

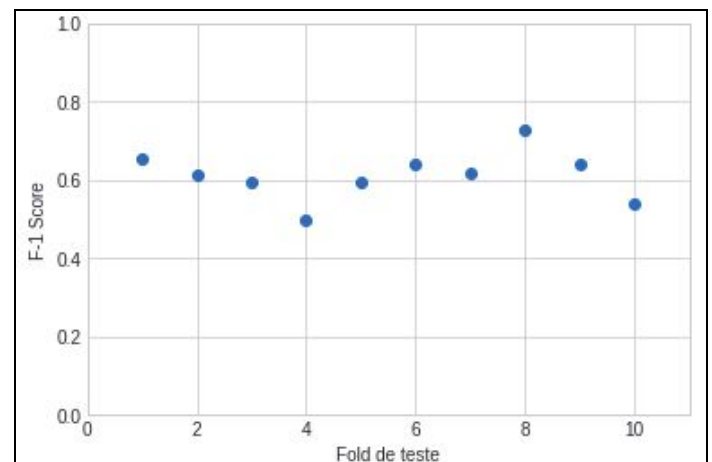
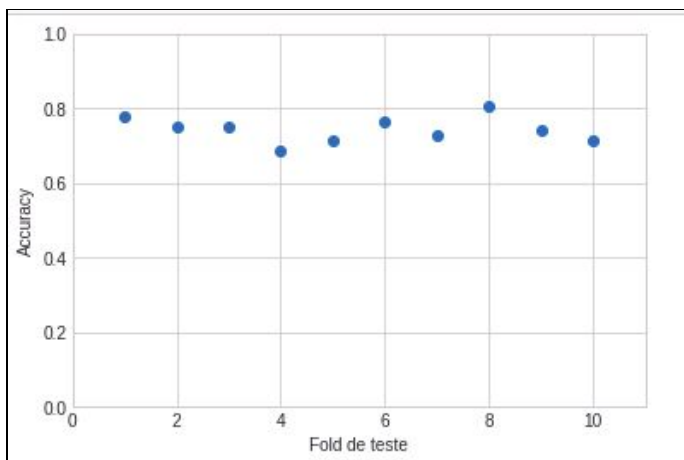
## Exercício: implementação do k-fold cross validation.

O exercício consiste na implementação do algoritmo de aprendizado supervisionado kNN (*k-nearest neighbors*) utilizando a técnica de *k-fold cross validation* com o conjunto de dados *Pima Indian Diabetes*. Foi utilizado  $k = 10$  para a validação cruzada, ou seja, o conjunto inteiro de dados é dividido em 10 partes de tamanho e distribuição aproximadamente igual, das quais nove são usados para treinamento e um é usado para teste.

Além de utilizarmos o  $k = 5$  para o número de vizinhos próximos no algoritmo kNN, que consiste no caso padrão, também avaliamos o desempenho do algoritmo com  $k = 3$  e  $k = 7$ . O desempenho das três variações é comparado abaixo.

A implementação do algoritmo e a geração dos gráficos foram feitas com a linguagem Python. Acompanham este relatório o código fonte em *.py* e o notebook Jupyter utilizado.

### Desempenho 3NN:



#### Acurácia

Média: 0.74

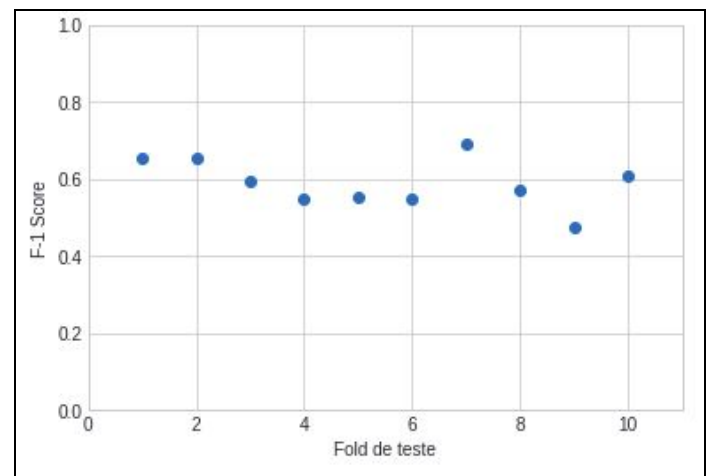
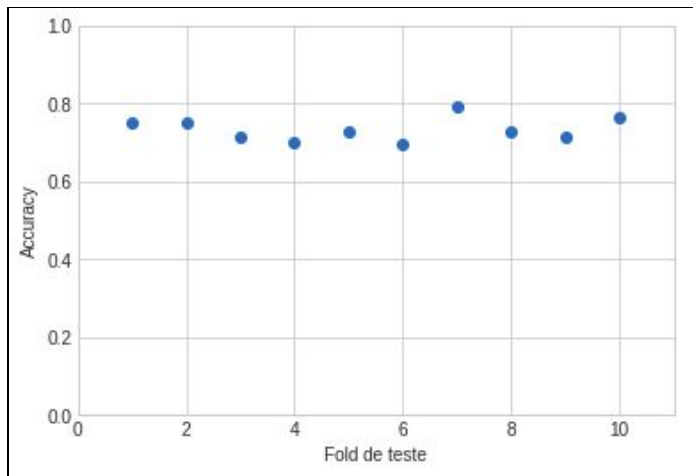
Desvio Padrão: 0.03

#### Escore F-1

Média: 0.61

Desvio Padrão: 0.06

## Desempenho 5NN:



### Acurácia

Média: 0.73

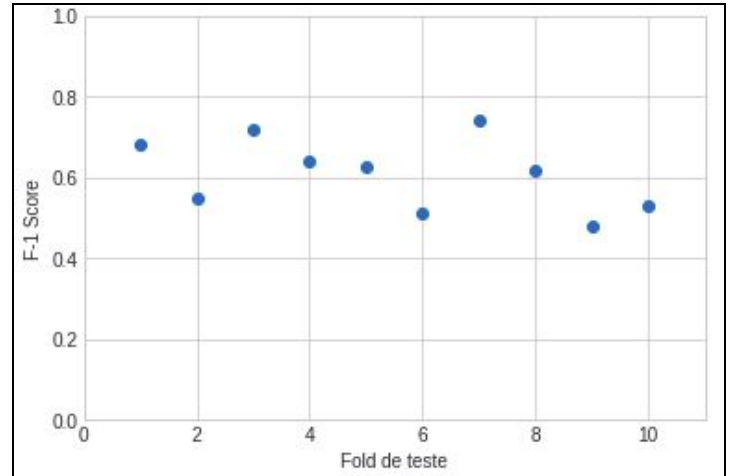
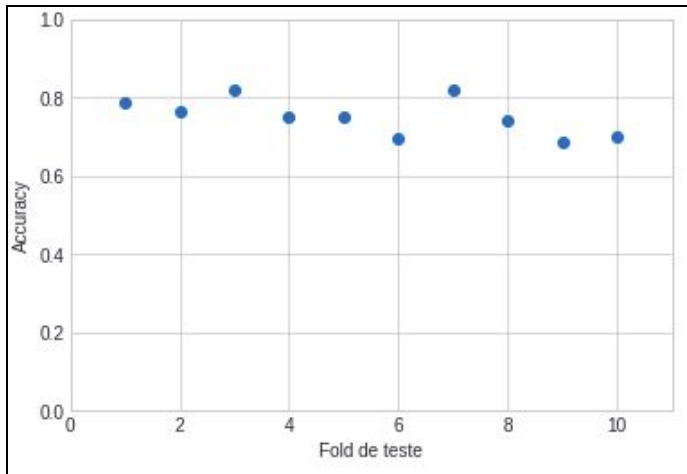
Desvio Padrão: 0.03

### Escore F-1

Média: 0.59

Desvio Padrão: 0.06

## Desempenho 7NN:



### Acurácia

Média: 0.75

Desvio Padrão: 0.04

### Escore F-1

Média: 0.61

Desvio Padrão: 0.09