Assessment 12 - P 4_3: Clustering

Sonia is a program director for a major health insurance provider. Recently she has been reading in medical journals and other articles, and found a strong emphasis on the influence of weight, gender and cholesterol on the development of coronary heart disease. The research she's read confirms time after time that there is a connection between these three variables, and while there is little that can be done about one's gender, there are certainly life choices that can be made to alter one's cholesterol and weight. She begins brainstorming ideas for her company to offer weight and cholesterol management programs to individuals who receive health insurance through her employer. As she considers where her efforts might be most effective, she finds herself wondering if there are natural groups of individuals who are most at risk for high weight and high cholesterol, and if there are such groups, where the natural dividing lines between the groups occur.

Sonia's goal is to identify and then try to reach out to individuals insured by her employer who are at high risk for coronary heart disease because of their weight and/or high cholesterol. She understands that those at low risk, that is, those with low weight and cholesterol, are unlikely to participate in the programs she will offer. She also understands that there are probably policy holders with high weight and low cholesterol, those with high weight and high cholesterol, and those with low weight and high cholesterol. She further recognizes there are likely to be a lot of people somewhere in between. In order to accomplish her goal, she needs to search among the thousands of policy holders to find groups of people with similar characteristics and craft programs and communications that will be relevant and appealing to people in these different groups.

1. Organizational Understanding:

- **Objective Clarification:** Define the goal of the clustering model: to identify high-risk individuals for coronary heart disease based on weight and cholesterol levels.
- **Stakeholder Involvement:** Engage stakeholders to understand their requirements and expectations from the clustering initiative.

2. Data Understanding:

- **Data Collection:** Acquire the dataset containing weight, cholesterol, and gender information for policyholders.
- **Exploratory Data Analysis (EDA):** Explore the dataset to comprehend the distribution, relationships, and potential patterns within the variables.

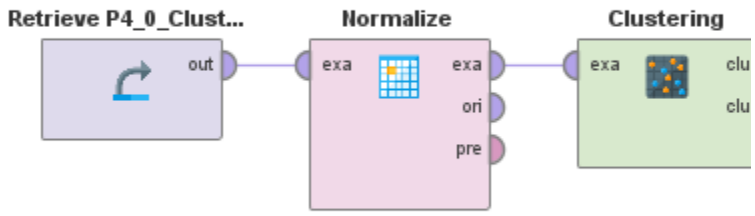
Assessment 12 - P 4_3: Clustering

Open in Turbo Prep Auto Model

Row No.	Weight	Cholesterol	Gender
1	102	111	1
2	115	135	1
3	115	136	1
4	140	167	0
5	130	158	1
6	198	227	1
7	114	131	1
8	145	176	0
9	191	223	0
10	186	221	1

3. Data Preparation:

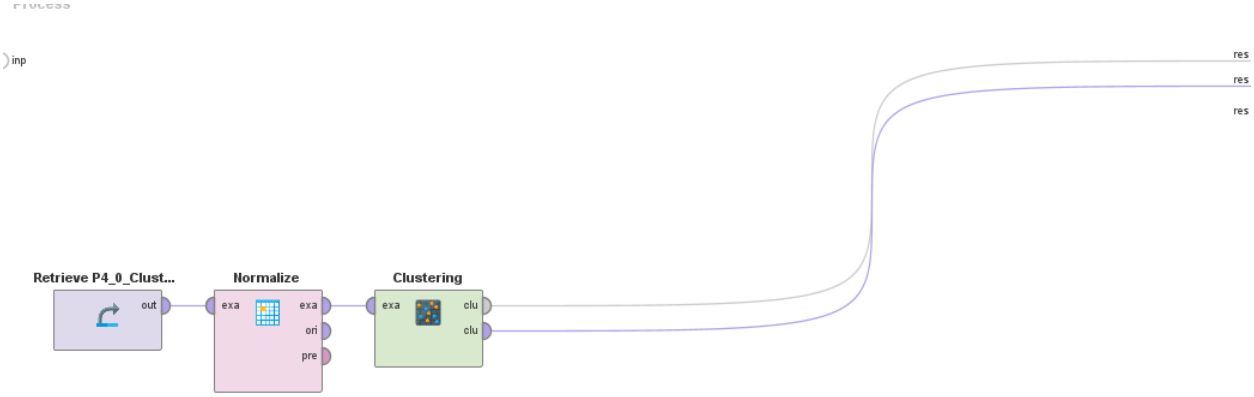
- **Data Cleaning:** Handle missing values, outliers, and inconsistencies in the dataset to ensure data quality.
- **Feature Selection:** Choose weight and cholesterol as relevant features for clustering.
- **Normalization/Standardization:** Normalize or standardize the features for consistent scaling.

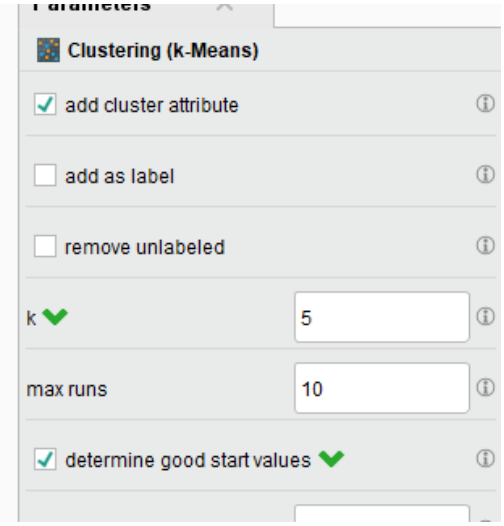


Assessment 12 - P 4_3: Clustering

4. Modeling:

- **Clustering Algorithm Selection:** Choose suitable clustering algorithms like K-Means or Hierarchical Clustering for grouping similar individuals.
- **Model Configuration:** Set parameters (e.g., number of clusters) based on data characteristics or using techniques like the elbow method.
- **Model Training:** Train the selected algorithm using the prepared dataset.





Assessment 12 - P 4_3: Clustering

5. Evaluation:

- **Internal Evaluation:** Use internal metrics (e.g., Silhouette Score) to assess the quality of clusters generated by the model.
- **External Evaluation:** If applicable, employ external metrics (e.g., Adjusted Rand Index) for validation using ground truth labels.

Average Within-Centroid Distance: This metric measures the average distance between data points within the same cluster. A lower value indicates that the data points within a cluster are closer to each other, which is desirable in clustering. It suggests that the members of each cluster are similar to each other in terms of the features being considered (weight and cholesterol in this case).

Cluster-Specific Within-Centroid Distances: The subsequent lines provide the average within-centroid distance for each individual cluster (e.g., Avg. within centroid distance_cluster_0, Avg. within centroid distance_cluster_1, and so on).

- Cluster 0 has an average within-centroid distance of 0.053, indicating that the data points within this cluster are very close to each other.
- Cluster 2 has a higher average distance (0.114), suggesting that the points in this cluster are somewhat more spread out compared to Cluster 0.
- Cluster 4, 8, and 9 also have relatively low average distances, indicating tightly grouped data points within these clusters.

6. Deployment:

- **Program Development:** Design targeted weight and cholesterol management programs for each identified cluster, tailoring interventions to specific group needs.

The clustering analysis has revealed distinct groups among policyholders based on weight and cholesterol levels. Leveraging this valuable insight, the deployment phase focuses on developing highly customized weight and cholesterol management programs for each cluster, ensuring interventions are precisely aligned with the unique characteristics of each group.

1. **Cluster 0 (Low Risk, Low Weight, and Low Cholesterol):**
    - Develop wellness programs focusing on maintaining healthy lifestyle choices.
    - Promote regular physical activity and balanced nutrition to sustain their low-risk status.
    - Provide educational materials about maintaining cardiovascular health.
2. **Cluster 2 (Moderate Risk, Higher Weight, and Moderate Cholesterol):**

Assessment 12 - P 4_3: Clustering

- Implement personalized diet and exercise plans to manage weight and cholesterol levels.
- Offer counseling sessions with nutritionists and fitness experts.
- Introduce group fitness classes and community-based activities to encourage peer support.

3. **Cluster 4, 8, and 9 (Low Risk, Normal Weight, and Moderate Cholesterol):**
    - Concentrate on sustaining their current healthy status with preventive measures.
    - Organize regular health check-ups and cholesterol screenings to monitor their cardiovascular health.
    - Provide incentives for maintaining a balanced lifestyle to prevent future risks.
4. **Cluster 1, 3, 5, 6, and 7 (Moderate to High Risk, Varying Weight, and High Cholesterol):**
    - Design intensive intervention programs with personalized meal plans and fitness routines.
    - Offer one-on-one counseling with healthcare professionals to address specific health concerns.
    - Conduct workshops on stress management, as high-stress levels might contribute to their health risks.