

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Instituto de Ciências Exatas e Informática

Curso de Ciência de Dados

Disciplina: **Aprendizado de Máquina**

Profa.: Cristiane Neri Nobre — [nobre@pucminas.br](mailto:nobre@pucminas.br)

Descrição do trabalho Prático

Valor: 20 pontos

**A partir da base selecionada por cada aluno, cada grupo deverá fazer:**

1. Fazer um pré-processamento criterioso na base de dados (selecionar atributos, codificar, balancear, eliminar outlier, etc).
2. Utilizar alguns algoritmos vistos em sala: árvore de decisão, Random Forest, K-means, Redes Neurais, Apriori, etc para extrair conhecimento da base de dados. O grupo poderá investigar outros métodos também. Isso inclusive é muito recomendado 😊
3. Gerar o relatório, com os resultados da avaliação

Todas as entregas deverão estar no **formato da ACM**, disponível no CANVAS. Favor conferir o template.

O **relatório final** deverá possuir as seguintes seções:

- 1) Uma pequena contextualização do problema abordado
- 2) Explicação da metodologia adotada para resolver o problema  
Descrição da base de dados, apresentando todos os atributos da base.
  - a. São atributos binários?
  - b. São atributos contínuos?
  - c. São atributos nominais?
  - d. Quais os possíveis valores para cada atributo?
  - e. Quais os valores máximo e mínimo de cada atributo?
  - f. A base é supervisionada? Ou seja, tem alguma classificação? Exemplo: base onde já tem uma classificação de pessoas gripadas ou não gripadas, com base nos atributos da base.
  - g. Que método está sendo utilizado para resolução do problema?

h. Que etapas de pré-processamento foram realizadas?

3) Resultados e discussões

Explicações dos resultados obtidos.

Apresentar os resultados das métricas de avaliação para o problema abordado.

- a. Se for uma base supervisionada, indicar os valores de precisão, sensibilidade, etc..
- b. Se for uma base não supervisionada, indicar os valores de métricas, como silhouette.

Os resultados, de acordo com os parâmetros alterados

- c. Rede neural necessita de alguns parâmetros e árvore de decisão de outros, por exemplo...

Discutir os resultados obtidos

4) Referências

O **relatório final** deverá ser entregue e apresentado, pelo CANVAS, até o dia **11 de dezembro**.

Por favor, para se inspirarem, vejam os artigos: **healthinf20241.pdf**, **healthinf20242.pdf** e **healthinf20243.pdf**, **ICMLA2019.pdf**, **SMC2019B.pdf** e **SMC2019A.pdf**, **health** que estão no **CANVAS**.

O trabalho será dividido nos **dois pontos de controle**:

Assunto	Valor	Data
<b>Implementação de um algoritmo de Aprendizado na base escolhida</b> Apresentar todas as etapas de pré-processamento realizadas:  <b>Exemplos:</b> Codificação dos atributos, identificação de outlier, inconsistência, visualização dos dados, balanceamento, normalização, redução de dimensionalidade, discretização, etc  Implementar algum algoritmo visto em sala na base de dados escolhida e verificar os resultados a partir da base de dados selecionada.	6 pontos	Até 20/10

<p>Avaliar o desempenho e fazer um pequeno relatório discutindo os resultados e discussões. O relatório deve conter todas as etapas realizadas.</p> <p><b>Para esta etapa, observe as recomendações contidas neste documento. Além disso, veja o documento abaixo para maiores informações:</b></p> <p>Template_ACM_Word_EXEMPLO.doc</p>		
<p><b>Relatório final</b></p> <p>O relatório final deverá conter:</p> <ol style="list-style-type: none"> <li>1) Uma contextualização do problema abordado</li> <li>2) Uma explicação da metodologia utilizada no trabalho. O que foi feito? Como? Que etapas de pré-processamento foram realizadas? Explicar, com detalhes, cada escolha realizada.</li> <li>3) Apresentar os resultados e discussões obtidos.</li> <li>4) O trabalho final precisa utilizar pelo menos 3 algoritmos de aprendizado de máquina diferentes. Como é uma competição, ganha quem melhor resolver o problema.</li> </ol>	14 pontos	Até 11/12

Qualquer dúvida, entre em contato com [nobre@pucminas.br](mailto:nobre@pucminas.br)