

Trabalho sobre validação cruzada

Aluno: Carlos Abel Córdova Sáenz

- Código também disponível em: <https://github.com/carlosabcs/KNNCrossValidation>

1) Resultados usando k-fold cross validation

a) k = 5:

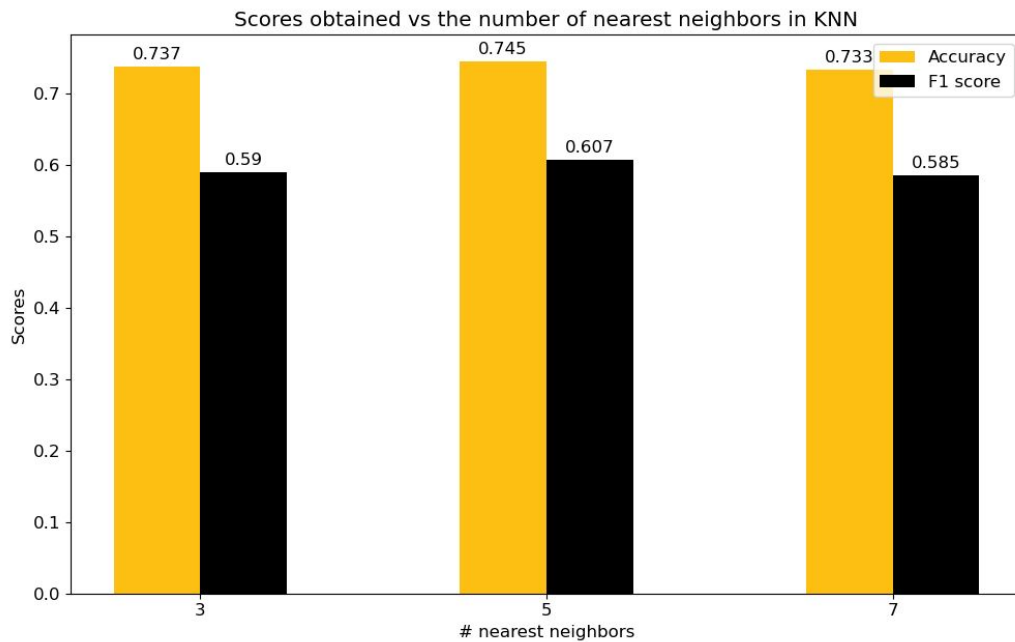


Figura 1. Resultados obtidos para distintos valores de “k” vizinhos mais próximos usando 5 folds

i) 3 vizinhos mais próximos:

Fold	Acurácia	F1-measure
1	0.778	0.660
2	0.747	0.614
3	0.706	0.554
4	0.734	0.539
5	0.721	0.583

Acurácia média (e desvio padrão): **0.737 (0.024)**

F1-measure média: **0.590 (0.043)**

ii) 5 vizinhos mais próximos:

Fold	Acurácia	F1-measure
1	0.719	0.598

2	0.721	0.538
3	0.719	0.566
4	0.825	0.733
5	0.740	0.600

Acurácia média (e desvio padrão): **0.745 (0.041)**

F1-measure média: **0.607 (0.067)**

iii) 7 vizinhos mais próximos:

Fold	Acurácia	F1-measure
1	0.732	0.568
2	0.714	0.577
3	0.771	0.632
4	0.734	0.577
5	0.714	0.569

Acurácia média (e desvio padrão): **0.733 (0.021)**

F1-measure média: **0.585 (0.024)**

b) k = 10:

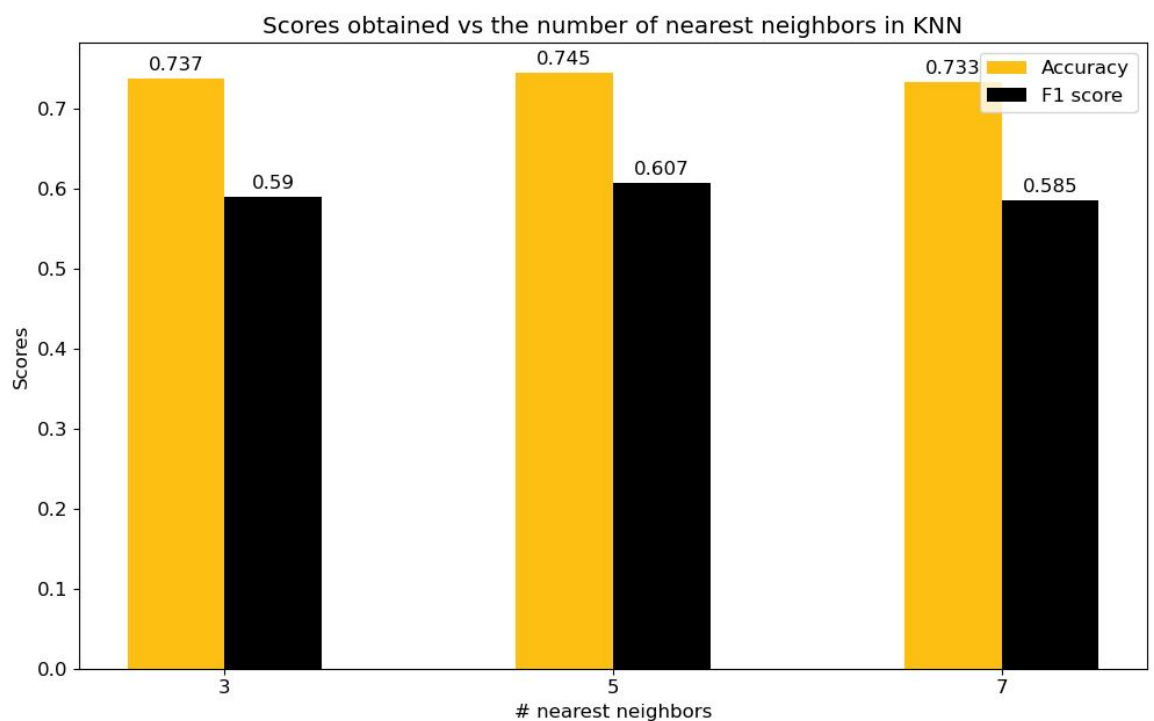


Figura 2. Resultados obtidos para distintos valores de “k” vizinhos mais próximos usando 10 folds

i) 3 vizinhos mais próximos:

Fold	Acurácia	F1-measure
1	0.724	0.632
2	0.779	0.638
3	0.779	0.638
4	0.636	0.517
5	0.727	0.553
6	0.737	0.565
7	0.740	0.524
8	0.805	0.737
9	0.688	0.556
10	0.753	0.655

Acurácia média (e desvio padrão): **0.737 (0.046)**

F1-measure média: **0.601 (0.066)**

ii) 5 vizinhos mais próximos:

Fold	Acurácia	F1-measure
1	0.750	0.612
2	0.779	0.691
3	0.753	0.612
4	0.753	0.655
5	0.636	0.500
6	0.763	0.625
7	0.714	0.560
8	0.714	0.500
9	0.753	0.578
10	0.701	0.549

Acurácia média (e desvio padrão): **0.732 (0.039)**

F1-measure média: **0.588 (0.060)**

iii) 7 vizinhos mais próximos:

Fold	Acurácia	F1-measure
1	0.776	0.605
2	0.779	0.622
3	0.688	0.520
4	0.805	0.681
5	0.753	0.642
6	0.750	0.558
7	0.714	0.560
8	0.688	0.538
9	0.740	0.600
10	0.727	0.632

Acurácia média (e desvio padrão): **0.742 (0.037)**

F1-measure média: **0.596 (0.048)**

2) Resultados usando repeated k-fold cross validation (r = 10, k = 10):

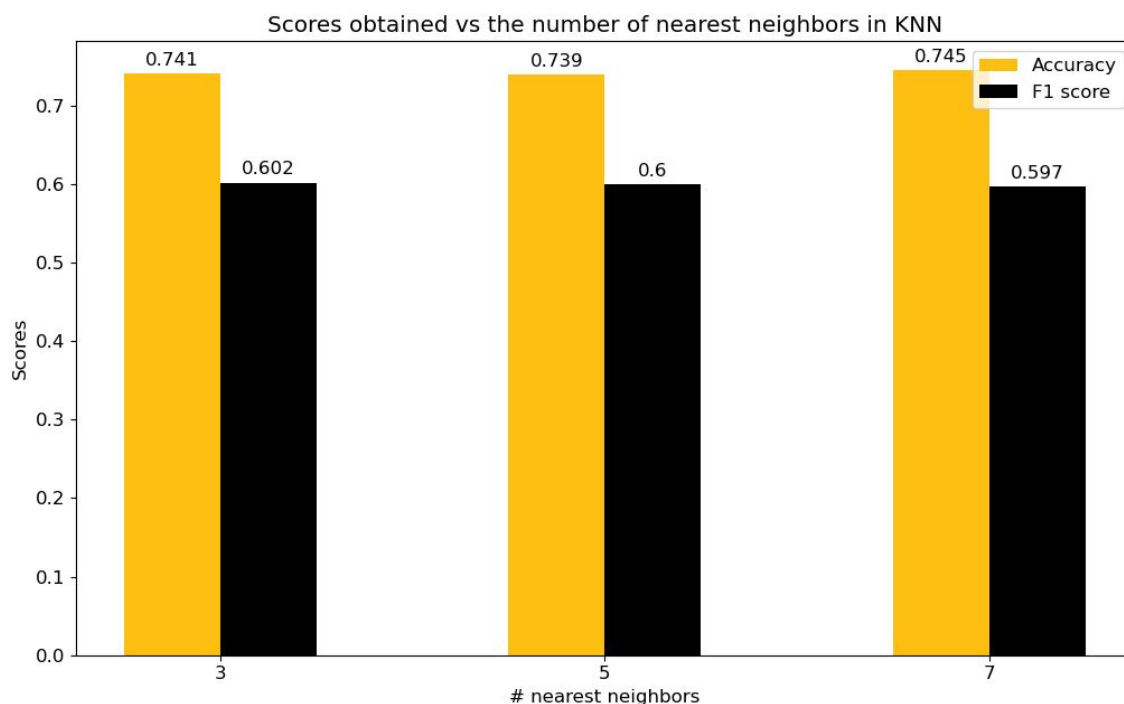


Figura 3. Resultados obtidos para distintos valores de “k” vizinhos mais próximos usando repeated cross validation com 10 folds e 10 repetições

a) 3 vizinhos mais próximos

Repetição	Acurácia média	F1-measure média
1	0.734	0.592
2	0.751	0.620
3	0.737	0.590
4	0.742	0.612
5	0.734	0.594
6	0.729	0.580
7	0.738	0.595
8	0.741	0.608
9	0.737	0.596
10	0.762	0.634

Acurácia média global (e desvio padrão): **0.741 (0.040)**

F1-measure média global: **0.602 (0.067)**

b) 5 vizinhos mais próximos:

Repetição	Acurácia média	F1-measure média
1	0.745	0.613
2	0.728	0.578
3	0.736	0.597
4	0.742	0.609
5	0.741	0.603
6	0.743	0.608
7	0.740	0.593
8	0.734	0.587
9	0.735	0.601
10	0.747	0.612

Acurácia média (e desvio padrão): **0.739 (0.050)**

F1-measure média: **0.600 (0.076)**

c) 7 vizinhos mais próximos

Repetição	Acurácia média	F1-measure média
1	0.751	0.596
2	0.730	0.564
3	0.747	0.600
4	0.741	0.593
5	0.743	0.600
6	0.742	0.588
7	0.750	0.611
8	0.745	0.600
9	0.743	0.601
10	0.754	0.614

Acurácia média (e desvio padrão): **0.745 (0.045)**

F1-measure média: **0.597 (0.074)**