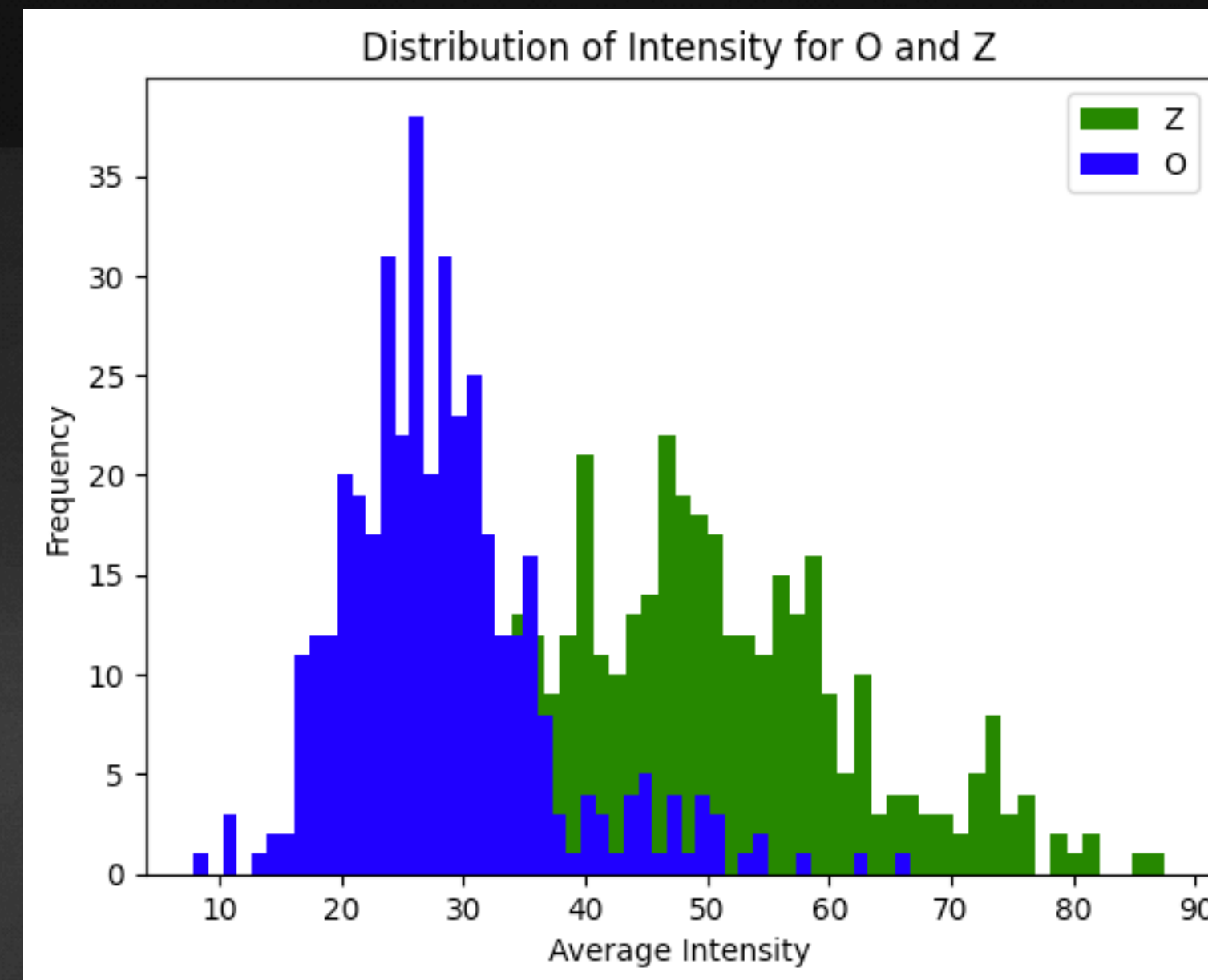
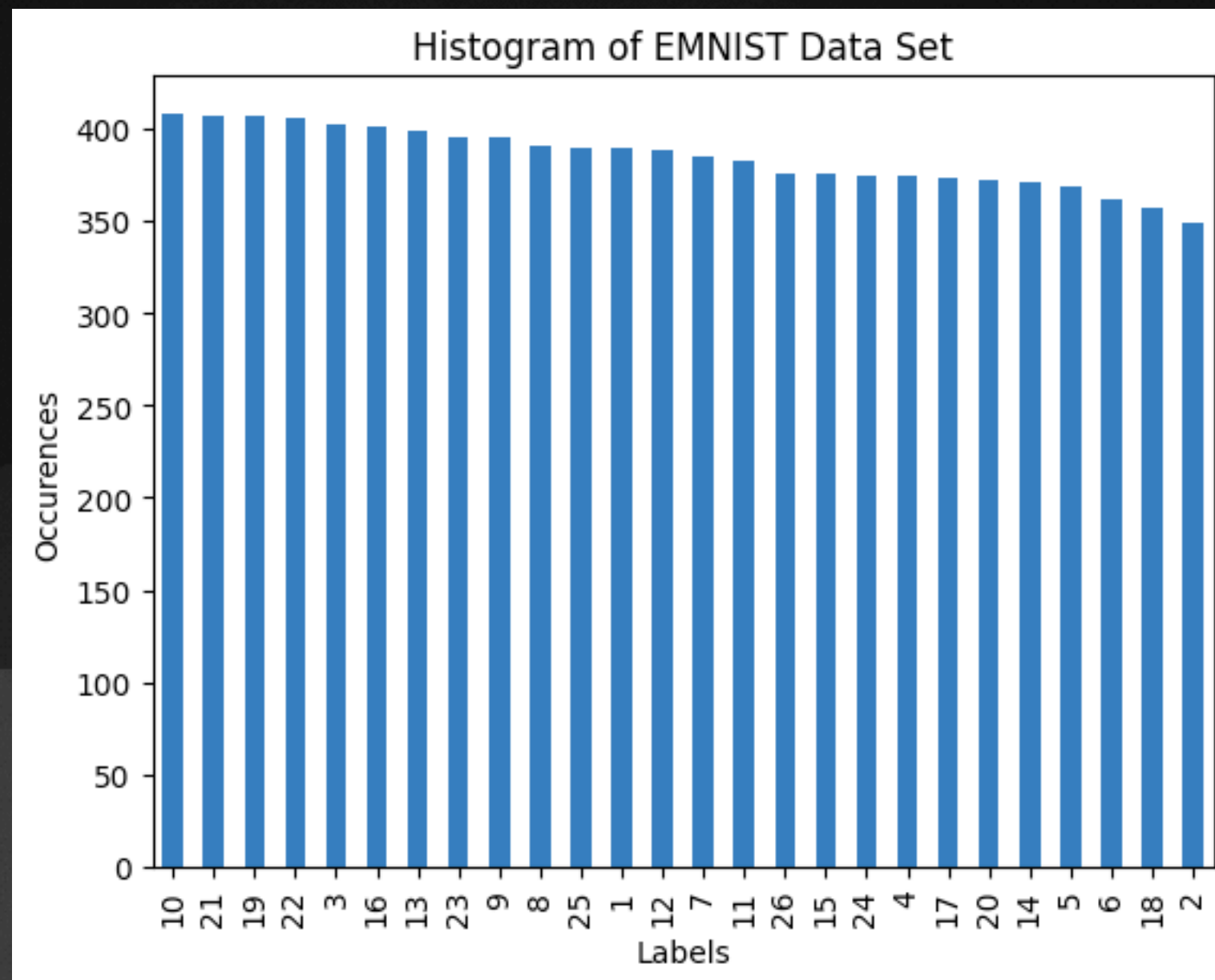
The background of the slide features a series of dark, layered, wavy shapes that resemble a mountain range or a series of hills. The colors are various shades of dark gray and black, creating a sense of depth and texture. The shapes are smooth and flowing, with some peaks and valleys. The overall effect is a minimalist, abstract landscape.

Problem Statement

- Image Classification is becoming increasingly important in many industries, including: Medicine, Agriculture, Autonomous Vehicles, Natural Language Processing, and many more. One specific use case this project will be focusing on is the classification of handwritten letters.
- Often times this type of classification uses computationally intensive Machine Learning models, like Convolutional Neural Nets. This project aims to show that it's possible to use basic, unsupervised and supervised learning techniques to train high performance classifiers, which can be ran on everyday computers.

EMNIST

An Overview

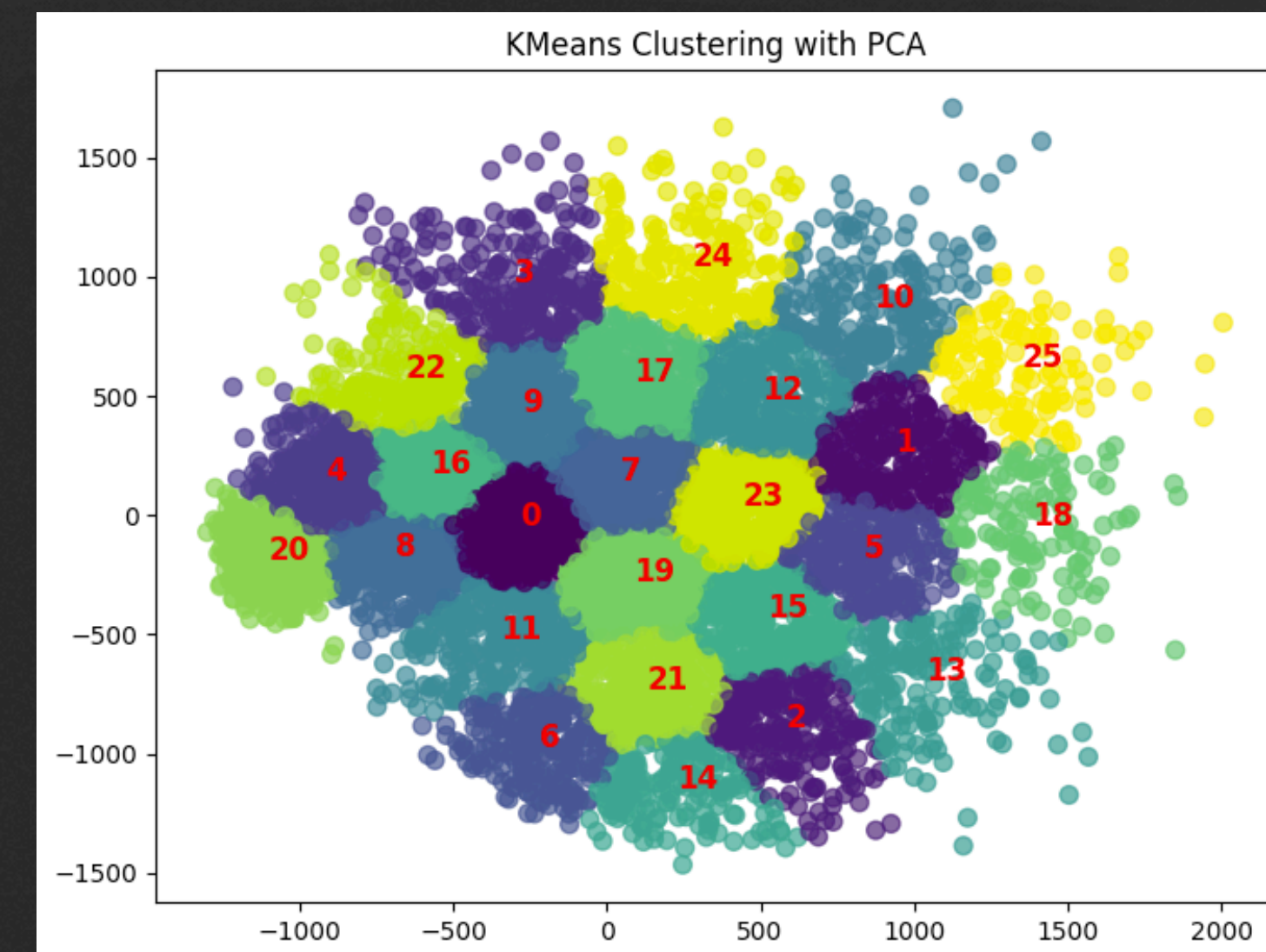
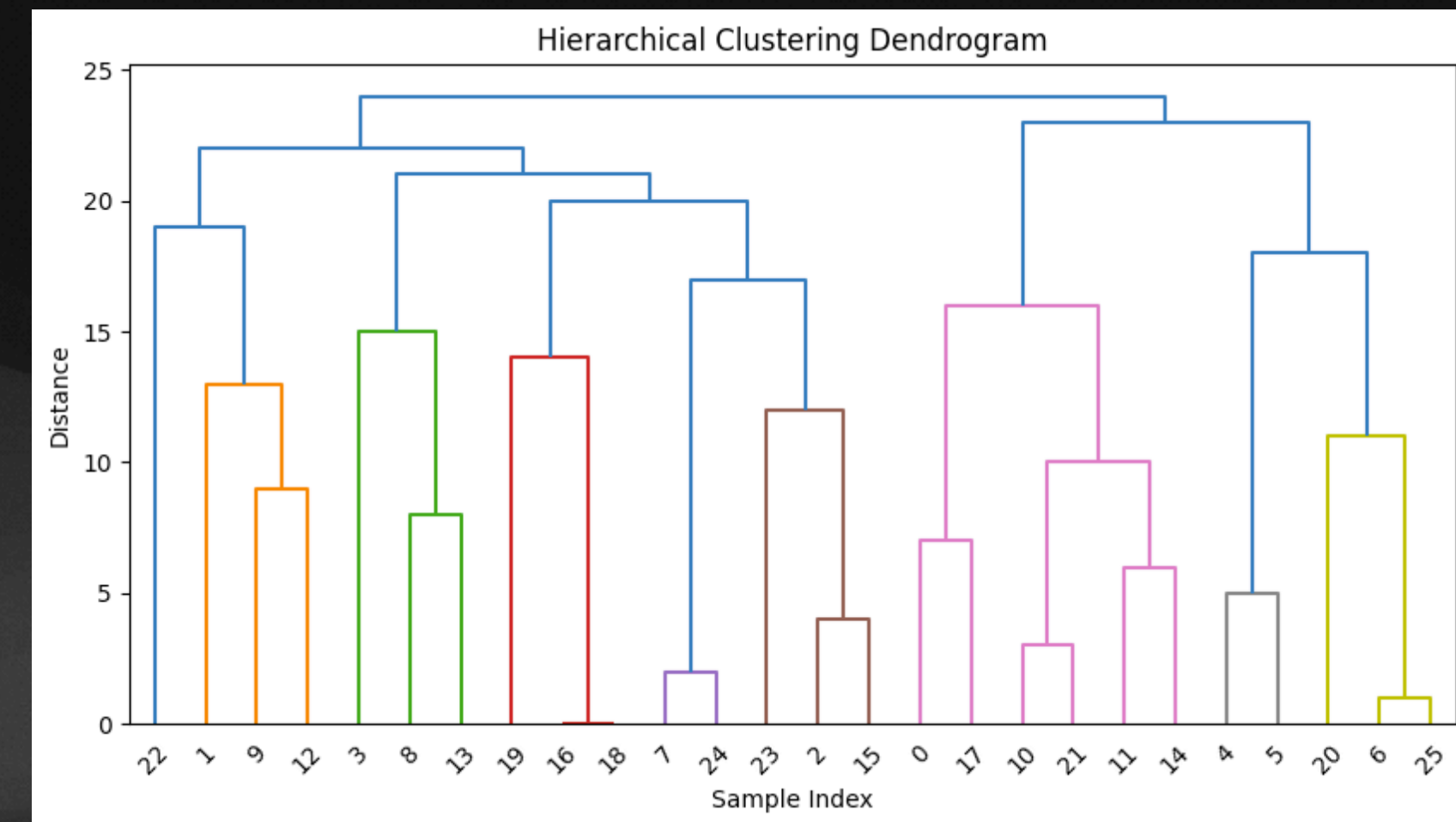


- This project uses the EMNIST Letters Dataset, which contains over 88,000 28x28 handwritten letters. This data was compiled from the NIST Special Database 19.
- Trained on a down-selected, 10,000 image, sample.

Models Used

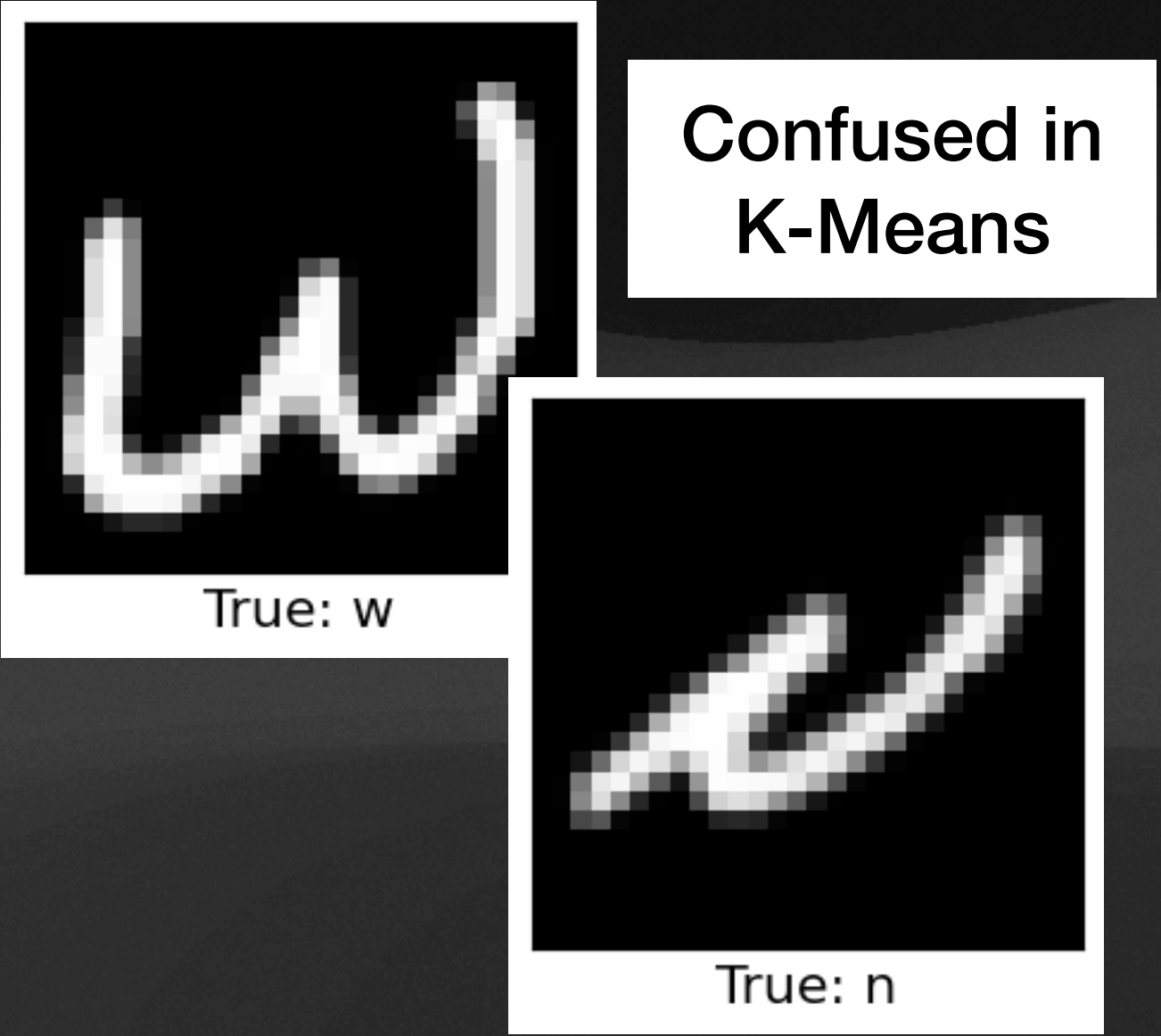
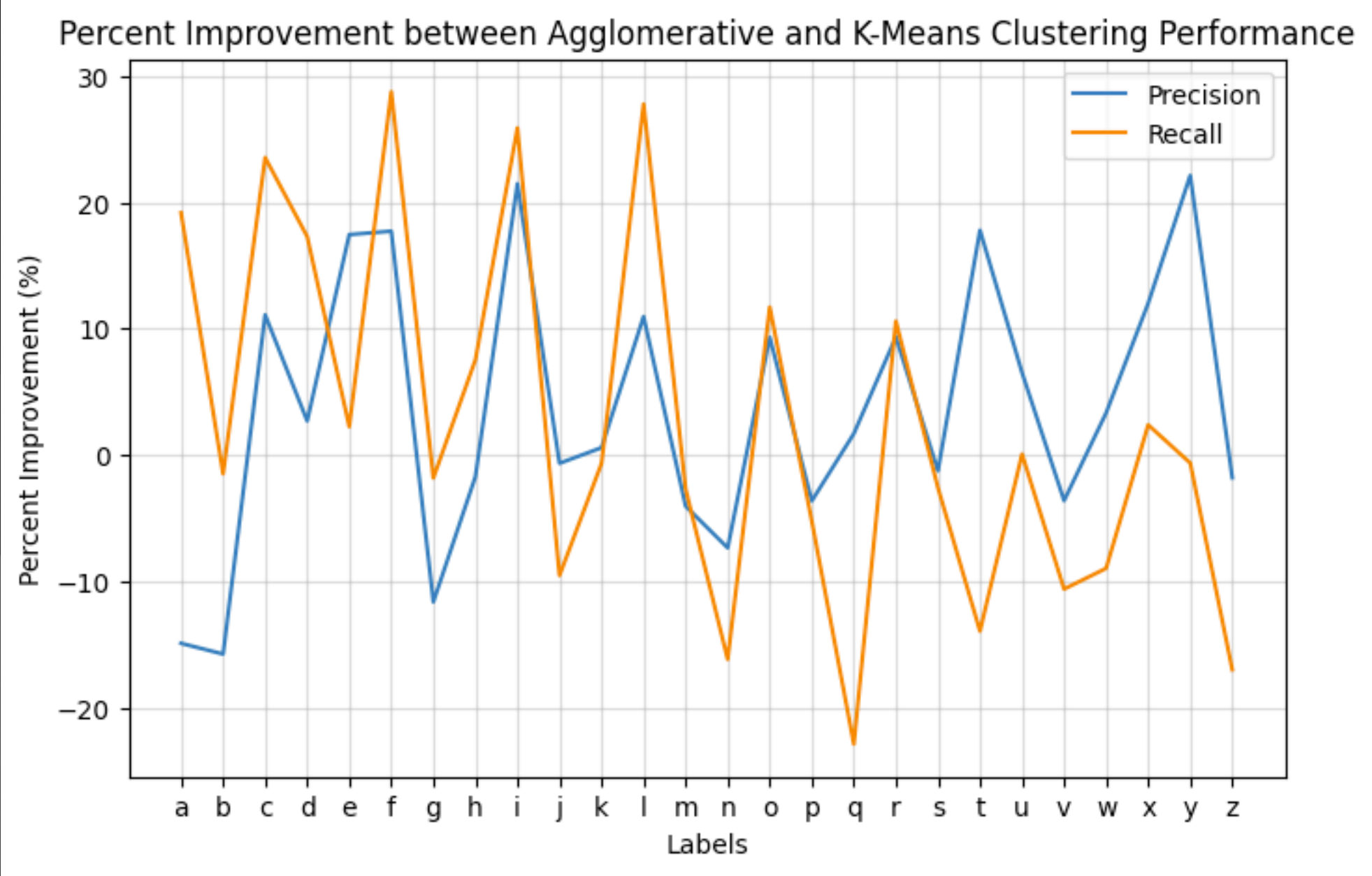
Agglomerative, K-Means Clustering, and K Nearest Neighbor

- In this project, I chose to use Agglomerative and K-Means Clustering as my Unsupervised Learning Models, along with KNN as a Supervised Model.
- These are perfect for building simple Classifiers that enable quick iterative designs. All of these fare well in datasets where features are diverse and clustered.
- As we'll see, Agglomerative Clustering performs the best, with K-Means and KNN not too far behind.



Performance Summary

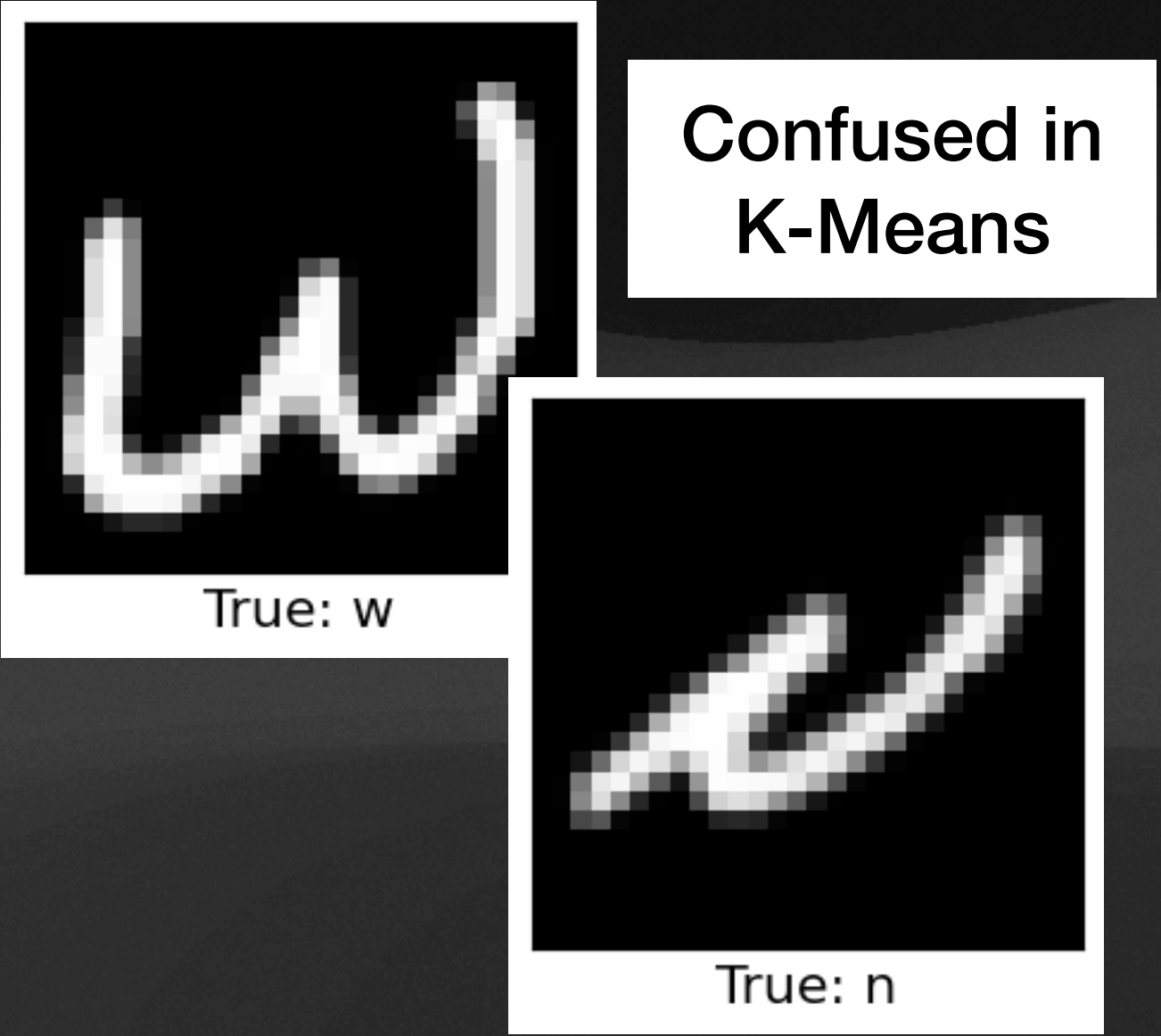
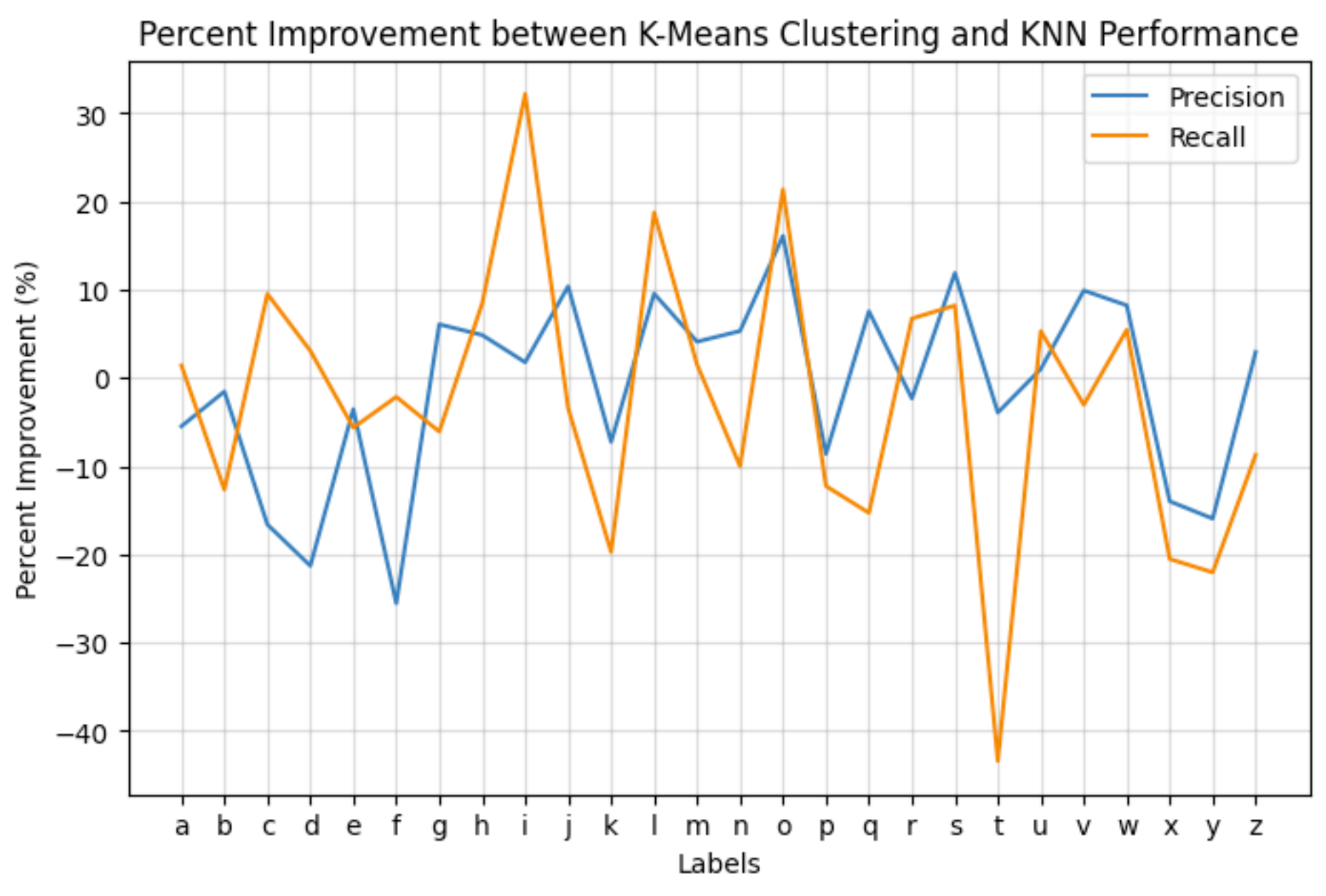
Agglomerative and K-Means Clustering



Model	Accuracy	Mean Recall	Mean Precision
Agglomerative	79%	80%	81%
K-Means	76%	76%	77%

Performance Summary

Agglomerative and K-Means Clustering



Model	Accuracy	Mean Recall	Mean Precision
K-Means	76%	76%	77%
KNN	76%	76%	78%

Takeaways

- Able to achieve up to 80% classification accuracy using minimal tuning
- Models are easily developed and iterated upon
- Allows for models to be developed on everyday computers and used on a larger scale.

The background features a stylized mountain range with multiple layers of peaks. The mountains in the foreground are a vibrant blue, while the subsequent layers become progressively darker, eventually blending into a solid black background at the top. The peaks are smooth and rounded, creating a layered, atmospheric effect.

Thank You!