

Relatório da atividade sobre Bernoulli Native Bayes de Aprendizado de Máquina Supervisionado

Vinicius Cordeiro Fonseca
20250082351
UFRN/IMD
07/08/2025

Sobre os dados utilizados: na primeira tarefa utilizamos o dataset de spam

1)

- Dividimos o conjunto em 3 subconjuntos disjuntos: treinamento(70%), validação(15%) e teste(15%)
- Testamos as métricas: minkowski, cityblock, cosine, euclidean e manhattan.

```
Modelo: minkowski
Precisão: 0.9151
Relatório de Classificação:
      precision  recall  f1-score  support
0         0.91     1.00     0.95     724
1         1.00     0.37     0.54     112

accuracy      0.92     836
macro avg     0.96     0.68     0.74     836
weighted avg  0.92     0.92     0.90     836

Matriz de Confusão:
[[724  0]
 [ 71 41]]

-----
Modelo: cityblock
Precisão: 0.9151
Relatório de Classificação:
      precision  recall  f1-score  support
0         0.91     1.00     0.95     724
1         1.00     0.37     0.54     112

accuracy      0.92     836
macro avg     0.96     0.68     0.74     836
weighted avg  0.92     0.92     0.90     836

Matriz de Confusão:
[[724  0]
 [ 71 41]]
```

```
Modelo: euclidean
Precisão: 0.9151
Relatório de Classificação:
      precision  recall  f1-score  support
0         0.91     1.00     0.95     724
1         1.00     0.37     0.54     112

accuracy      0.92     836
macro avg     0.96     0.68     0.74     836
weighted avg  0.92     0.92     0.90     836

Matriz de Confusão:
[[724  0]
 [ 71 41]]

-----
Modelo: manhattan
Precisão: 0.9151
Relatório de Classificação:
      precision  recall  f1-score  support
0         0.91     1.00     0.95     724
1         1.00     0.37     0.54     112

accuracy      0.92     836
macro avg     0.96     0.68     0.74     836
weighted avg  0.92     0.92     0.90     836

Matriz de Confusão:
[[724  0]
 [ 71 41]]
```

```
Modelo: cosine
Precisão: 0.9761
Relatório de Classificação:
      precision  recall  f1-score  support
0         0.97     1.00     0.99     724
1         0.99     0.83     0.90     112

accuracy      0.98     836
macro avg     0.98     0.91     0.94     836
weighted avg  0.98     0.98     0.98     836

Matriz de Confusão:
[[723  1]
 [ 19 93]]
```

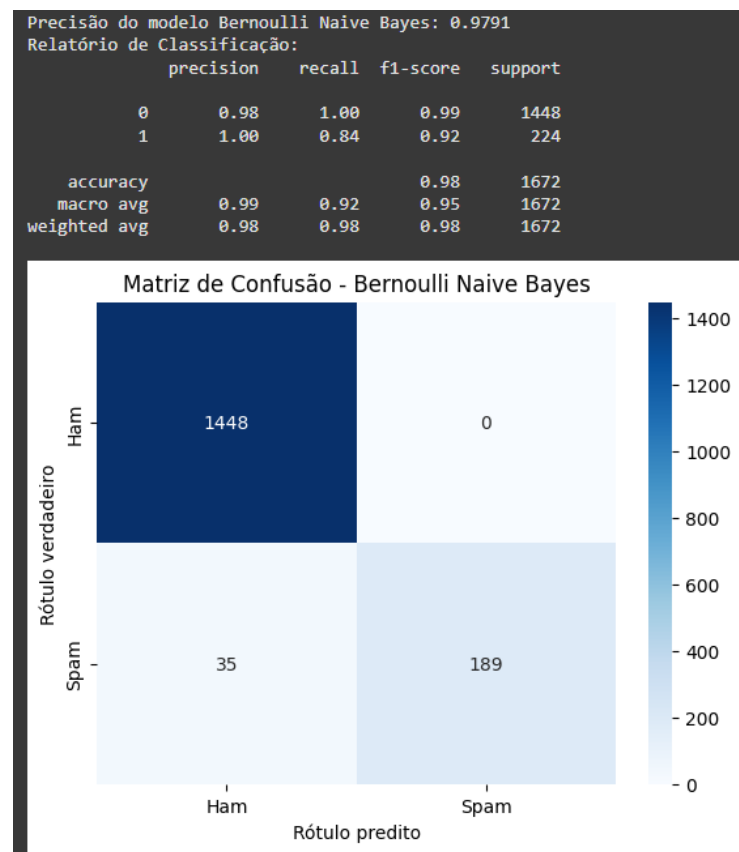
- Observamos que a métrica cosine tem melhor precisão e acurácia então treinamos o modelo com a junção dos conjuntos: treino + validação, e reportamos a performance obtida no conjunto teste como mostramos a seguir:

```
Modelo: cosine
Precisão: 0.8792
Relatório de Classificação:
      precision  recall  f1-score  support
0         0.95     0.91     0.93     724
1         0.54     0.71     0.61     112

accuracy      0.88     836
macro avg     0.74     0.81     0.77     836
weighted avg  0.90     0.88     0.89     836

Matriz de Confusão:
[[656 68]
 [ 33 79]]
```

- Também utilizamos o modelo Bernoulli Naive Bayes nesse conjunto de dados (treino+validação) e obtivemos os resultados a seguir:

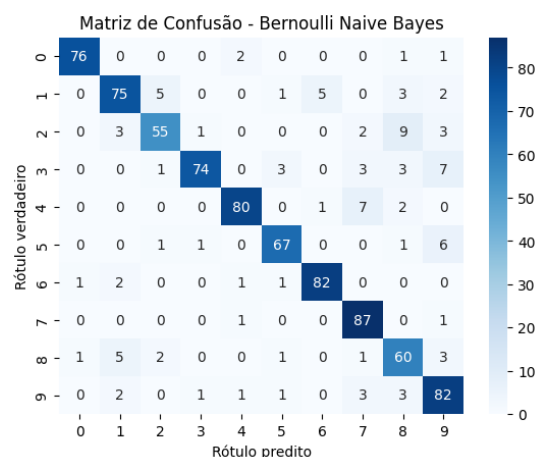


2)

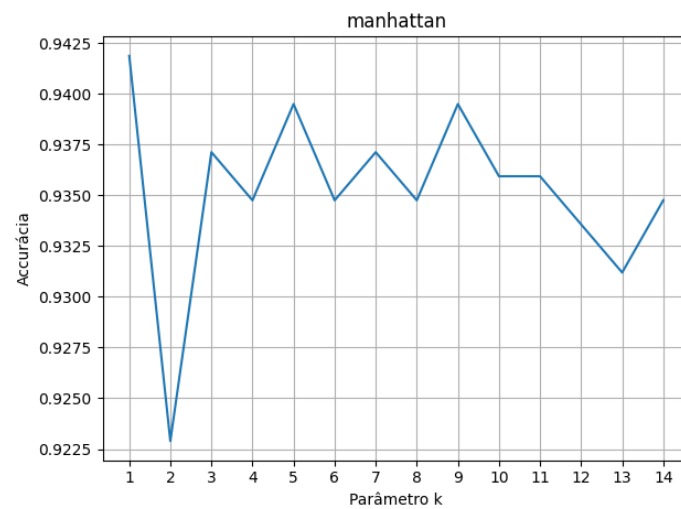
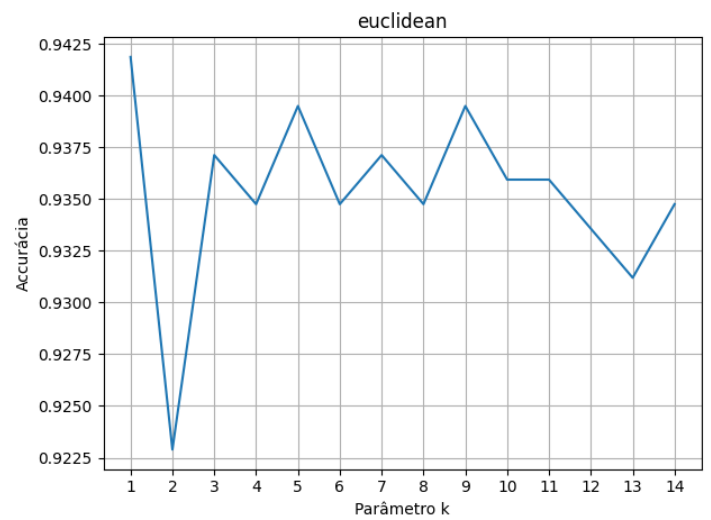
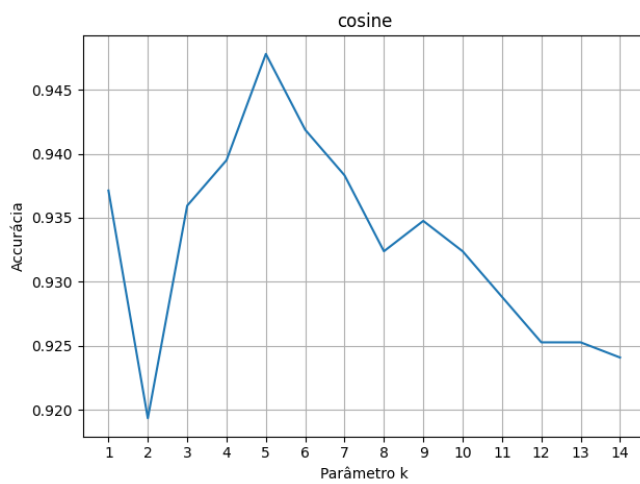
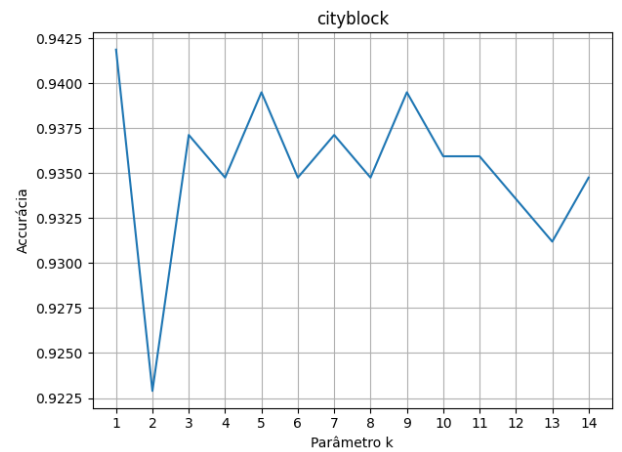
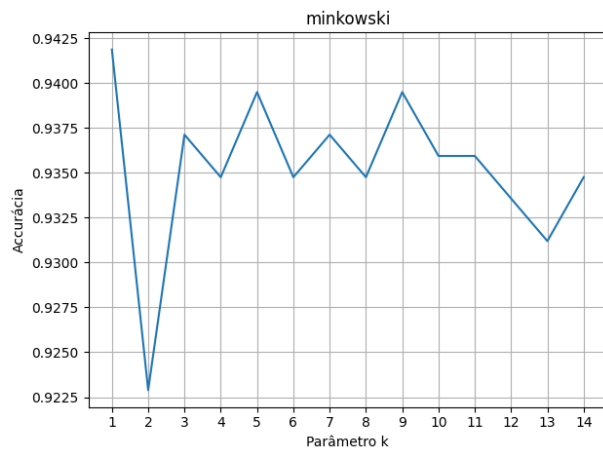
- Carregamos o conjunto de dados Optical Recognition of Handwritten Digits(ORHD) do UCI Machine Learning.
- Transformamos as características dos dados em valores binários.
- Dividimos o conjunto de dados em 3 subconjuntos disjuntos: treinamento(70%), validação(15%) e teste(15%). Treinamos um modelo Naive Bayes (Bernoulli) com o conjunto treino + validação e calculamos sua acurácia e matriz de confusão como mostramos abaixo:

Precisão do modelo Bernoulli Naive Bayes: 0.8754
Relatório de Classificação:

	precision	recall	f1-score	support
0	0.97	0.95	0.96	80
1	0.86	0.82	0.84	91
2	0.86	0.75	0.80	73
3	0.96	0.81	0.88	91
4	0.94	0.89	0.91	90
5	0.91	0.88	0.89	76
6	0.93	0.94	0.94	87
7	0.84	0.98	0.91	89
8	0.73	0.82	0.77	73
9	0.78	0.88	0.83	93
accuracy			0.88	843
macro avg	0.88	0.87	0.87	843
weighted avg	0.88	0.88	0.88	843



- Então buscamos as melhores métricas e valores de vizinhos próximos para treinar o KNN e comparar sua performance com o Bernoulli Naive Bayes:



- Observamos nos gráficos que a acurácia é maior na métrica cosine e com K=5, então treinamos nosso KNN com esses parâmetros e obtivemos o seguinte resultado:

```
Modelo: cosine
Precisão: 0.9549
Relatório de Classificação:
      precision    recall  f1-score   support

     0         1.00      0.99      0.99         80
     1         0.88      0.99      0.93         91
     2         0.97      0.97      0.97         73
     3         0.98      0.93      0.96         91
     4         0.98      0.98      0.98         90
     5         0.95      0.93      0.94         76
     6         0.99      0.98      0.98         87
     7         1.00      0.99      0.99         89
     8         0.90      0.89      0.90         73
     9         0.91      0.89      0.90         93

 accuracy          0.95         843
 macro avg         0.96         0.95         0.95         843
 weighted avg      0.96         0.95         0.96         843

Matriz de Confusão:
[[79  1  0  0  0  0  0  0  0  0]
 [ 0 90  0  0  0  0  0  0  1  0]
 [ 0  1 71  0  0  0  0  0  0  1]
 [ 0  0  1 85  0  1  0  0  2  2]
 [ 0  1  0  0 88  0  1  0  0  0]
 [ 0  0  0  0  1 71  0  0  0  4]
 [ 0  1  0  0  0  1 85  0  0  0]
 [ 0  1  0  0  0  0  0 88  0  0]
 [ 0  4  1  0  1  1  0  0 65  1]
 [ 0  3  0  2  0  1  0  0  4 83]]
```