Data Visualization of Hemingway's novels

Using Unsupervised Machine Learning Method

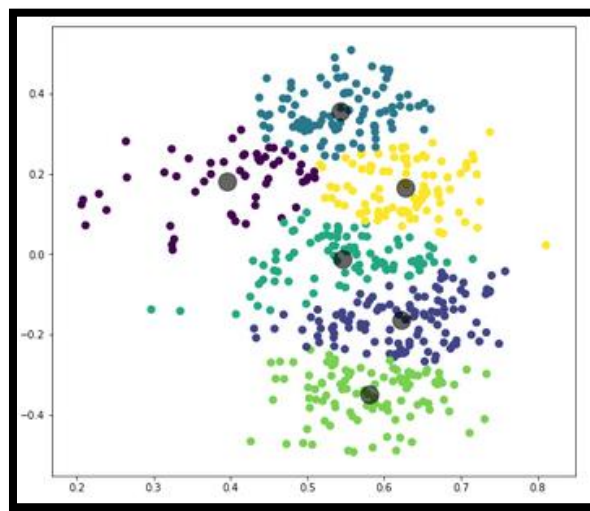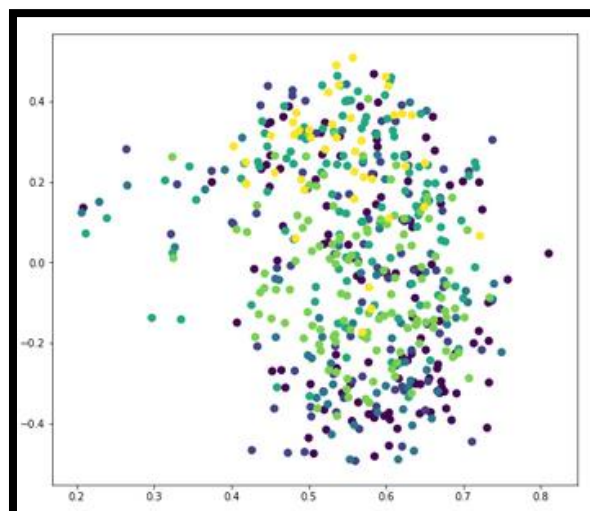*Louis - Kaiyang Liu*

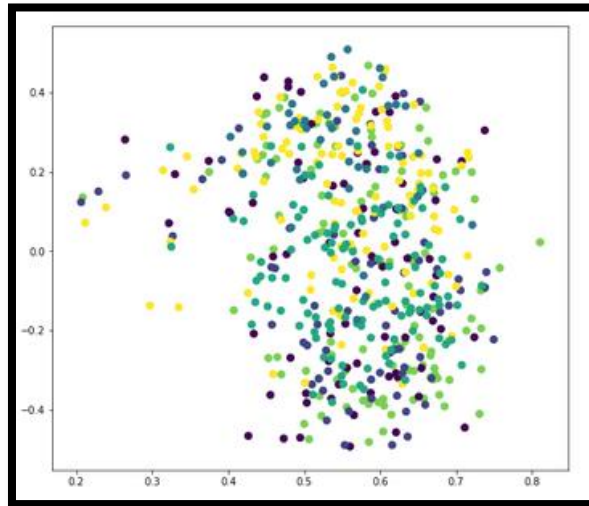*Figure 1 Cluster-SVD_labels of df*

Figure 2 Cluster-id of df
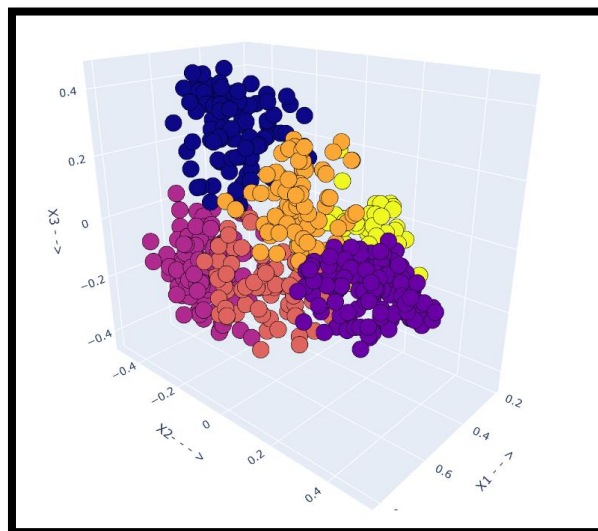
Figure 3 Cluster-transformed id of plot_df

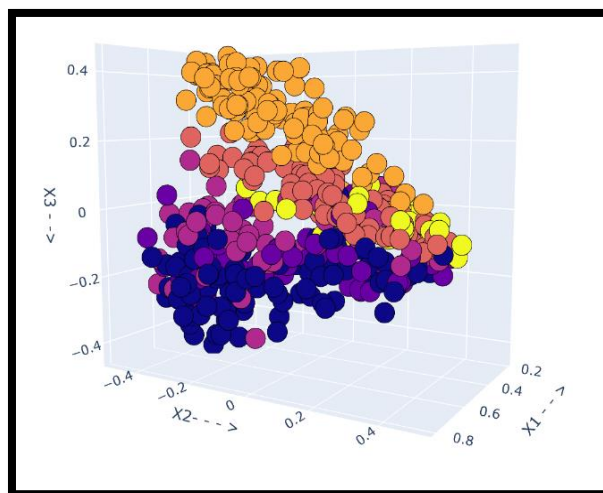*Figure 4 3D Plot- SVD_labels of df*

Figure 5 3D Plot-id of df

Analysis of Hemingway's six novels

using Unsupervised Machine Learning

Louis, Kaiyang Liu

In this report, we mainly analyze the relationship between the text of Hemingway's six novels based on the data visualization that we get from the unsupervised machine learning method – Cluster analysis and explain the insights in it.

First of all, we look at Figure 1, it shows that the cluster plot of texts in six novels after a Single Value Decomposition process. It has 6 different clusters and each cluster has its own center point. The center has its own feature so that the machine can recognize it as a category, and the scatter around them have very similar feature compared to the center point so that they become a cluster. Then we look at Figure 2, we find that there are many overlapping in this cluster plot, which means that the six books have many similar features. However, the yellow plot is a little different, this cluster is located in the relatively upper position, which means that the book *The Old Man and the Sea* have some different features from the other five books, it lacks some of their common feature. We can see more clearly if we compare Figure 3 and Figure 1 to see the position of the cluster corresponding to the color.

Moreover, to observe the relationship between the six novels more intuitively, we can look at the 3D plot that shows in Figure 3 and Figure 4. Figure 3 is the three-dimension plot of the clusters of six books after a Single Value Decomposition process, and Figure 4 is the cluster based on the id. We can find that the orange plot which stands for "*Across the River and Into the Trees*" is mostly independent, which means this book has less similarity feature to other novels. Besides, the other five books intersect together and all these six novels converge to a point, which means that the end of the converge is probably a very general writing style or method of Hemingway.