Flowchart for submissions

Data Collection  
(Log data + Error data)

Exact error (Root cause)

Error data attributes

Unsupervised Learning

Dataset with important attributes

Identify important attributes

1. Data Collection (log data + error data)
2. Identify imp attributes
3. Dataset
4. Unsupervised Learning
5. 2 clusters: Good agents and Bad agents
6. Focus on Bad agents features
7. Error data attributes
8. What exactly is the error?

[https://www.quora.com/How-do-I-understand-the-characteristics-of-each-cluster-when-doing-a-K-Means-clustering-algorithm]

When you perform a K-Means clustering algorithm, you end up with a set of clusters where each data point is assigned to one of the clusters. To understand the characteristics of each cluster, you can follow these steps:

1. **Calculate Centroids**: In K-Means clustering, each cluster is represented by a centroid, which is the mean of all the data points assigned to that cluster. Calculate the centroid for each cluster.
2. **Feature Analysis**: Once you have the centroids, you can analyze the features of the data points within each cluster. You can calculate the mean, median, standard deviation, or any other relevant statistics for each feature within each cluster.
3. **Visualization**: Visualizing the clusters can provide a clear understanding of their characteristics. You can use scatter plots or other visualization techniques to plot the data points in each cluster and see how they are distributed in the feature space.
4. **Interpretation**: Analyze the results to understand the characteristics of each cluster. Look for patterns in the data points within each cluster. Are there any common traits or behaviors that define the data points in a particular cluster?
5. **Compare Clusters**: Compare the characteristics of different clusters to identify similarities and differences. This can help you understand how the data points are grouped together and what distinguishes one cluster from another.
6. **Validation**: It's important to validate the results of your clustering analysis. You can use external validation measures such as silhouette score or internal validation measures like the within-cluster sum of squares (WCSS) to evaluate the quality of the clustering and ensure that the clusters are meaningful.
7. **Iterate**: If the initial clustering results are not satisfactory, you may need to iterate the process by adjusting the number of clusters (k) or trying different initialization methods to see if you can achieve better separation and clearer characteristics for each cluster.

By following these steps, you can gain a better understanding of the characteristics of each cluster in a K-Means clustering algorithm and extract meaningful insights from your data.