

# UNIVATES CENTRO UNIVERSITÁRIO

## Inteligência Artificial

### Mineração de dados

Número máximo de integrantes: 3

Professor: Evandro Franzen

## Descrição do trabalho

O presente trabalho consiste na aplicação de métodos de aprendizado de máquina supervisionado para classificação de dados organizados em diferentes formatos, tabulares, texto e imagens. O principal objetivo é explorar diferentes técnicas, configurações e arquiteturas para avaliar quais delas permitem obter os melhores resultados.

## Atividades

### 1. Seleção das bases de dados

Selecionar as bases de dados que serão utilizadas no trabalho (mínimo três *datasets*). Todas as bases de dados devem ser adequadas para a tarefa de classificação, ou seja, devem possuir um rótulo associado a cada entrada. Os formatos dos dados para cada uma das bases deve seguir as seguintes definições:

- Base 1: Dados tabulares, com atributos numéricos ou categóricos. O grupo pode decidir desconsiderar alguns dos atributos, se entender que os mesmos não são relevantes.
- Base 2: Imagens coloridas, relacionadas a uma determinada temática, como, por exemplo, plantas, animais, desenhos ou outras.
- Base 3: Dados textuais, com a existência de um atributo de texto e um rótulo, como, por exemplo, um comentário que pode ser classificado como positivo ou negativo.

Buscar as bases em repositórios públicos. Um documento com a descrição dos integrantes do grupo, das bases de dados e atributos deverá ser entregue até 26/05.

### 2. Pré-processamento ou preparação dos dados

Podem ser realizadas transformações, conversões de formato, limpeza e outras atividades de acordo com a necessidade. Caso a base esteja adequada para a aplicação dos métodos, não é obrigatória a realização destes processos.

### 3. Aplicação dos métodos de aprendizado de máquina

A aplicação dos métodos deve seguir as seguintes definições:

- Escolher um dos métodos estudados na primeira parte da disciplina, árvore de decisão, random forest, naive bayes ou outro para ser aplicado nos 3 conjuntos de dados. Efetuar testes com 3 diferentes combinações de hiperparâmetros para avaliar os resultados.
- Aplicar métodos baseados em redes neurais nos 3 conjuntos de dados. Utilizar tanto o pacote scikit learn, quanto o keras+tensorflow. Criar diferentes arquiteturas (no mínimo 5), com o objetivo de avaliar quais apresentam os melhores resultados. Redes com

camadas convolucionais devem ser aplicados, pelo menos, em imagens e dados textuais.

#### **4. Avaliação e interpretação dos resultados**

Os resultados obtidos devem ser interpretados e analisados pelo grupo. Realizar comparações dos resultados obtidos considerando o desempenho (acurácia, recall, precision, etc) e os tempos para execução. Elaborar tabelas e descrições comparativas, que permitam compreender quais métodos e arquiteturas mostraram melhores resultados e o tempo necessário para obter os resultados.

#### **Entrega e apresentação**

Descrição parcial da base, atributos: 26/05.

Aulas para desenvolvimento do trabalho: 02/06 e 23/06.

Seminário (apresentação do trabalho): 07/07.

Entrega da documentação completa das atividades desenvolvidas, arquivos de dados, resultados: 07/07.