



UNIVERSIDADE FEDERAL DO CEARÁ – UFC SOBRAL
INTERNET DAS COISAS – 2024.1 - PROF. WENDLEY S. SILVA

Trabalho prático

Introdução a Machine Learning com Scikit-Learn - Python (individual)

Entrega até 16/04/2024 (terça-feira), 23h55, via SIGAA

Modo de entrega: A solução para o trabalho prático (individual) descrito abaixo deve ser entregue via relatório no SIGAA, com um único arquivo .PDF até o prazo descrito acima. O código-fonte utilizado precisa ficar disponível em um repositório e o seu link deve constar no relatório.

DESCRIÇÃO DO TRABALHO:

Machine learning ou Aprendizado de máquina [Ref. 1] é um campo de pesquisa em ciência da computação, inteligência artificial, e estatística que objetiva treinar algoritmos para aprender padrões e fazer previsões a partir de um conjunto de dados previamente fornecido.

Encontramos aplicações de Machine learning em diversos serviços populares: Netflix, Amazon e Youtube, por exemplo, usam *machine learning (ML)* para fazer novas recomendações de produtos; instituições bancárias utilizam ML para detectar atividades fraudulentas em transações de cartões de crédito, e empresas de assistência à saúde estão começando a usar ML para monitorar e diagnosticar pacientes.

Este trabalho se divide em duas partes: a **Parte 1** consiste na reprodução de um tutorial disponível online, e a **Parte 2** agrega novos conjuntos de dados, com ênfase na investigação. Os algoritmos de ML devem ser implementados em Python e utilizar o Scikit-learn [Ref. 2], uma ferramenta de *machine learning* para Python. O Scikit-Learn pode ser instalado e executado na sua máquina local, ou pode ser usado remotamente, via Google Colaboratory.

PARTE 1:

Reproduzir o Tutorial presente no site do Scikit-Learn [Ref. 3], com a adição de outros dois algoritmos de Machine Learning no experimento, demonstrando em Tabelas e gráficos qual obteve o melhor resultado.

PARTE 2:

Nesta parte você deve escolher um conjunto de dados do Scikit-Learn, dentre os disponíveis no site [Ref. 4], entender (e explicar brevemente no relatório) quais são as suas características e usar pelo menos 4 (quatro) algoritmos de *machine learning* para fazer um comparativo de qual obteve o melhor resultado. Utilizar Tabelas e Gráficos para avaliar o comparativo. Se preferir, pode-se optar por um conjunto de dados disponíveis no Kaggle [Ref. 5].

Tanto na Parte 1 como na Parte 2, o relatório deve conter um link para o seu código-fonte, que pode ficar hospedado em um repositório tipo GitHub, ou no Google Colaboratory, dentre outros. Todos os gráficos também devem ser gerados no Python [Ref. 6] e o código-fonte (para gerar os gráficos) também deve ficar disponível.

O RELATÓRIO:

O relatório deve conter a seguinte estrutura:

- 1 Cabeçalho de identificação (UFC, disciplina, nome, matrícula, data)
- 2 PARTE 1 (até 6 páginas)
 - 2.1 Introdução
 - 2.2 Conjunto de dados utilizados
 - 2.3 Algoritmos utilizados
 - 2.4 Metodologia (*quais as entradas e saídas? quais os parâmetros para o treinamento? quais as métricas utilizadas?*)
 - 2.5 Resultados obtidos (*colocar as tabelas e gráficos obtidos, com uma breve descrição textual de cada resultado*)
 - 2.6 Link para o código-fonte
- 3 PARTE 2 (até 7 páginas)
 - 3.1 Introdução
 - 3.2 Conjunto de dados utilizados
 - 3.3 Algoritmos utilizados
 - 3.4 Metodologia
 - 3.5 Resultados obtidos
 - 3.6 Link para o código-fonte
- 4 Referências

REFERÊNCIAS:

[Ref. 1] O que é machine learning: <https://scikit-learn.org/stable/tutorial/basic/tutorial.html>

[Ref. 2] Como instalar o Scikit-Learn? <https://scikit-learn.org/stable/install.html>

[Ref. 3] Tutorial: <https://scikit-learn.org/stable/tutorial/basic/tutorial.html>

[Ref. 4] Dataset: <https://scikit-learn.org/stable/modules/classes.html#module-sklearn.datasets>

[Ref. 5] Repositório Kaggle: <https://www.kaggle.com/datasets>

[Ref. 6] Gráficos: <https://matplotlib.org/3.1.0/tutorials/introductory/pyplot.html>