

Fluxograma para clusterização

GUIA DE ETAPAS E TOMADA DECISÃO

NECESSIDADE DE CLUSTERIZAR

Por análise exploratória ou conhecimento prévio do pesquisador é estabelecido o objetivo de checar a existência de possíveis grupos

QUAL MÉTODO DEVE SER USADO

Que tipo de dados estão disponíveis?

(Numéricos, binários, fatores nominais/ordinais...)

Com base nos dados que tipo de agrupamento pode ser feito? particionado em k grupos ou múltiplos níveis de clusters?

Existem muitas variáveis? é necessário reduzir dimensionalidade?

CLUSTERIZAÇÃO HIERÁRQUICA

CLUSTERIZAÇÃO NÃO HIERÁRQUICA

MÉTODO DIVISIVO

MÉTODO AGLOMERATIVO

MÉTODO DE PARTICIONAMENTO

O Método necessita de calculo de métrica?

Sim

APLICAÇÃO DE MÉTRICAS TIPOS DE DISTÂNCIAS

com base na análise exploratória determinar melhor distância

FORMAÇÃO DA MATRIZ DE DISSIMILARIDADE

DETERMINAÇÃO DO Nº DE GRUPOS E ALTURA

Gráfico do cotovelo, Gráfico da Silhueta ou conhecimento prévio do nº de grupos

HIERÁRQUICA

NÃO HIERÁRQUICA

APLICAÇÃO DA FUNÇÃO

Algoritmos como: Divisive Analysis Clustering, Agglomerative Nesting, hclust...

APLICAÇÃO DA FUNÇÃO

Algoritmos como: Kmeans, DBscans, Partitioning Around Medoids, Fuzzy Analysis...

GRÁFICO DE DISPERSÃO

aconselhável para clusterização com até 3 variáveis

TABELA COM MEDIDAS DOS GRUPOS

Medidas das médias dos grupos por atributo, sua quantidades e variâncias

FORMAÇÃO/DESCRIÇÃO DO PERFIL DE CADA CLUSTER

é interessante comparar as médias e variâncias dos grupos.

AJUSTES?

Caso não tenham sido formados grupos satisfatórios para o objetivo, é interessante testar outros métodos e métricas voltando em uma das etapas e ajustando o método

FIM

Vale ressaltar que nem sempre será possível formar bons grupos, mas isso se dará pelos objetivos e variáveis disponíveis.