Names: Daniel Rodriguez, Sriharsha Aitharaju

Project Title: Comparative Analysis of Machine Learning Clustering Methods and Neural networks in PyTorch for Image Generation on the Fashion MINST Dataset

Project Proposal:

This project aims to evaluate the performance and effectiveness of traditional machine learning clustering methods, such as k-Nearest Neighbors (KNN), from the scikit-learn library against neural networks implemented in PyTorch. The study will provide a comparison starting with a basic conceptual overview of clustering, followed by their practical application in the context of image generation using the Fashion MNIST dataset. By leveraging pixel clustering, this project will highlight the strengths and weaknesses of each approach with a visual representation highlighting the key differences, if any.

The Fashion MNIST dataset, consisting of 70,000 grayscale images of clothing items, will be used to assess the performance of each clustering method and the neural network. The dataset will be split into training and test sets, with methods subsequently being trained on the data. Methods include k-Nearest Neighbors with and without PCA selection, and PyTorch-based neural networks. The project will explore various architectures and hyperparameters to identify the optimal configuration for image generation.

To visualize and evaluate the effectiveness of each clustering method, t-distributed Stochastic Neighbor Embedding (t-SNE) will be employed to plot the results of the clusters. This will allow for a clear comparison of the methods’ abilities to group similar images together, as well as the quality of the generated images. As a result, we aim to not only compare the results off of visual inspection but also on data grouping being displayed post training of the sets.