

Algoritmos de Classificação



Descrição completa da base de dados escolhidos.

Explicação dos algoritmos de aprendizado de máquina escolhidos.

Explicação do conceito e justificativa dos aperfeiçoamentos escolhidos.

Explicação dos códigos em Python para criação dos modelos.

Discussão sobre o desempenho atingido pelos modelos (acurácia) com e sem o pré-processamento dos dados.

Respostas às perguntas

Conclusões



Banco de dados

O banco de dados foi retirado do site Kaggle.

Dados referente a informações cardíacas ou relacionadas.

Determina se é provável o acontecimento de um ataque cardíaco.

O Banco possui 14 colunas, sendo uma delas o desfecho.



Algoritmos

Árvores de Decisão

- O algoritmo tenta encontrar a melhor variável para começar a construção do nó, depois ele escolhe o melhor ponto para separar os dados.

Naïve Bayes

- Tomando como premissa a suposição de independência entre as variáveis do problema, o modelo de Naïve Bayes realiza uma classificação probabilística de observações, caracterizando-as em classes pré-definidas.

Máquina de Vetores de Suporte — SVM

- Responsável por encontrar a melhor fronteira de separação entre classes/rótulos possível para um dado conjunto de dados que sejam linearmente separáveis. Para o SVM, as diversas fronteiras de separação possíveis que são capazes de separar completamente as classes são chamadas de hiperplanos.



Aperfeiçoamento dos Dados

Exclusão de alguns atributos

- Foram removidas as informações sobre o colesterol obtido através do sensor de IMC, o de glicemia em jejum

Remoção de outliers

- Um outlier é um valor que foge da normalidade e que pode (e provavelmente irá) causar anomalias nos resultados obtidos por meio de algoritmos e sistemas de análise.



Resultados

Sem tratamento de dados

Execução	Árvore de busca	Naive Bayes	SVM
1°	79.120%	84.615%	69.207%
2°	83.516%	83.813%	67.039%
3°	84.714%	83.256%	70.304%

Com tratamento de dados

Execução	Árvore de busca	Naive Bayes	SVM
1°	82.365%	85.265%	72.956%
2°	81.695%	87.582%	72.585%
3°	88.569%	84.623%	74.845%



Perguntas

Qual modelo apresentou maior e menor acurácia? Justifique este comportamento considerando o funcionamento dos algoritmos de aprendizagem supervisionada utilizados.

O pré-processamento dos dados ajudou a melhorar a acurácia dos modelos? Justifique sua resposta.

Houve alguma diferença na acurácia quando duas melhorias foram adicionadas ao mesmo tempo e quando cada uma delas foi feita separadamente? Justifique sua resposta.

Algoritmos de Classificação

Guilherme Baccarin