

Prática - 18

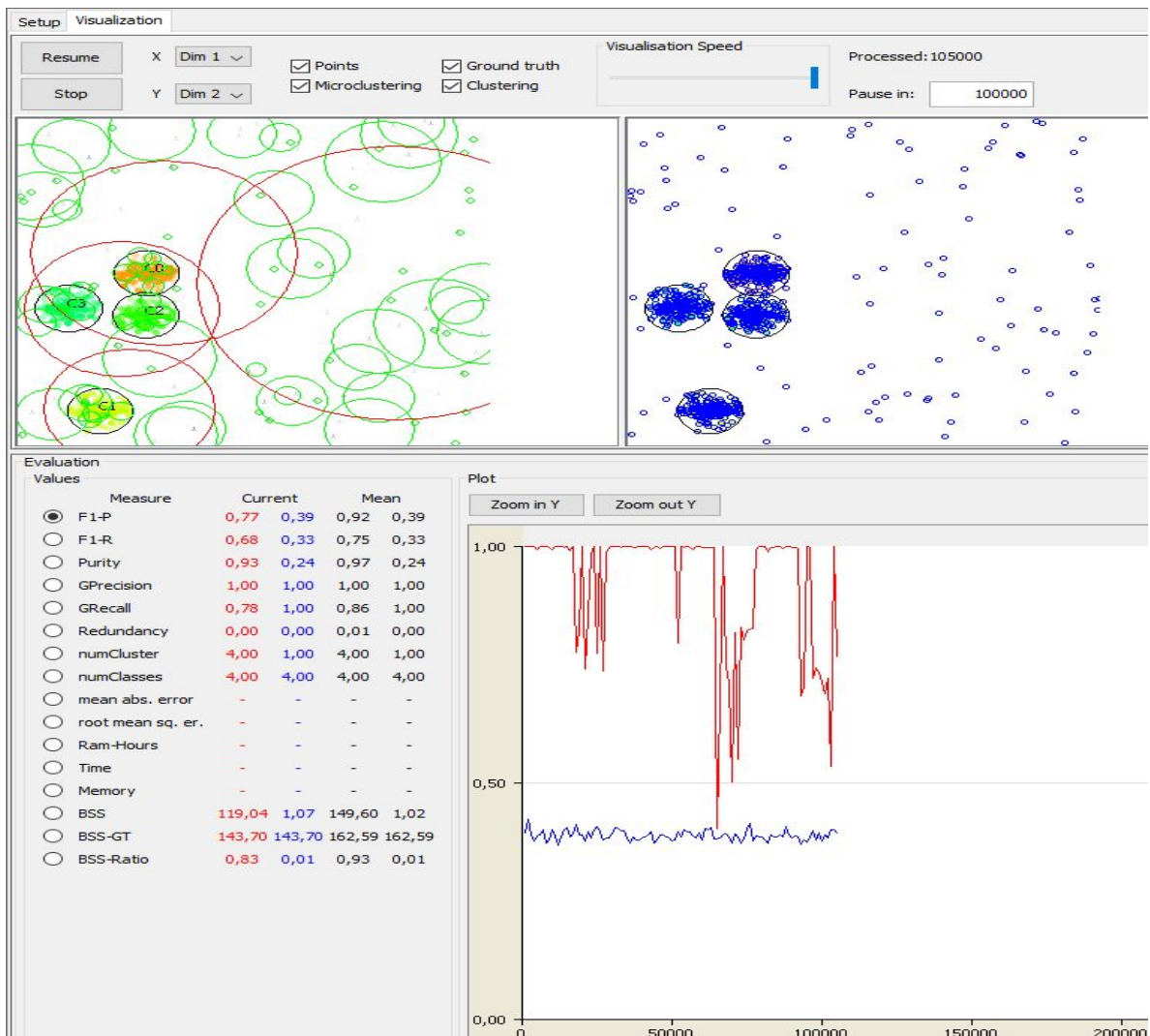
Alunos:

- Pedro Henrique Faria Teixeira - 11621BCC025
- João Daniel de Aquino Rufino - 11621BCC033

1- Usando a ferramenta MOA, faça:

a. Gere uma base de dados com 4 atributos e 4 clusters, habilite a opção de surgimento/desaparecimento de novos clusters a cada 5.000 exemplos. A base deve

conter ruído de 0.1 e 100.000 exemplos. Avalie dois algoritmos de agrupamento usando essa base e duas medidas de validação disponíveis. Discuta os resultados obtidos.



Cluster Algorithm Setup

Stream:

Algorithm1:

Algorithm2:

R:

Podemos observar que o algoritmo CluStream varia mais na seleção de valores para cada cluster em relação ao DsStream.

Também observamos que a **pureza** do Clu Stream é maior que a do DsStream.

Há também uma diferenciação dos clusters, em quanto o Clu Stream manteve o padrão de 4 clusters, o DsStream terminou com apenas 1. Nos dois algoritmos, houve muitos itens que não conseguiram ser clusterizados.

b. Gere uma nova base com as mesmas configurações, mas dessa vez sem ruído. Avalie os mesmos algoritmos usando as mesmas medidas de validação. Discuta os resultados obtidos. Os algoritmos tiveram o mesmo comportamento nas bases com e sem ruído?



Não tiveram o mesmo comportamento, no sem ruído todas as amostras ficaram englobadas dentro de cada clusters, enquanto no com ruído tiveram amostras fora dos clusters. A **pureza** também aumentou em ambos, porém o DsStream ainda continua na desvantagem em comparação ao CluStream.

Durante o processo de clusterização o CluStream ainda continuou variando mais que o DsStream na escolha dos valores, e é por isso que ele consegue obter resultados melhores do que o Ds Stream, pois consegue obter uma ampla região e pegar diversos valores e assim fazer uma melhor comparação. Enquanto o DsStream é pouco seletivo em seus valores, variando bem pouco, se mostra graficamente mais estável porém no resultado final não é tão bom quanto o CluStream. O padrão de clusters finais também seguiu a mesma quantidade.