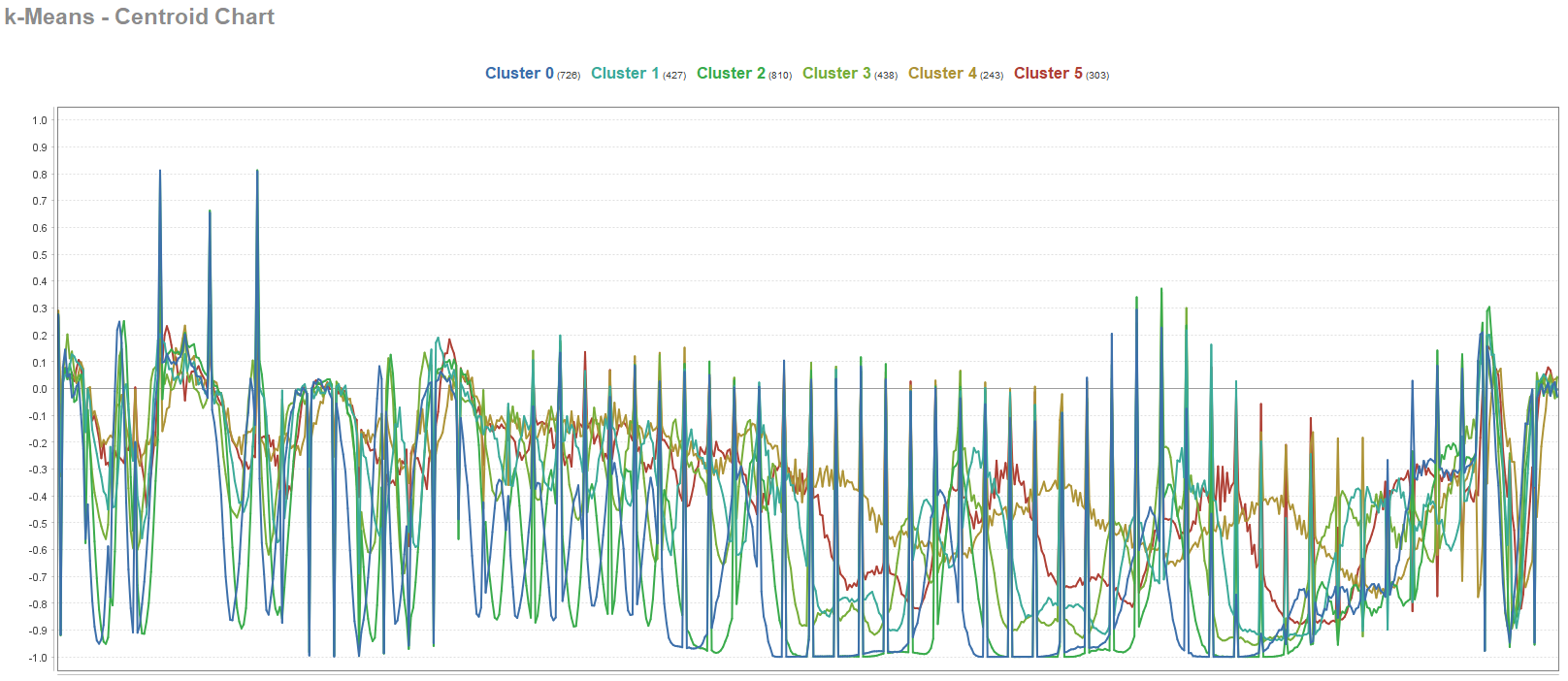


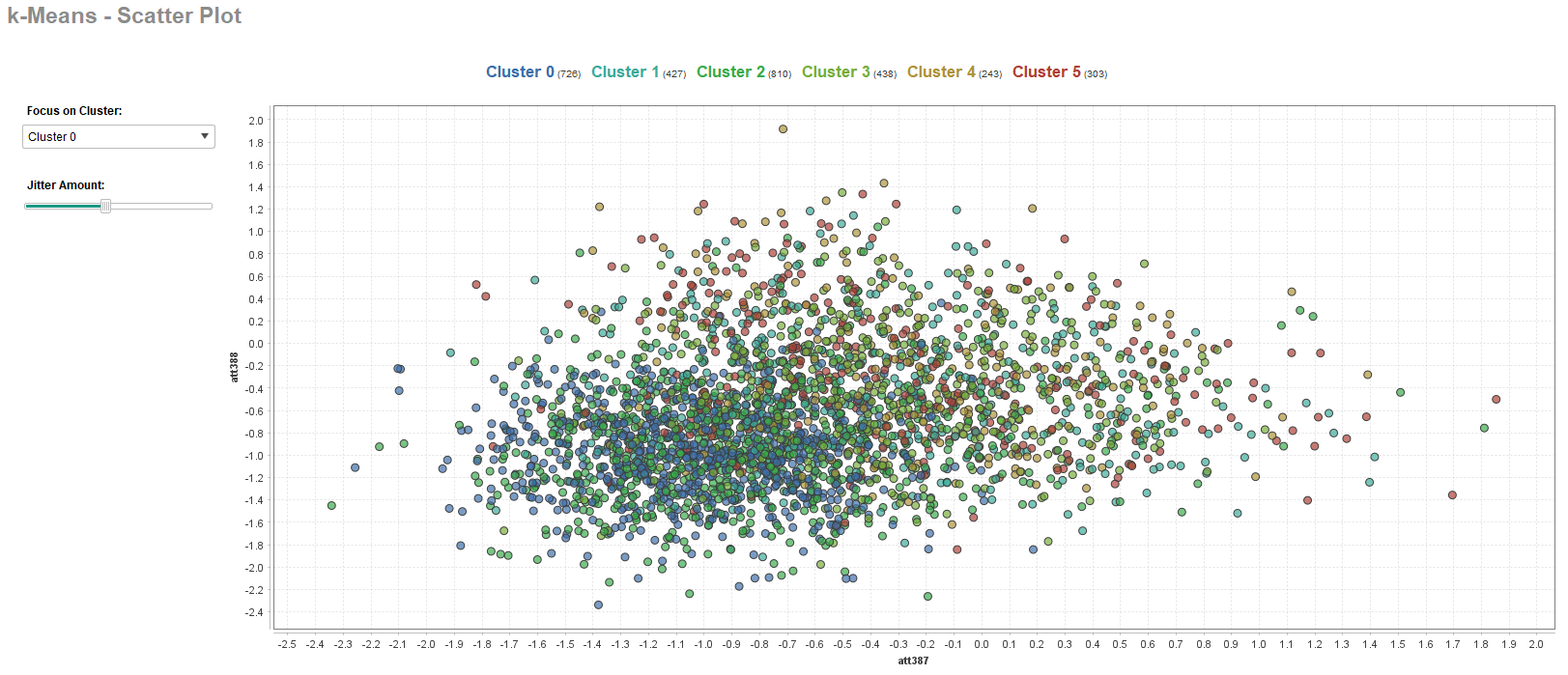
**How do you think the values used serve as the clustering points?**

At each split, a value acts like a rule, deciding which branch the data points go to. The closer the values are between data points, the more likely they end up in the same group. As you move down the tree, the splits become more detailed, making the clusters more specific. These values are what create the boundaries between the different clusters, and as you move further down the tree, the groups become more specific and detailed.



**Review the Cetroid Chart.** **What does this chart show us about the clusters?**

Clusters like 0 and 2 have more ups and downs, meaning their data points vary a lot, while clusters like 1 and 4 have smoother lines, indicating more consistency. The points where the lines spread apart show where the clusters are most different. The numbers next to each cluster tell us how many data points are in each one, showing which patterns are more common.

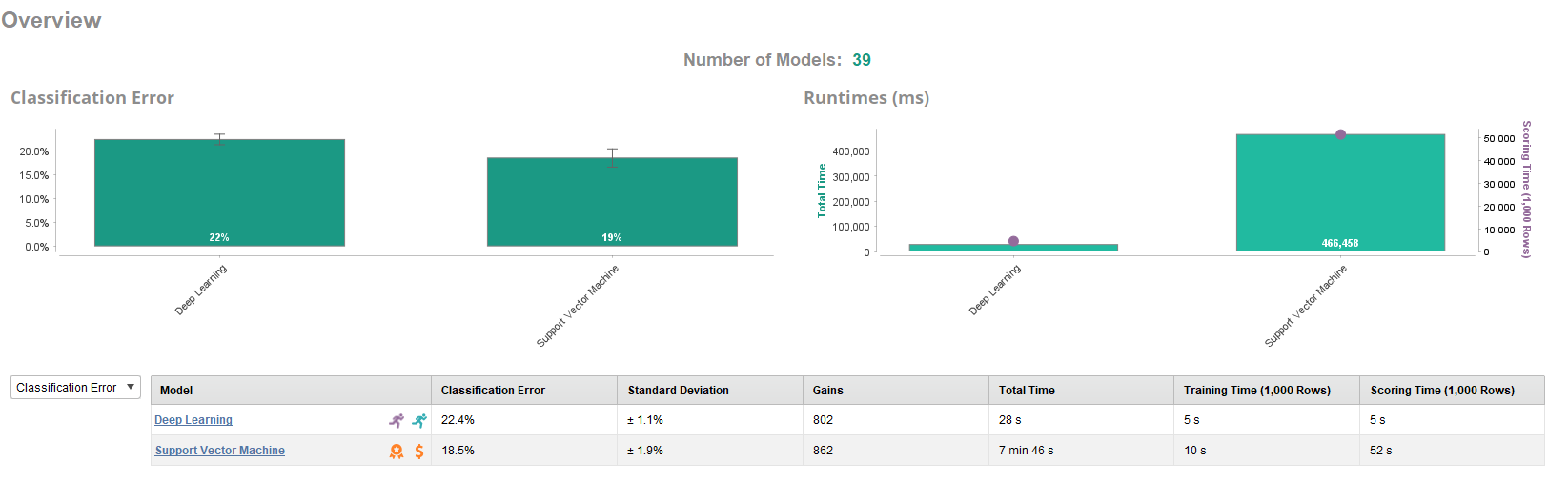


**Focus on Cluster 0 and then** **describe what this scatter plot shows**.

This scatter plot shows the results of a K-Means clustering, where each dot represents a data point based on two attributes, att387 and att388. The different colors show the six clusters, labeled from 0 to 5. Some clusters are bigger, like Cluster 0 with more points, and others are smaller, like Cluster 5 with fewer points. The dots in the middle are packed closely together, meaning the clusters overlap, so the points are pretty similar. Some clusters, like Cluster 5, are spread out more, showing more variation. Overall, this plot helps visualize how the K-Means algorithm grouped the data.

Explain why Cluster Analysis is a useful tool and how might it be used to classify groups of customers.

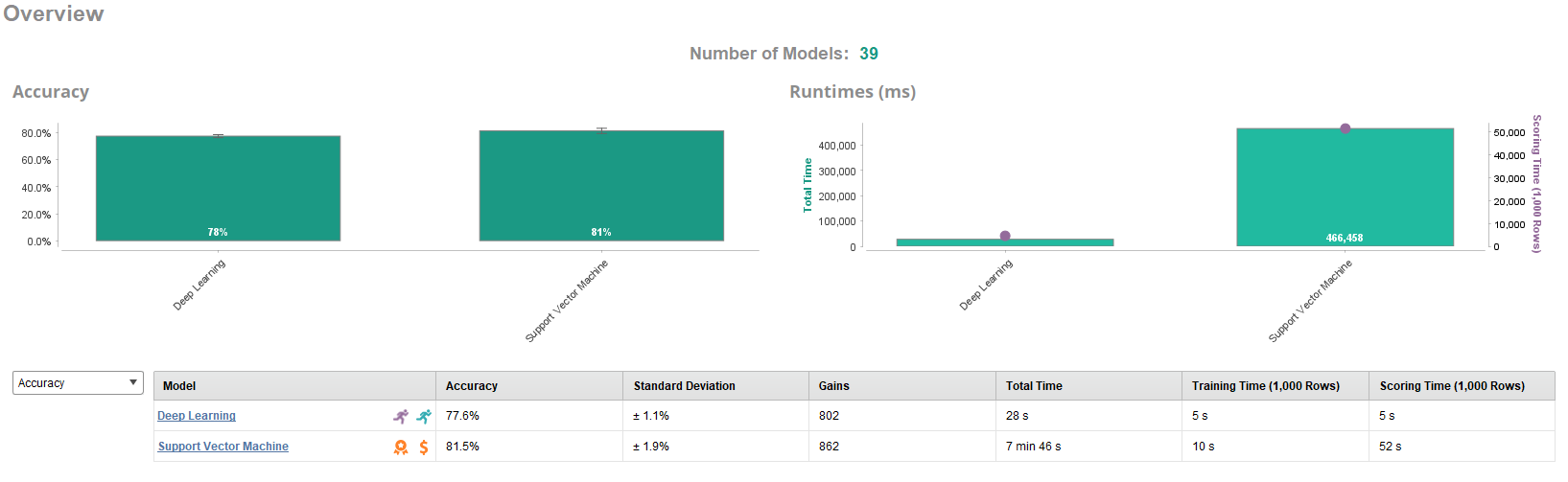
Cluster analysis is a useful tool because it helps find patterns in data and groups similar things together. For businesses, it’s great for organizing customers into different groups based on things like buying habits or interests. For example, one group might like expensive products, while another prefers discounts. By knowing these groups, companies can create better ads, offers, and products that match what each group wants. This helps businesses connect with their customers in a more effective way.



**Which model has a lower Standard Deviation? Is a lower standard deviation a good condition? Why? How does that relate to the Classification error?**

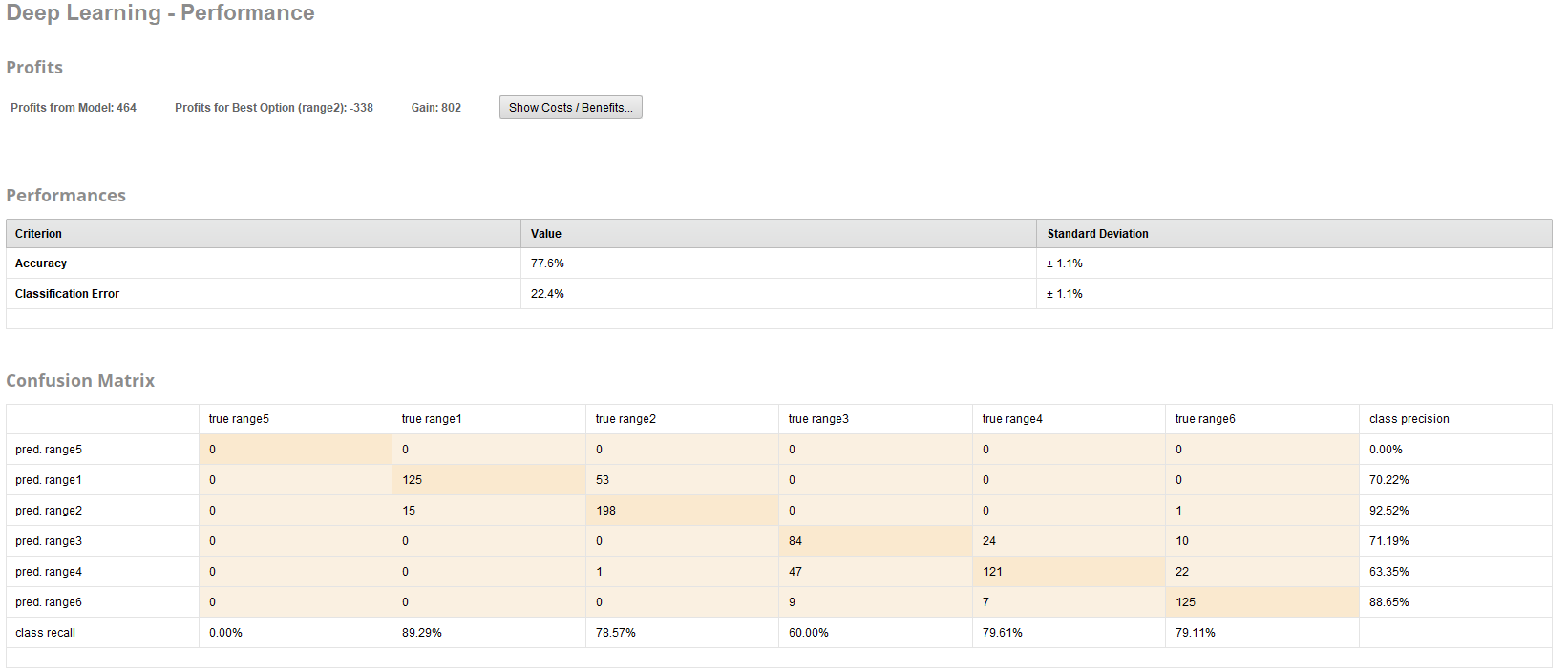
The Deep Learning model has the lower standard deviation, at ±1.1%, compared to the Support Vector Machine's ±1.9%. A lower standard deviation is a good condition because it means the model’s performance is more consistent and reliable across different runs or datasets. This consistency indicates that the model is stable and less likely to produce fluctuating results when exposed to new data.

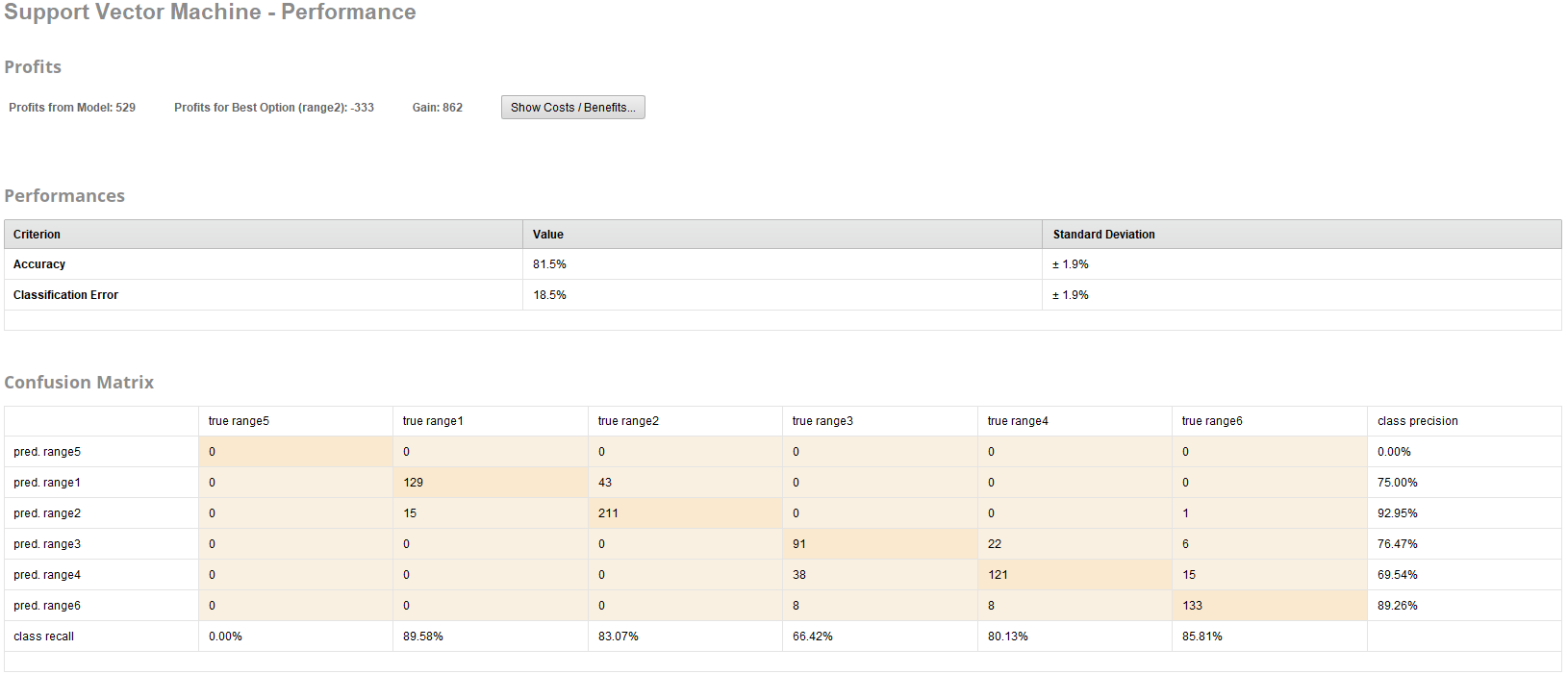
In relation to the classification error, while the Support Vector Machine has a lower classification error rate of 18.5% compared to the Deep Learning model’s 22.4%, the Deep Learning model shows more consistency due to its lower standard deviation. This means that while the SVM may be more accurate in this instance, the Deep Learning model's performance is more predictable and less likely to vary significantly with different datasets. In an ideal situation, you'd want a model with both a low classification error and a low standard deviation for both accuracy and stability.



**Change the tab to Accuracy as shown below. Which model has the higher accuracy?**

The Support Vector Machine model has the higher accuracy, at 81.5%, compared to the Deep Learning model, which has an accuracy of 77.6%.





**Look at the confusion matrix for the Deep Learning (ANN) model versus the SVM model. Which model has a higher precision rate? Explain the lift chart and how it is used.**

The Support Vector Machine (SVM) model has better precision than the Deep Learning model. For example, SVM has a precision of 89.26% for class range6, while Deep Learning has 88.65%. This means SVM is better at correctly identifying true positives.

A lift chart shows how well a model finds important results compared to random guessing. The X-axis shows the percentage of people the model picks as positive, and the Y-axis shows how much better the model is at finding real positives. A higher lift means the model is doing a good job early on, helping you focus on the most important cases.

**Finally, compare and contrast the three model types that we used in this lab (ANN, SVM and k-Means).**

ANNs, SVMs, and k-Means are all different tools for different jobs. ANNs are great for finding complex patterns in large datasets, but they can be slow and hard to understand. SVMs are good at separating data into categories, especially when there’s a clear divide between groups, but they might struggle with bigger datasets. k-Means helps group data based on similarities without needing labels, making it useful for finding patterns, but it doesn’t predict outcomes. In short, ANNs handle complex data, SVMs classify, and k-Means clusters similar data points.