

## Bacharelado em Sistemas de Informação

### Trabalho sobre Mineração de Dados

Disciplina: Inteligência de Negócios - 2020/1 - Profa Kelly

#### Previsão de Ataque Cardíaco

##### 1.0 Introdução

Um hospital deseja prever ataques cardíacos nos pacientes que dão entrada no seu setor de emergência. Para isso, reuniu informações a respeito de 303 pacientes atendidos previamente, indicando se sofreram, ou não, um ataque cardíaco.

O objetivo é dizer se um paciente irá sofrer ou não um ataque cardíaco, utilizando os seguintes atributos:

- Idade (age): em anos
- Sexo (sex): 1 = masculino; 0 = feminino
- Tipo de dor no peito (cp): 1 = angina típica; 2 = angina atípica; 3=dor não angina; 4=assintomático
- Pressão sanguínea em repouso (trestbps): em mm Hg na admissão do hospital
- Colesterol (chol): em mg/dl
- Glicose em jejum > 120 mg/dl (fbs): 1= sim; 0 = não
- Resultado do eletrocardiograma em repouso (restecg): 0 = normal; 1 = anormal
- Frequência cardíaca máxima alcançada (thalach)
- Angina induzida por exercícios (exang): 1 = sim; 0 = não
- Depressão de ST induzida por exercício em relação ao repouso (oldpeak)
- Inclinação do segmento ST de pico do exercício (slope): 1 = subida; 2 = plana; 3 = descida
- Número de vasos principais (0-3) coloridos por fluoroscopia (ca)
- (thal): 3 = normal; 6 = defeito corrigido; 7 = defeito reversível

Indicador a ser previsto:

- diagnóstico de ataque cardíaco (num) : 1= sim; 0 = não

A base de dados está disponível em <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>

## 2.0 Objetivo do Trabalho

O objetivo do trabalho é realizar um estudo comparativo das técnicas de classificação KNN e SVM (*Support Vector Machine*) na previsão de ataques cardíacos, utilizando métricas e técnicas de validação de modelos.

Para esse fim, os seguintes passos devem ser executados

### I - Implementação dos cenários:

#### KNN:

- Para cada divisão treino/teste obtida pela técnica de validação cruzada k-fold com 5 divisões (k=5):
  - Realizar o treinamento do algoritmo.
  - Executar os testes.
- calcular a média das acurácias obtidas em cada divisão treino/teste
- criar a matriz de confusão com os resultados acumulados (soma de todas as matrizes de confusão geradas a cada divisão treino/teste)
- Calcular a precisão (*precision*) e a revocação (*recall*) a partir da matriz de confusão gerada.
- O procedimento acima deve ser executado para k (número de vizinhos) = 3,5,7 e 9.

#### SVM:

##### -Kernel Linear

- Para cada divisão de treino/teste obtida pela técnica de validação cruzada k-fold com 5 divisões (k=5):
  - Treinar o modelo utilizando o kernel linear
  - Executar os testes.
- criar a matriz de confusão com os resultados acumulados (soma de todas as matrizes de confusão geradas a cada divisão treino/teste)
- Calcular a precisão (*precision*) e a revocação (*recall*) a partir da matriz de confusão gerada.

- Kernel RBF

- Para cada divisão de treino/teste obtida pela técnica de validação cruzada k-fold com 5 divisões (k=5):
  - Treinar o modelo utilizando o kernel RBF
  - Executar os testes.
- criar a matriz de confusão com os resultados acumulados (soma de todas as matrizes de confusão geradas a cada divisão treino/teste)
- Calcular a precisão (*precision*) e a revocação (*recall*) a partir da matriz de confusão gerada.

**As divisões (folds) geradas a partir da validação cruzada devem ser as mesmas em todos os cenários, ou seja, você fará a divisão apenas uma vez..**

### II - Elaboração de um artigo científico apresentando os resultados obtidos.

### 3.0 Implementação

O código deve ser implementado em Python e será executado no Jupyter Notebook através do Anaconda. É permitida a utilização dos pacotes distribuídos pelo Anaconda, tais como, sklearn, pandas e numpy.

O programa gerado não deverá solicitar nenhuma entrada ao usuário. Vale dizer que os códigos que não executarem serão desconsiderados.

### 4.0 Artigo

O artigo deverá conter as seguintes seções:

- Resumo: onde o conteúdo do artigo deve ser descrito de forma sucinta.
- Introdução: contextualizando o trabalho.
- Referencial teórico: onde deve ser apresentada a teoria acerca das técnicas de classificação KNN e SVM.
- Metodologia: onde devem ser descritos os testes realizados e a base utilizada (não precisa descrever cada atributo). Além disso, deve ser abordada a teoria sobre matriz de confusão, validação cruzada e métricas (acurácia, precisão e revocação)
- Resultados: onde serão expostos os resultados alcançados.
- Conclusão: que deve estar relacionada ao objetivo aqui proposto para o trabalho.

O documento deverá ser produzido seguindo o modelo do IEEE transaction, a ser disponibilizado no AVA, ser entregue no formato PDF, e ter no **máximo 6 páginas** (incluindo imagens, tabelas e referências).

Observações:

1. Trabalhos que excederem o número de páginas serão penalizados.
2. Consulte as fontes, referencie-as, mas escreva o texto com suas próprias palavras. Caso seja detectado o plágio, o trabalho será avaliado com nota zero.
3. Documentos desacompanhados do código fonte serão avaliados com nota zero.

### 5.0 Produção e entrega

- O artigo e os códigos produzidos devem ser entregues em um arquivo compactado por meio do AVA até as 23h do dia 13/09/20.
- Trabalho individual
- Nota máxima: 46 pontos
- É possível que o autor seja convocado para responder a questionamentos relativos ao trabalho desenvolvido.
- Caso seja detectada algum tipo de fraude, o trabalho será avaliado com nota zero.
- Envios sem a codificação ou sem o artigo serão avaliados com nota zero.
- Além do conteúdo técnico, será avaliada a qualidade do texto produzido, que deve ser claro, objetivo e ter as informações apresentadas de forma organizada.