# Trabalho da disciplina
# Extração do Conhecimento em Fluxos Contínuos de Dados

Elaine Faria
elaine@.ufu.br

Pós-Graduação em Ciência da Computação
Faculdade de Computação
Universidade Federal de Uberlândia

## 1. Introdução

O objetivo principal do trabalho é realizar um estudo detalhado sobre métodos de classificação, regressão, agrupamento, mudança de conceito ou detecção de *outliers* para Fluxos Contínuos de Dados e analisar o seu desempenho em diferentes bases de dados.

- Cada grupo poderá ser composto por no máximo 3 alunos.
- Valor: 40 pontos.
- Data de Entrega do Documento Impresso: 20/06.
- Data da Apresentação: 27/06 e 04/07/17.

## 2. Descrição do Trabalho

Os alunos deverão escolher e comparar pelo menos 3 algoritmos para uma das tarefas de mineração de Fluxo Contínuo de Dados (*data streams*): classificação, regressão, agrupamento, mudança de conceito ou detecção de *outliers.* Os alunos podem usar as implementações, disponíveis na ferramenta MOA, buscar outras implementações na Internet ou implementar o seu próprio código para alguma destas tarefas.

Os algoritmos escolhidos deverão ser executados em pelo menos 4 bases de dados diferentes. É necessário usar bases de dados artificiais e reais. As bases de dados artificiais, podem, por exemplo, serem geradas usando o gerador de dados do *framework* MOA, ou algum gerador de sua preferência. As bases de dados reais podem ser obtidas na Internet (por exemplo, no repositório UCI). Os pré-processamentos necessários deverão ser executados em cada base de dados.

O resultado da tarefa de mineração de dados em cada base de dados deve ser analisado usando medidas de validação adequadas ao problema. O *framework* MOA disponibiliza um conjunto de medidas de validação. Se o aluno quiser, poderá usar estas medidas, ou bucar/ implementar novas.

Cada grupo produzirá um documento (ver Seção 3) que apresentará os algoritmos utilizados, as bases de dados e discutirá os resultados encontrados. É importante que a discussão dos

algoritmos seja crítica, no sentido de apontar vantagens e desvantagens do mesmo, bem como suas principais deficiências.

Se o aluno quiser, poderá usar uma base de dados própria, que esteja relacionada ao seu trabalho na pós-graduação. Nesse caso, converse com o professor antes, para definirem um novo escopo para o projeto.

Cada grupo deverá apresentará o trabalho desenvolvido para o professor.

3. Material a ser entregue

   Documento impresso contendo: i) uma descrição sucinta de cada algoritmo utilizado e o seu respectivo pseudocódigo, ii) uma descrição sucinta de cada base de dados utilizada, e iii) uma descrição sucinta de cada medida de validação. O documento também deve conter os resultados da execução de cada algoritmo em cada base de dados, bem como uma discussão dos resultados obtidos. Lembre-se, se algum pré-processamento foi feito na base de dados, ele deve ser descrito no documento.

4. Apresentação

   Cada grupo terá 30 minutos para apresentar o seu trabalho. Todos os integrantes do grupo deverão participar da apresentação do trabalho.

5. Regras

- Não serão aceitos trabalhos atrasados. Se o grupo não entregar o trabalho no dia combinado, ele receberá nota zero.
- Em caso de projetos copiados de colegas todos os envolvidos recebem nota zero. Lembre-se é muito improvável que haja trabalhos iguais, afinal há várias bases de dados e diferentes algoritmos de agrupamento. Se os alunos usarem códigos disponíveis na Inernet, é preciso entendê-lo antes da apresentação.
- O professor em hipótese alguma verificará ou ajudará na construção do código fonte.
- O professor poderá tirar dúvidas conceituais sobre o trabalho em horário de aula ou horário de atendimento.
- A interpretação dos resultados e o entendimento dos algoritmos fazem parte da avaliação e devem ser realizados pelos alunos.
- O professor poderá questionar cada um dos integrantes do grupo no momento da apresentação.
- A nota dos integrantes não necessariamente será a mesma. Se durante a apresentação o professor detectar que algum integrante do grupo não tem domínio sobre o projeto, ele poderá receber uma nota menor que os demais integrantes.