

LAB2: Comparação de Classificadores

Alceu Eilert Nascimento

Table of contents

1	Introdução	1
2	Scripts e Metodologia	2
3	Em qual ponto o tamanho da base de treinamento deixa de ser relevante?	39
4	Qual é o classificador que tem o melhor desempenho com poucos dados (< 1000 exemplos)?	40
5	Qual é o classificador que tem melhor desempenho com todos os dados?	40
6	Qual é o classificador mais rápido para classificar os 58.000 exemplos de teste?	40
7	Os erros são os mesmos para todos os classificadores quando todos eles utilizam toda a base de treinamento?	40
8	Conclusão sobre as Matrizes de Confusão:	41

1 Introdução

Neste trabalho, utilizamos os classificadores K-Nearest Neighbors (KNN), Naïve Bayes (Bernoulli, Gaussian e Multinomial), Linear Discriminant Analysis (LDA) e Regressão Logística para comparar seus desempenhos com uma base de treinamento crescente.

A base de treinamento começou com 1000 exemplos e aumentou até 20.000 exemplos, em blocos de 1.000.

As perguntas principais abordadas são:

- Qual classificador tem o melhor desempenho com menos de 1000 exemplos?
- Qual classificador tem o melhor desempenho com todos os dados?

- Qual é o classificador mais rápido para classificar os 58.000 exemplos de teste?
- Os erros são os mesmos para todos os classificadores ao utilizar toda a base de treinamento?

2 Scripts e Metodologia

Para o treinamento e avaliação dos classificadores, utilizamos o código de Python dado pelo professo (lab2.py) com algumas adaptações:

```
import sys
import numpy as np
import time
import matplotlib.pyplot as plt

from sklearn import linear_model
from sklearn.neighbors import KNeighborsClassifier
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis

from sklearn.metrics import confusion_matrix
from sklearn.datasets import load_svmlight_file

from sklearn.naive_bayes import GaussianNB, BernoulliNB, MultinomialNB

def evaluate_classifiers(X_train, y_train, X_test, y_test, batchsize=1000):
    classifiers = {
        'KNN': KNeighborsClassifier(n_neighbors=3),
        'Logistic Regression': linear_model.LogisticRegression(),
        'Linear Discriminant Analysis': LinearDiscriminantAnalysis(),
        'GaussianNB': GaussianNB(),
        'BernoulliNB': BernoulliNB(),
        'MultinomialNB': MultinomialNB(),
    }

    results = {name: [] for name in classifiers}
    times = {} # Dicionário para armazenar o tempo de classificação dos 58k exemplos de teste

    for size in range(batchsize, X_train.shape[0] + 1, batchsize):
        X_train_subset = X_train[:size].toarray()
        y_train_subset = y_train[:size]
```

```

    print(f"Tamanho da base de treinamento: {size}")
    for name, clf in classifiers.items():
        start_time = time.time()
        clf.fit(X_train_subset, y_train_subset)
        score = clf.score(X_test.toarray(), y_test)
        elapsed_time = time.time() - start_time
        print(f"Classificador: {name} | Acurácia: {score:.4f} | Tempo: {elapsed_time:.2f}")
        results[name].append(score)

    # Predição dos dados de teste e tempo de classificação para 58.000 exemplos
    start_pred_time = time.time()
    y_pred = clf.predict(X_test.toarray())
    pred_time = time.time() - start_pred_time

    # Imprime a matriz de confusão para o classificador atual
    print(f"Matriz de Confusão para {name}:")
    cm = confusion_matrix(y_test, y_pred)
    print(cm)

    # Armazena o tempo de classificação no dicionário
    if size == X_train.shape[0]: # Apenas para o último bloco de treinamento (todos
        times[name] = pred_time

    return results, times

def plot_results(results, batchsize, total_samples):
    """
    Função para plotar os resultados de acurácia em função do tamanho da base de treinamento
    """
    training_sizes = list(range(batchsize, total_samples + 1, batchsize))

    # Plotar o gráfico para cada classificador
    plt.figure(figsize=(10, 6))

    for name, scores in results.items():
        plt.plot(training_sizes, scores, label=name)

    # Configurações do gráfico
    plt.title('Desempenho dos Classificadores em Função do Tamanho da Base de Treinamento')
    plt.xlabel('Tamanho da Base de Treinamento (número de exemplos)')
    plt.ylabel('Acurácia na Base de Testes')
    plt.legend(loc='lower right')

```

```

plt.grid(True)
plt.show()

def print_classification_times(times):
    """
    Função para exibir os tempos de classificação em uma tabela.
    """
    print("\nTabela de Tempos de Classificação (58.000 exemplos de teste):")
    print("{:<30} {:<15}".format('Classificador', 'Tempo (segundos)'))
    print("-" * 45)
    for classifier, time_taken in times.items():
        print("{:<30} {:<15.6f}".format(classifier, time_taken))

if __name__ == "__main__":
    # Carregar os dados
    print("Carregando dados...")
    X_train, y_train = load_svmlight_file('train.txt')
    X_test, y_test = load_svmlight_file('test.txt')

    # Avaliar os classificadores e coletar tempos de classificação
    results, times = evaluate_classifiers(X_train, y_train, X_test, y_test)

    # Plotar os resultados
    plot_results(results, batchsize=1000, total_samples=X_train.shape[0])

    # Exibir a tabela de tempos de classificação
    print_classification_times(times)

```

Carregando dados...

Tamanho da base de treinamento: 1000

Classificador: KNN | Acurácia: 0.7918 | Tempo: 1.11 segs.

Matriz de Confusão para KNN:

```

[[5279    2   40   51   13   22  122    0   20   11]
 [  48 3155 1372  142   71   14  583   95  455  720]
 [  18    5 5460  189   10    3   37  136   22    8]
 [  12    3  526 4534   47   62    3  421   12  199]
 [  22    3   68    2 5156    0  166   21   12  272]
 [  12   11  236  396  145 4298   71   49  118  203]
 [  75   28   18    1   26  109 5484    0  117    0]
 [   0   16   92   54  227    2    0 4748   22  936]
 [ 604   30  370   57  192   75  152   71 3120 1024]
 [  36   14   44  157  174    5    4  153   26 5200]]

```

Classificador: Logistic Regression | Acurácia: 0.7291 | Tempo: 0.09 segs.

Matriz de Confusão para Logistic Regression:

```
[[4982    5  218   42   44   21  218    3   20    7]
 [  12 2882  733  257  456   19 1097   42  548  609]
 [   29   25 5498   83   18    4   51  155   16    9]
 [    9    6 1096 3899  190   54   15  334    4  212]
 [   32    6   39    3 5306    0  140   15    8  173]
 [   17   11  387  835  532 3268  222  123   52   92]
 [   87   36   68    1   61  161 5349    0   95    0]
 [    0   62   41   42  175    2    0 4873   23  879]
 [  907   63  334   48  332   83  462  127 1976 1363]
 [   62   30   22  121  671    0    3  162   17 4725]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.7873 | Tempo: 0.08 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5151  112   16   49   29  108   33   14    9   39]
 [    1 4155 1046  133  156  178  338  254  321   73]
 [   29   50 5367   23    4    0   13  384   18    0]
 [    2   53   56 4393    1  112   10 1103   72   17]
 [   65  401  208    0 4622    2   68   17   25  314]
 [   11   84   10  566    4 4633   63   58   81   29]
 [  190  952   56  110   72   65 4295    5  109    4]
 [    0 1142   59   11   40   10    0 4773   36   26]
 [  188  132   74   62   48   49  349   79 3630 1084]
 [  121   81   58  147   19   41   48  107   37 5154]]
```

Classificador: GaussianNB | Acurácia: 0.6744 | Tempo: 0.73 segs.

Matriz de Confusão para GaussianNB:

```
[[4157    1  890   85   11   98   34   37  246    1]
 [   16 2901 2288  130  119   36  596  158  302  109]
 [    5    4 5182  616   29    5    4   34    3    6]
 [    2    1  752 4336   47   27    0  646    1    7]
 [   12    0 1933   35 3251   24   16   96   62  293]
 [    3    7 1325  100    9 3856   12   39   95   93]
 [    7    1 1689    3    0  256 3654    4  244    0]
 [    2   12  613  261   98    2    0 4934   29  146]
 [  361    0 1687  273   66   88   27   16 3136   41]
 [   13    0  691  178  539   46    0   82  123 4141]]
```

Classificador: BernoulliNB | Acurácia: 0.7857 | Tempo: 0.16 segs.

Matriz de Confusão para BernoulliNB:

```
[[4616    1  411   31  155   88  180    7   58   13]
 [  895 3020 1835   18   26   27  307   54  152  321]
 [    6    9 5655   67   11    2   36   87   10    5]
 [    3    5  769 4681   11  140    1  173    4   32]
 [   76    1  182    0 4931    2  167   12   19  332]]
```

```
[ 4 26 32 169 57 5117 37 17 42 38]
[ 140 17 204 0 159 146 5115 0 77 0]
[ 0 15 601 51 211 4 0 5126 20 69]
[ 897 16 523 4 329 92 148 29 3226 431]
[ 140 8 135 79 637 56 21 95 53 4589]]
```

Classificador: MultinomialNB | Acurácia: 0.7369 | Tempo: 0.05 segs.

Matriz de Confusão para MultinomialNB:

```
[[5006 8 53 63 92 4 297 1 31 5]
 [ 128 3036 544 59 110 2 1489 30 510 747]
 [ 46 41 5440 92 6 2 50 173 28 10]
 [ 19 13 1494 3516 18 58 13 459 13 216]
 [ 74 17 7 0 5133 0 213 7 20 251]
 [ 17 22 270 886 160 3655 242 73 91 123]
 [ 191 79 1 0 40 87 5310 0 150 0]
 [ 0 98 78 11 118 0 1 4977 36 778]
 [1257 59 121 5 144 24 473 57 2216 1339]
 [ 145 37 20 80 465 1 19 74 45 4927]]
```

Tamanho da base de treinamento: 2000

Classificador: KNN | Acurácia: 0.8668 | Tempo: 0.48 segs.

Matriz de Confusão para KNN:

```
[[5334 2 14 49 14 2 111 0 29 5]
 [ 15 5242 204 334 53 9 303 94 157 244]
 [ 16 10 5449 267 5 1 32 92 14 2]
 [ 6 4 72 5389 2 45 4 233 31 33]
 [ 19 9 44 25 5159 0 154 16 7 289]
 [ 9 8 26 1009 67 4157 64 32 81 86]
 [ 35 24 4 6 17 76 5604 0 92 0]
 [ 2 17 73 178 123 2 2 4870 14 816]
 [ 196 31 51 248 80 43 94 51 4553 348]
 [ 29 22 16 266 187 2 7 176 28 5080]]
```

Classificador: Logistic Regression | Acurácia: 0.7822 | Tempo: 0.09 segs.

Matriz de Confusão para Logistic Regression:

```
[[4968 6 78 103 32 15 282 5 66 5]
 [ 6 3275 193 445 287 68 1424 58 254 645]
 [ 26 23 5368 137 15 3 96 186 24 10]
 [ 8 8 116 5271 55 43 15 226 7 70]
 [ 42 6 18 7 5276 0 177 21 4 171]
 [ 13 11 64 1448 281 3328 213 82 40 59]
 [ 71 35 5 12 30 194 5420 0 91 0]
 [ 0 64 28 81 120 5 0 5164 5 630]
 [ 535 66 92 85 221 66 472 144 3043 971]
 [ 75 29 6 178 552 1 11 199 5 4757]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.8733 | Tempo: 0.09 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5268  37  11  22  49  18  61  9  81  4]
 [  1 5487 226 152  3 107 164 386 76 53]
 [ 19  48 5298 21  1  6 12 463 17  3]
 [  3  32  46 4971  0 38 10 681 24 14]
 [ 28 507  67  0 4544  2 78 27 37 432]
 [ 11  25  8 379  3 4920 68 38 62 25]
 [116  88 44 116 62 220 5094  0 115  3]
 [  0 131 51  4 17  1  0 5829 27 37]
 [125 110 45 24 41 54 182 63 4483 568]
 [ 32  62  5 151 21 37 18 147 20 5320]]
```

Classificador: GaussianNB | Acurácia: 0.7820 | Tempo: 0.64 segs.

Matriz de Confusão para GaussianNB:

```
[[4474  1 563 128  7 39 50 20 276  2]
 [  2 3948 1327 565 100 26 251 222 113 101]
 [  6  14 4495 1157  5  7 10 144 20 30]
 [  6  1 205 5199 19 18  0 336  8 27]
 [  3  2 765  47 3974 28 71 60 104 668]
 [  6 13 420 96 56 4600 33 20 102 193]
 [  8  3 546  6  5 157 4575  0 557  1]
 [  0 23 370 383 41  0  0 5221 12 47]
 [307  2 256 577 37 71 34 15 4310 86]
 [ 12  2 103 326 116 16  3 50 119 5066]]
```

Classificador: BernoulliNB | Acurácia: 0.8222 | Tempo: 0.17 segs.

Matriz de Confusão para BernoulliNB:

```
[[4595  1 309 41 215 49 224  8 102 16]
 [699 3467 1429 26 19 36 206 83 218 472]
 [ 10  14 5563 68 11  3 59 134 17  9]
 [  4  4 236 5286  6 60  1 176 12 34]
 [ 90  2 115  0 4951  1 172 13 28 350]
 [  3 23 18 333 87 4939 45 12 40 39]
 [131 32 49  1 179 82 5305  0 79  0]
 [  0 18 262 95 153  4  0 5465 17 83]
 [679 20 220 11 247 56 141 24 3978 319]
 [145 11 54 102 615 55 27 97 35 4672]]
```

Classificador: MultinomialNB | Acurácia: 0.7814 | Tempo: 0.05 segs.

Matriz de Confusão para MultinomialNB:

```
[[4962  7 27 103 89  2 321  2 42  5]
 [ 82 3276 327 115 110 16 1581 88 300 760]
 [ 28  31 5472 37  5  3 79 201 23  9]
 [ 14 13 131 5037  5 46  9 481 14 69]
 [ 84 16  6  0 5174  0 220 11  9 202]
 [ 12 24 16 1674 87 3383 197 44 37 65]]
```

```
[ 107  81  0  0  39 112 5377  0 142  0]
[  0  91 54 30 130  5  1 5487 16 283]
[ 866 68 64  4 154 27 484 75 2843 1110]
[ 162 39  8 106 537  1 23 113 11 4813]]
```

Tamanho da base de treinamento: 3000

Classificador: KNN | Acurácia: 0.8901 | Tempo: 0.71 segs.

Matriz de Confusão para KNN:

```
[[5367  3  7 34 17 16 74  1 38  3]
 [ 16 5756 186 159 44  6 208 66 140 74]
 [ 18  20 5345 375  8  2 24 79 16  1]
 [  5  6 29 5443  5 57  4 210 30 30]
 [ 17 19 37  3 5200  0 148 28  6 264]
 [ 12 13 14 645 62 4550 50 35 79 79]
 [ 39 29  4  1 17 104 5604  0 60  0]
 [  2 26 71 179 120  1  1 5036 16 645]
 [ 136 90 44 140 80 36 88 43 4813 225]
 [ 12 26 18 215 175  7  6 238 29 5087]]
```

Classificador: Logistic Regression | Acurácia: 0.8067 | Tempo: 0.12 segs.

Matriz de Confusão para Logistic Regression:

```
[[4966  8 41 109 34 35 280  7 77  3]
 [  2 3527 204 340 397 97 1250 83 165 590]
 [ 19  20 5332 124 20  3 108 223 33  6]
 [  8  7 55 5260 63 62 11 289  8 56]
 [ 37  6 16  1 5283  1 162 30  4 182]
 [ 12 11 28 1108 278 3786 156 99 33 28]
 [ 67 31  0 10 30 231 5414  0 75  0]
 [  0 64 22 78 149  7  0 5320  5 452]
 [ 352 57 66 60 217 79 317 176 3655 716]
 [ 58 28  7 165 532  1 10 239  5 4768]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.8935 | Tempo: 0.10 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5228 21 23 32 44  3 76 12 120  1]
 [  1 5678 224 125  5 89 133 289 79 32]
 [ 16 43 5383 17  2  1 22 385 18  1]
 [  5 37 35 5032  1 19  3 653 20 14]
 [ 12 317 45  0 4899  0 73 21 19 336]
 [  9 32  8 262  4 5050 57 36 55 26]
 [ 98 52 29 15 133 81 5347  0 100  3]
 [  0 139 40  4 32  1  0 5832 23 26]
 [ 92 115 38 14 46 80 117 52 4638 503]
 [ 12 57  6 135 46 33 18 165 25 5316]]
```

Classificador: GaussianNB | Acurácia: 0.8384 | Tempo: 0.63 segs.

Matriz de Confusão para GaussianNB:


```

[[4733    1  271   66    3    9  149    1  326    1]
 [   1 4883  770  262  119   12  215  207  130   56]
 [  16   29 5032  561   15    0   39  123   59   14]
 [   7    1  372 5058    3   11    3  301   42   21]
 [   7    3  262    4 4722   19  133   70  101  401]
 [   4   33  388   92   48 4579   61    6   62  266]
 [   7    4  111    5    7  134 5402    0  187    1]
 [   4   44  342  317  41    0    1 5269   61   18]
 [ 225    2  382  272   38   69  130   12 4470   95]
 [   8    3  228  130  137   14   11  151  109 5022]]

```

Classificador: BernoulliNB | Acurácia: 0.8285 | Tempo: 0.17 segs.

Matriz de Confusão para BernoulliNB:

```

[[4582    1  220   44  225   88  273   18   95   14]
 [ 566 3602 1177   16   25   34  193   91  347  604]
 [   8   16 5480   50   12    4  122  172   16    8]
 [   4    5  236 5200   15   79    2  239   11   28]
 [  72    2   99    0 5019    1  154   23   25  327]
 [   3   24   15  248   92 5032   48   22   30   25]
 [ 114   34   14    0  133   92 5387    0   84    0]
 [   0   21  197   47  158    3    0 5615   13   43]
 [ 466   21  202    7  312   70  168   44 4084  321]
 [ 107   11   40   84  723   54   22  145   41 4586]]

```

Classificador: MultinomialNB | Acurácia: 0.7954 | Tempo: 0.05 segs.

Matriz de Confusão para MultinomialNB:

```

[[4932    5   31  124  104    2  311    5   42    4]
 [  42 3319  373   81  158   27 1488  123  332  712]
 [  21   31 5441   61    5    2   95  205   22    5]
 [  13   12   76 5137    8   47   10  451   15   50]
 [  67   13    9    0 5206    0  200   15   10  202]
 [  11   21   10 1621   94 3482  182   52   28   38]
 [  75   77    0    0   40  154 5365    0  147    0]
 [   0   90   47   39  122    8    2 5604   20  165]
 [ 665   51   62    2  169   42  439   79 3256  930]
 [  87   39    9  103  511    1   18  123   18 4904]]

```

Tamanho da base de treinamento: 4000

Classificador: KNN | Acurácia: 0.8938 | Tempo: 0.91 segs.

Matriz de Confusão para KNN:

```

[[5323    3    7   89   21    3   82    3   29    0]
 [  18 5781  179  131   32   10  216   92  149   47]
 [  19   21 5442  270    5    3   29   72   25    2]
 [   4    7   24 5518    6   44    3  165   21   27]
 [  16   20   38    3 5152    0  148   34   13  298]
 [  13   11   21  626   49 4570   46   39   88   76]

```

```
[ 38  27  2  2  18  99 5609  0  63  0]
[  1  27  60 168 174  1  1 5093  59 513]
[ 148  77  54 146  74  30  81  42 4862 181]
[  12  25  14 217 156  7  5  275  32 5070]]
```

Classificador: Logistic Regression | Acurácia: 0.8141 | Tempo: 0.15 segs.

Matriz de Confusão para Logistic Regression:

```
[[4969  7  43  93  37  47 280  11  71  2]
 [  2 3504 262 304 353 112 1268 115 166 569]
 [ 15  19 5380  94  17  3 112 199  44  5]
 [  7  7  70 5220  64  56  11 324  8  52]
 [ 37  6  18  1 5258  1 154  43  5 199]
 [ 11  9  26 967 249 3961 139 101  41  35]
 [ 66 24  0  6  41 215 5429  0  77  0]
 [  0 65 25  73 133  6  0 5358 12 425]
 [ 337 60 71  55 191  71 288 193 3857 572]
 [ 56 29  8 161 430  1  8 302  8 4810]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.8854 | Tempo: 0.10 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5080 16 16 46 43  6  81 15 256  1]
 [  0 5503 355 108  4  87 163 313 69 53]
 [ 13 43 5432 10  0  1  33 329 27  0]
 [  4 36 41 5022  1 27  3 654 18 13]
 [ 15 290 52  0 4804  1 88 28 37 407]
 [ 10 32 12 285  3 4994 72 49 58 24]
 [ 91 48 31 13 148 72 5363  0 89  3]
 [  0 131 44  3 28  1  0 5831 34 25]
 [ 115 98 36 11 48 64 150 65 4655 453]
 [ 19 49 10 138 37 14 23 251 33 5239]]
```

Classificador: GaussianNB | Acurácia: 0.8449 | Tempo: 0.64 segs.

Matriz de Confusão para GaussianNB:

```
[[4421  1 221 103  34  20 109  1 649  1]
 [  4 4773 830 257 104 19 194 232 174 68]
 [ 11  21 5092 577 16  0 48 70 50  3]
 [  4  1 485 5051  4 19  2 230  9 14]
 [  2  3 206 13 4921  6 77 50 113 331]
 [  4 12 139 94 25 4879 67  3 78 238]
 [  6  3 113  5 24 123 5348  0 236  0]
 [  5 32 378 339 34  2  1 5219 71 16]
 [ 196  2 182 196 29 48 92 11 4853 86]
 [ 12  3 181 176 133 14  8 159 134 4993]]
```

Classificador: BernoulliNB | Acurácia: 0.8302 | Tempo: 0.17 segs.

Matriz de Confusão para BernoulliNB:

```
[[4580  1 193  56 232  96 269  24  94 15]
```

```
[ 762 3566 1018 15 14 37 255 85 342 561]
[ 8 15 5495 49 8 10 125 147 23 8]
[ 4 4 207 5227 10 80 2 245 11 29]
[ 63 1 113 0 4968 1 147 26 37 366]
[ 3 21 18 283 85 5002 46 18 33 30]
[ 108 30 14 0 140 96 5376 0 94 0]
[ 0 20 189 48 132 6 0 5639 23 40]
[ 506 22 187 10 262 68 153 46 4146 295]
[ 97 14 41 84 585 53 22 163 64 4690]]
```

Classificador: MultinomialNB | Acurácia: 0.8015 | Tempo: 0.05 segs.

Matriz de Confusão para MultinomialNB:

```
[[4910 5 29 138 100 2 327 9 34 6]
[ 42 3311 422 65 104 30 1436 131 332 782]
[ 17 29 5485 44 5 2 98 179 22 7]
[ 10 11 80 5091 7 46 9 494 14 57]
[ 68 12 13 0 5159 0 190 19 13 248]
[ 11 20 9 1614 83 3483 176 61 32 50]
[ 63 71 0 0 44 152 5382 0 146 0]
[ 0 86 55 38 109 7 2 5649 33 118]
[ 610 48 60 5 154 33 422 83 3493 787]
[ 71 37 9 99 369 1 17 130 37 5043]]
```

Tamanho da base de treinamento: 5000

Classificador: KNN | Acurácia: 0.8977 | Tempo: 1.06 segs.

Matriz de Confusão para KNN:

```
[[5423 5 0 30 9 2 65 2 24 0]
[ 5 5751 188 123 43 7 184 72 184 98]
[ 18 16 5427 306 4 2 26 62 24 3]
[ 6 7 23 5564 8 43 4 112 20 32]
[ 14 11 37 3 5217 0 152 21 14 253]
[ 10 10 9 619 46 4538 59 41 128 79]
[ 41 17 3 1 16 94 5600 0 86 0]
[ 1 23 59 172 182 1 1 5065 46 547]
[ 152 60 49 157 72 26 87 42 4871 179]
[ 12 15 14 214 145 2 8 180 30 5193]]
```

Classificador: Logistic Regression | Acurácia: 0.8193 | Tempo: 0.18 segs.

Matriz de Confusão para Logistic Regression:

```
[[5016 4 32 94 30 29 268 12 73 2]
[ 2 3526 310 294 362 85 1143 86 208 639]
[ 19 18 5396 81 18 3 116 190 42 5]
[ 8 7 63 5289 65 49 12 261 9 56]
[ 37 6 19 2 5241 1 156 39 6 215]
[ 13 9 26 949 257 3970 142 87 45 41]
[ 76 21 0 4 39 177 5417 0 124 0]]
```

```
[ 0 62 26 74 138 4 0 5343 9 441]
[ 353 50 72 56 188 58 247 162 3916 593]
[ 55 30 8 152 387 0 7 233 7 4934]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.8905 | Tempo: 0.11 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5240 11 14 43 40 1 89 17 104 1]
[ 0 5413 475 107 9 81 133 304 68 65]
[ 19 48 5422 9 1 0 29 327 32 1]
[ 4 30 41 5118 0 27 5 557 20 17]
[ 21 177 45 0 4895 0 88 9 60 427]
[ 10 20 17 289 3 4978 80 40 72 30]
[ 105 38 31 20 155 77 5311 0 120 1]
[ 0 119 44 1 35 1 0 5836 35 26]
[ 132 72 36 9 68 44 129 70 4650 485]
[ 26 39 11 135 58 8 20 111 42 5363]]
```

Classificador: GaussianNB | Acurácia: 0.8551 | Tempo: 0.67 segs.

Matriz de Confusão para GaussianNB:

```
[[5113 1 35 65 24 11 79 0 231 1]
[ 4 4570 857 327 108 19 135 366 166 103]
[ 14 19 5008 676 12 1 22 85 46 5]
[ 5 1 379 5178 3 18 0 207 8 20]
[ 13 2 118 13 4979 4 59 60 123 351]
[ 4 14 114 93 12 4877 65 3 89 268]
[ 11 5 129 5 23 116 5305 0 264 0]
[ 8 42 241 483 31 1 2 5215 54 20]
[ 250 1 148 245 27 25 72 10 4815 102]
[ 47 3 124 164 110 14 6 120 137 5088]]
```

Classificador: BernoulliNB | Acurácia: 0.8266 | Tempo: 0.17 segs.

Matriz de Confusão para BernoulliNB:

```
[[4593 1 169 72 228 82 274 24 101 16]
[ 809 3448 779 14 13 29 317 81 316 849]
[ 8 17 5445 52 8 2 174 148 24 10]
[ 4 6 192 5270 12 75 2 214 11 33]
[ 59 1 110 0 4952 1 155 24 41 379]
[ 3 19 19 295 90 4974 49 16 38 36]
[ 109 30 8 0 124 82 5384 0 121 0]
[ 0 22 185 63 146 6 0 5597 28 50]
[ 555 21 176 11 258 58 160 43 4061 352]
[ 83 15 31 81 565 48 21 145 71 4753]]
```

Classificador: MultinomialNB | Acurácia: 0.8017 | Tempo: 0.05 segs.

Matriz de Confusão para MultinomialNB:

```
[[4914 5 32 135 92 2 329 9 36 6]
[ 39 3271 437 71 116 24 1387 130 382 798]]
```

```
[ 16  27 5472  56  7  2 105 177  20  6]
[ 10  11  56 5227  5 42  8 401  11 48]
[ 56  11  13  0 5156  0 194  18  19 255]
[  9  20  8 1670  88 3438 174  52  33 47]
[ 60  67  0  0  39 133 5374  0 185  0]
[  0  85  57  45 113  6  2 5644  36 109]
[ 623 42  66  7 159  29 412  92 3436 829]
[ 58  33  9 104 344  1  18 115  48 5083]]
```

Tamanho da base de treinamento: 6000

Classificador: KNN | Acurácia: 0.9205 | Tempo: 1.35 segs.

Matriz de Confusão para KNN:

```
[[5436  5  1  22  6  1  57  3  29  0]
 [  2 5958 170 129  58  4  82  76 114 62]
 [ 20  14 5448 300  6  2  19  50  26  3]
 [  6  7  19 5595  6 36  3 109  20 18]
 [ 13  9  26  4 5309  0 157  25  13 166]
 [  7  6  8  578 36 4612  52  33 135 72]
 [ 42 17  3  0  11 101 5579  0 105  0]
 [  1 23  56 103  78  1  1 5732  21 81]
 [ 55 50  43 124  51 20  63  36 5075 178]
 [ 13 15  13 157 160  5  7 174  32 5237]]
```

Classificador: Logistic Regression | Acurácia: 0.8420 | Tempo: 0.22 segs.

Matriz de Confusão para Logistic Regression:

```
[[5023  6  33  96  29  28 243  9  92  1]
 [  1 3935 283 316 447  72 878 156 141 426]
 [ 21  18 5447  92  12  1  87 156  50  4]
 [  5  7  43 5387  42  40  6 221  13 55]
 [ 36  7  18  2 5268  1 157  46  7 180]
 [ 10  9  26 920 196 4069 114  74  57 64]
 [ 70 22  0  4  27 156 5421  0 158  0]
 [  0 56  25  76 148  3  0 5538  14 237]
 [186 54  59  57 160  42 163 145 4332 497]
 [ 54 29  8 141 372  0  7 238  6 4958]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9025 | Tempo: 0.12 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5260 14  13  23  36  1  75  22 115  1]
 [  0 5642 337  86  5  46  91 322  50 76]
 [ 20  38 5528  9  1  0  23 236  32  1]
 [  2  26  34 5224  0  24  3 463  23 20]
 [ 20 141  41  0 4934  0  92  12  60 422]
 [ 10  19  13 295  3 4960  62  47  89 41]
 [112 42  36  30 123  51 5296  0 166  2]
 [  0  66  40  2  31  1  0 5903  31 23]]
```

```
[ 86 66 34 5 57 28 83 71 4766 499]
[ 28 32 9 108 50 1 22 112 35 5416]]
Classificador: GaussianNB | Acurácia: 0.8694 | Tempo: 0.66 segs.
```

Matriz de Confusão para GaussianNB:

```
[[5088 1 29 51 23 11 60 0 296 1]
[ 3 4791 747 257 103 20 115 420 140 59]
[ 12 19 5042 661 13 1 18 79 41 2]
[ 4 1 337 5212 2 27 0 203 9 24]
[ 8 2 96 6 5143 3 52 57 104 251]
[ 2 9 69 93 5 4817 51 4 129 360]
[ 12 6 124 4 20 102 5338 0 252 0]
[ 3 30 187 340 24 0 1 5412 63 37]
[ 194 2 99 197 26 16 55 12 5014 80]
[ 36 3 63 171 185 6 3 72 147 5127]]
```

Classificador: BernoulliNB | Acurácia: 0.8383 | Tempo: 0.20 segs.

Matriz de Confusão para BernoulliNB:

```
[[4582 1 233 51 203 79 255 18 121 17]
[ 568 3668 1073 10 18 28 233 81 377 599]
[ 8 18 5585 50 6 0 85 106 22 8]
[ 4 5 223 5271 5 59 0 205 11 36]
[ 57 2 112 0 4956 1 148 24 49 373]
[ 3 18 25 282 79 4987 45 13 45 42]
[ 104 31 15 0 102 82 5392 0 132 0]
[ 0 21 182 43 115 6 0 5652 30 48]
[ 437 20 190 8 232 48 136 42 4285 297]
[ 82 15 32 70 553 47 19 132 78 4785]]
```

Classificador: MultinomialNB | Acurácia: 0.8122 | Tempo: 0.05 segs.

Matriz de Confusão para MultinomialNB:

```
[[4913 5 33 129 87 2 335 9 39 8]
[ 38 3402 579 71 129 27 1261 209 281 658]
[ 17 32 5537 55 6 2 61 154 21 3]
[ 8 11 62 5291 2 41 5 351 14 34]
[ 57 11 11 0 5184 0 182 25 20 232]
[ 7 19 8 1742 73 3399 146 50 38 57]
[ 56 68 0 0 35 132 5371 0 196 0]
[ 0 82 58 47 102 8 1 5691 30 78]
[ 526 49 72 8 145 31 326 96 3774 668]
[ 58 38 11 106 354 1 15 120 37 5073]]
```

Tamanho da base de treinamento: 7000

Classificador: KNN | Acurácia: 0.9269 | Tempo: 1.49 segs.

Matriz de Confusão para KNN:

```
[[5434 4 1 22 6 1 41 3 48 0]
[ 2 6021 156 115 59 6 69 72 95 60]]
```

```
[ 16 13 5492 245 3 1 14 76 26 2]
[ 5 8 24 5614 2 43 1 82 22 18]
[ 13 8 27 5 5328 0 144 24 27 146]
[ 7 6 7 586 18 4633 52 18 159 53]
[ 39 14 3 0 8 59 5662 0 73 0]
[ 2 20 37 127 61 1 0 5753 11 85]
[ 46 40 46 105 47 21 53 37 5145 155]
[ 12 11 21 126 164 5 8 136 52 5278]]
```

Classificador: Logistic Regression | Acurácia: 0.8622 | Tempo: 0.23 segs.

Matriz de Confusão para Logistic Regression:

```
[[5102 6 34 65 25 21 206 9 91 1]
[ 0 4385 244 290 548 66 559 219 157 187]
[ 22 19 5442 104 16 1 73 156 53 2]
[ 5 6 47 5416 39 40 3 199 18 46]
[ 38 6 22 2 5291 1 146 44 5 167]
[ 8 9 29 761 182 4285 90 54 69 52]
[ 78 27 1 1 37 118 5461 0 135 0]
[ 0 45 24 83 177 2 0 5593 15 158]
[ 145 61 51 53 132 43 91 117 4645 357]
[ 56 25 7 138 408 0 6 221 9 4943]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9124 | Tempo: 0.15 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5313 12 14 16 27 0 61 22 93 2]
[ 0 5713 274 92 7 52 71 341 45 60]
[ 21 40 5530 10 1 0 22 235 28 1]
[ 1 16 30 5326 0 26 2 371 25 22]
[ 22 85 44 0 5114 0 87 9 37 324]
[ 9 9 12 291 4 4988 59 40 87 40]
[ 107 46 36 29 120 43 5330 0 147 0]
[ 0 56 40 2 46 1 0 5910 22 20]
[ 84 54 36 5 65 33 67 68 4866 417]
[ 30 33 9 81 75 6 19 104 35 5421]]
```

Classificador: GaussianNB | Acurácia: 0.8742 | Tempo: 0.63 segs.

Matriz de Confusão para GaussianNB:

```
[[5193 1 6 41 8 1 42 0 267 1]
[ 0 4825 725 259 116 23 117 453 95 42]
[ 12 15 5115 586 12 1 17 83 47 0]
[ 1 1 299 5234 3 33 0 190 38 20]
[ 13 2 64 11 5197 1 32 53 108 241]
[ 10 6 39 91 66 4750 54 4 268 251]
[ 74 6 92 5 20 84 5327 0 250 0]
[ 1 24 152 335 25 0 1 5449 75 35]
[ 157 3 57 193 29 16 49 13 5104 74]]
```

[35 2 56 139 209 4 2 79 210 5077]]
Classificador: BernoulliNB | Acurácia: 0.8439 | Tempo: 0.18 segs.

Matriz de Confusão para BernoulliNB:

```
[[4582 1 237 43 199 82 243 14 149 10]
 [ 500 3763 1078 9 23 23 227 84 445 503]
 [ 7 19 5617 48 7 0 66 97 23 4]
 [ 3 4 220 5291 8 62 0 185 13 33]
 [ 59 2 106 0 4992 1 139 21 59 343]
 [ 3 15 23 233 95 5029 40 13 50 38]
 [ 107 31 16 0 111 66 5395 0 132 0]
 [ 0 20 176 60 130 4 0 5631 30 46]
 [ 371 20 182 8 236 47 105 38 4460 228]
 [ 86 13 35 61 605 50 16 124 93 4730]]
```

Classificador: MultinomialNB | Acurácia: 0.8209 | Tempo: 0.05 segs.

Matriz de Confusão para MultinomialNB:

```
[[4921 6 35 121 87 2 333 8 39 8]
 [ 25 3490 723 78 236 22 1073 298 244 466]
 [ 16 33 5532 68 6 2 51 154 25 1]
 [ 9 10 60 5342 3 41 2 304 16 32]
 [ 54 11 13 0 5238 0 165 21 24 196]
 [ 7 19 9 1707 78 3453 137 39 39 51]
 [ 53 75 0 0 40 120 5381 0 189 0]
 [ 0 82 52 57 114 10 1 5692 29 60]
 [ 485 52 77 10 143 31 235 92 4065 505]
 [ 66 38 14 106 396 1 13 116 34 5029]]
```

Tamanho da base de treinamento: 8000

Classificador: KNN | Acurácia: 0.9309 | Tempo: 1.70 segs.

Matriz de Confusão para KNN:

```
[[5451 3 1 16 5 2 37 3 41 1]
 [ 1 6032 185 116 63 4 44 69 74 67]
 [ 16 10 5556 197 3 1 14 69 20 2]
 [ 5 4 21 5634 1 50 1 63 19 21]
 [ 13 13 25 7 5316 0 141 22 17 168]
 [ 8 4 12 568 10 4692 47 14 111 73]
 [ 28 11 4 3 8 51 5701 0 52 0]
 [ 2 20 38 133 66 1 0 5729 10 98]
 [ 45 25 45 108 44 31 55 29 5122 191]
 [ 12 9 23 109 124 6 10 131 30 5359]]
```

Classificador: Logistic Regression | Acurácia: 0.8805 | Tempo: 0.23 segs.

Matriz de Confusão para Logistic Regression:

```
[[5208 5 31 53 20 9 160 9 64 1]
 [ 1 5114 158 287 288 60 360 115 93 179]
 [ 21 18 5490 99 11 1 65 129 52 2]]
```



```
[ 5 5 39 5493 22 41 4 149 15 46]
[ 37 8 26 2 5288 1 137 39 5 179]
[ 8 10 28 740 136 4359 87 42 66 63]
[ 74 26 0 1 27 103 5507 0 120 0]
[ 0 42 28 104 186 2 0 5534 15 186]
[ 146 52 59 63 115 48 102 99 4621 390]
[ 55 25 9 134 355 0 6 197 7 5025]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9208 | Tempo: 0.14 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5333 14 13 15 19 0 62 19 83 2]
[ 0 5940 218 97 8 34 44 246 26 42]
[ 21 38 5568 13 1 0 14 206 26 1]
[ 1 14 30 5405 1 20 2 302 22 22]
[ 21 90 51 0 5133 0 88 6 32 301]
[ 8 10 8 326 5 4970 63 34 71 44]
[ 81 42 38 34 77 34 5438 0 113 1]
[ 0 59 42 3 49 1 0 5894 25 24]
[ 81 59 42 6 56 28 70 56 4885 412]
[ 31 34 9 79 70 7 20 96 34 5433]]
```

Classificador: GaussianNB | Acurácia: 0.8825 | Tempo: 0.66 segs.

Matriz de Confusão para GaussianNB:

```
[[5209 1 9 31 3 1 51 0 254 1]
[ 1 5138 543 236 91 20 101 403 86 36]
[ 14 19 5170 540 10 1 15 71 46 2]
[ 1 1 268 5304 2 31 0 146 37 29]
[ 14 2 67 9 5204 1 31 50 111 233]
[ 10 6 31 93 55 4776 45 2 238 283]
[ 71 7 95 6 15 81 5309 0 274 0]
[ 1 33 149 398 22 0 1 5376 75 42]
[ 146 3 49 171 25 7 41 12 5167 74]
[ 29 3 58 135 195 5 2 64 221 5101]]
```

Classificador: BernoulliNB | Acurácia: 0.8503 | Tempo: 0.17 segs.

Matriz de Confusão para BernoulliNB:

```
[[4599 1 254 36 174 84 244 12 142 14]
[ 340 4095 925 11 22 27 206 82 312 635]
[ 8 17 5630 50 5 0 65 81 25 7]
[ 3 4 208 5328 4 56 0 166 13 37]
[ 62 2 113 0 4976 1 135 20 59 354]
[ 3 14 23 235 81 5034 40 10 49 50]
[ 108 31 19 0 97 70 5398 0 135 0]
[ 0 22 199 73 132 6 0 5576 36 53]
[ 365 19 186 9 228 43 103 29 4444 269]
[ 85 13 32 64 547 50 17 119 102 4784]]
```

Classificador: MultinomialNB | Acurácia: 0.8246 | Tempo: 0.05 segs.

Matriz de Confusão para MultinomialNB:

```
[[4939    6   35  119   80    2  329    7   31   12]
 [  19 3585  812   84  174   20  951  222  178  610]
 [  16   32 5544   74    5    2   51  139   24    1]
 [   8    9   60 5387    2   43    3  257   15   35]
 [  46   12   13    0 5227    0  158   22   25  219]
 [   7   19   11 1672   55 3498  129   27   40   81]
 [  49   72    0    0   37  119 5388    0  193    0]
 [   0   82   58   68  113   10    1 5659   36   70]
 [ 482   46   87   13  130   28  237   78 4022  572]
 [  56   38   13  101  325    2   12  108   50 5108]]
```

Tamanho da base de treinamento: 9000

Classificador: KNN | Acurácia: 0.9346 | Tempo: 1.82 segs.

Matriz de Confusão para KNN:

```
[[5469    3    1   16    5    2   25    2   36    1]
 [   0 6079  165  113   61    2   38   72   61   64]
 [  15   15 5593  167    3    1   14   59   19    2]
 [   4    3   25 5638    1   51    1   60   17   19]
 [  13   14   16    4 5313    0  136   21   16  189]
 [   9    4   11  584    7 4718   44   12   93   57]
 [  37   11    4    3    8   49 5700    0   46    0]
 [   1   27   42  129   69    1    0 5729    8   91]
 [  39   25   47   99   31   26   56   30 5166  176]
 [  17    9   23  108   83    7    8  126   29 5403]]
```

Classificador: Logistic Regression | Acurácia: 0.8886 | Tempo: 0.36 segs.

Matriz de Confusão para Logistic Regression:

```
[[5275    5   26   39   18    5  123    7   61    1]
 [   1 5237  137  283  275   58  294   98  103  169]
 [  22   19 5531   96   11    1   50  111   45    2]
 [   5    5   39 5537   19   41    2  120   15   36]
 [  37    8   25    1 5300    1  129   40    7  174]
 [   7   10   30  744  127 4386   76   29   68   62]
 [  84   32    1    1   23  106 5485    0  126    0]
 [   0   41   31  117  194    2    0 5507   16  189]
 [ 117   51   56   59  107   43   78   83 4777  324]
 [  56   24    8  151  304    0    5  172   14 5079]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9237 | Tempo: 0.20 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5352   13   11   16   20    0   50   17   79    2]
 [   0 5982  205   92    8   21   42  234   31   40]
 [  20   37 5595   12    1    0   11  186   25    1]
 [   1   12   31 5436    1   18    2  277   21   20]
```

```
[ 25  88  47   0 5143   0  87   7  31 294]
[  9  11   6 340   5 4970  56  26  69  47]
[ 84  46  37  28  74  37 5419   0 130   3]
[  0  60  48   4  52   1   0 5880  28  24]
[ 87  56  39   4  50  27  56  52 4958 366]
[ 35  31   9  88  61   6  18  96  32 5437]]
```

Classificador: GaussianNB | Acurácia: 0.8874 | Tempo: 0.66 segs.

Matriz de Confusão para GaussianNB:

```
[[5212   1  10  32   2   1  44   0 257   1]
 [   1 5175 565 213  88  16  82 383  98  34]
 [  14  22 5255 464   7   1  13  65  44   3]
 [   1   1 196 5391   2  29   0 139  35  25]
 [  14   2  66  10 5220   3  33  45 104 225]
 [   9   6  30 108  36 4896  46   2 187 219]
 [  78   7  94   6  14  84 5289   0 286   0]
 [   1  40 170 406  21   1   1 5341  73  43]
 [148   3  47 175  26  11  38  13 5157  77]
 [ 28   3  44 155 221   4   1  54 199 5104]]
```

Classificador: BernoulliNB | Acurácia: 0.8540 | Tempo: 0.18 segs.

Matriz de Confusão para BernoulliNB:

```
[[4602   1 263  29 171  83 242  12 142  15]
 [263 4274 914  11  24  27 194  77 332 539]
 [   8  19 5651  41   7   0  60  74  22   6]
 [   3   4 214 5347   5  48   0 147  13  38]
 [ 65   1 104   0 4983   1 133  17  58 360]
 [   3  16  19 291  68 4999  39  11  47  46]
 [111  33  22   0  88  76 5389   0 139   0]
 [   0  22 220  81 137   7   0 5535  36  59]
 [329  20 191  10 227  51  99  28 4488 252]
 [ 86  13  30  69 499  53  15 112 118 4818]]
```

Classificador: MultinomialNB | Acurácia: 0.8290 | Tempo: 0.05 segs.

Matriz de Confusão para MultinomialNB:

```
[[4956   6  33 111  76   2 329   7  29  11]
 [ 18 3631 831  82 194  16 901 212 171 599]
 [ 16  31 5564  73   4   2  45 129  21   3]
 [   8   9  59 5436   1  41   3 216  14  32]
 [ 42  13  12   0 5235   0 153  22  25 220]
 [   7  20  12 1683  57 3505 118  25  41  71]
 [ 46  76   0   0  34 117 5382   0 203   0]
 [   0  84  69  83 115   7   1 5629  36  73]
 [450  47  90  14 126  26 210  67 4142 523]
 [ 51  38  13 107 293   2  11 104  59 5135]]
```

Tamanho da base de treinamento: 10000

Classificador: KNN | Acurácia: 0.9382 | Tempo: 2.11 segs.

Matriz de Confusão para KNN:

```
[[5476    3    2   16    4    2   24    2   30    1]
 [   0 6121  167  116   57    4   35   62   34   59]
 [  14   14 5607  161    3    1   15   51   20    2]
 [   4    3   26 5645    2   52    1   52   18   16]
 [  15   13   15    5 5307    2  130   22   11  202]
 [  10    4   11  499    4 4835   39   13   81   43]
 [  33   11    5    2    3   45 5719    0   40    0]
 [   1   28   47  122   60    1    0 5755    7   76]
 [  39   25   51  112   33   35   48   30 5155  167]
 [  19   10   22  111   79    6    8  129   26 5403]]
```

Classificador: Logistic Regression | Acurácia: 0.8962 | Tempo: 0.53 segs.

Matriz de Confusão para Logistic Regression:

```
[[5296    5   27   32   16    5  118    7   53    1]
 [   0 5404  135  265  245   65  270   68   65  138]
 [  22   19 5544  100   12    1   50   91   47    2]
 [   5    5   43 5568   21   42    2   83   17   33]
 [  33    7   26    1 5302    1  121   42    8  181]
 [   7   11   31  657  103 4526   64   25   66   49]
 [  75   31    0    1   21  112 5498    0  120    0]
 [   0   42   35  123  182    2    0 5521   16  176]
 [ 107   52   56   64  101   47   76   78 4791  323]
 [  55   25    8  154  278    0    5  164   16 5108]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9279 | Tempo: 0.18 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5358   10   11   15   19    0   47   17   80    3]
 [   0 6027  222   85    9   22   38  199   31   22]
 [  22   41 5605   12    1    0    4  175   27    1]
 [   1   12   29 5470    1   19    1  247   23   16]
 [  20   71   42    0 5208    0   86    5   29  261]
 [   9   11    6  314    4 5015   50   24   67   39]
 [  77   49   37   15   56   36 5460    0  125    3]
 [   0   58   47    6   58    1    0 5882   22   23]
 [  80   59   38    5   51   29   54   57 4961  361]
 [  34   31    9   91   69    7   16   98   29 5429]]
```

Classificador: GaussianNB | Acurácia: 0.8891 | Tempo: 0.69 segs.

Matriz de Confusão para GaussianNB:

```
[[5220    1   11   32    2    1   41    0  251    1]
 [   1 5184  583  238   86   22   85  340   80   36]
 [   9   24 5289  447    4    1    8   52   53    1]
 [   2    1  212 5390    1   33    0  127   31   22]
 [  14    2   44   12 5273    0   32   44   90  211]]
```

```
[ 9 6 29 103 31 4958 46 2 169 186]
[ 78 7 89 8 15 90 5286 0 285 0]
[ 1 47 175 426 21 1 1 5323 60 42]
[ 175 5 53 182 23 7 38 13 5112 87]
[ 25 5 62 151 221 4 0 55 184 5106]]
```

Classificador: BernoulliNB | Acurácia: 0.8602 | Tempo: 0.18 segs.

Matriz de Confusão para BernoulliNB:

```
[[4600 1 284 28 170 81 240 9 136 11]
 [ 125 4553 839 11 28 29 187 72 280 531]
 [ 6 22 5659 43 7 0 55 68 23 5]
 [ 3 4 207 5360 9 49 0 138 13 36]
 [ 65 2 100 0 5019 1 121 16 55 343]
 [ 3 16 18 260 69 5038 38 11 45 41]
 [ 110 35 25 0 76 79 5394 0 139 0]
 [ 0 25 224 89 137 5 0 5523 34 60]
 [ 305 22 191 10 228 56 94 27 4520 242]
 [ 84 18 26 68 529 58 13 108 128 4781]]
```

Classificador: MultinomialNB | Acurácia: 0.8339 | Tempo: 0.05 segs.

Matriz de Confusão para MultinomialNB:

```
[[4949 5 40 108 79 2 331 7 26 13]
 [ 13 3709 858 83 257 16 793 182 140 604]
 [ 15 34 5578 77 5 2 38 115 21 3]
 [ 8 9 53 5489 2 41 2 170 14 31]
 [ 28 13 12 0 5278 0 138 23 23 207]
 [ 6 22 12 1650 53 3563 108 22 41 62]
 [ 44 79 1 0 31 119 5382 0 202 0]
 [ 0 90 63 101 112 10 1 5613 32 75]
 [ 421 52 86 23 121 25 190 65 4217 495]
 [ 48 42 14 109 296 4 8 107 58 5127]]
```

Tamanho da base de treinamento: 11000

Classificador: KNN | Acurácia: 0.9381 | Tempo: 2.29 segs.

Matriz de Confusão para KNN:

```
[[5476 3 2 16 4 2 23 2 31 1]
 [ 0 6109 171 115 56 5 37 65 35 62]
 [ 13 14 5606 161 2 1 15 52 22 2]
 [ 4 4 27 5640 2 51 1 56 18 16]
 [ 13 13 16 5 5308 2 132 21 12 200]
 [ 10 4 11 494 4 4827 40 13 88 48]
 [ 33 10 5 2 3 45 5717 0 43 0]
 [ 1 19 44 123 54 1 0 5771 7 77]
 [ 38 26 49 114 34 34 46 30 5159 165]
 [ 18 9 22 111 80 6 8 128 29 5402]]
```

Classificador: Logistic Regression | Acurácia: 0.8952 | Tempo: 0.60 segs.

Matriz de Confusão para Logistic Regression:

```
[[5300    5    26    34    16     6   112     7    53     1]
 [   1 5310   145   316   274    65   269    71    88   116]
 [   22    19 5562    99    10     1    42    86    45     2]
 [    5     5    43 5562    21    43     2    87    17    34]
 [   35     5    26     1 5296     2   126    43     8   180]
 [    7    10    30   633   101 4560    60    27    68    43]
 [   78   28     1     1   22   113 5500     0   115     0]
 [    0   42    36   122   178     2     0 5522    16   179]
 [  108   50   55    69   103    51    76    78 4791   314]
 [   56   25     8   157   288     0     5   161    14 5099]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9261 | Tempo: 0.19 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5367    10    11    17    21     0    48    17    66     3]
 [    0 6002   234    88     9    29    40   201    28    24]
 [   22    42 5601    12     1     0     4   179    27     0]
 [    1    13    27 5447     1    19     2   273    21    15]
 [   21    80    46     0 5203     0    82     6    25   259]
 [    9    12     7   321     5 5005    50    26    65    39]
 [   78    47    37    19    58    38 5451     0   127     3]
 [    0    69    43     4    54     1     0 5880    22    24]
 [   81    62    39     4    50    28    58    60 4926   387]
 [   34    32     9    96    65     7    16    99    27 5428]]
```

Classificador: GaussianNB | Acurácia: 0.8874 | Tempo: 0.66 segs.

Matriz de Confusão para GaussianNB:

```
[[5210     1    13    37     2     1    41     0   254     1]
 [    2 5129   632   250    82    20   101   327    72   40]
 [    9    23 5312   425     5     1     8    51    53     1]
 [    2     1  221 5380     2    31     0   134    26    22]
 [   14     2    59    14 5250     0    33    46    93   211]
 [   10     6    34   105    37 4957    44     2   165   179]
 [   74     7    98     5    17    91 5285     0   281     0]
 [    1    51   187   433    20     1     1 5306    55    42]
 [  180     4    57   184    20     8    38    13 5106    85]
 [   27     3    69   147   223     6     0    52   178 5108]]
```

Classificador: BernoulliNB | Acurácia: 0.8566 | Tempo: 0.18 segs.

Matriz de Confusão para BernoulliNB:

```
[[4599     1   311    29   168    78   225     9   130    10]
 [  172 4376   943    11    27    27   188    71   276   564]
 [    7    20 5662    46     7     0    49    71    21     5]
 [    3     4  220 5340    10    54     0   140    12    36]
 [   63     1   107     0 5017     1   125    16    48   344]
 [    3    17    18   242    67 5056    39    11    44    42]
```

```
[ 111  33  27   0  80  79 5398   0 130   0]
[   0  24 226  81 138   6   0 5526  34  62]
[ 319  21 208  11 230  59  99  27 4464 257]
[  84  17  28  69 526  56  13 108 113 4799]]
```

Classificador: MultinomialNB | Acurácia: 0.8325 | Tempo: 0.05 segs.

Matriz de Confusão para MultinomialNB:

```
[[4935   4  40 119  80   2 333   7  27  13]
 [  14 3653 911  85 249  15 804 166 164 594]
 [  14  30 5594  75   5   2  33 111  21   3]
 [   8   9  64 5481   2  35   2 172  14  32]
 [  29  12  12   0 5273   0 141  21  24 210]
 [   7  20  12 1627  53 3585 107  21  42  65]
 [  41  70   1   0  33 118 5393   0 202   0]
 [   0  84  69  90 112   9   1 5617  33  82]
 [ 432  45  96  21 123  25 203  64 4169 517]
 [  47  39  14 109 302   4   9 107  58 5124]]
```

Tamanho da base de treinamento: 12000

Classificador: KNN | Acurácia: 0.9387 | Tempo: 2.48 segs.

Matriz de Confusão para KNN:

```
[[5475   3   3  16   4   2  23   1  32   1]
 [   0 6113 155 128  56   5  37  65  34  62]
 [  10  13 5605 155   4   1  18  54  26   2]
 [   4   4  26 5630   2  60   1  53  22  17]
 [  13  13  17   5 5310   2 129  22  12 199]
 [  10   5   9 431   3 4891  41  12  85  52]
 [  32  10   5   2   4  43 5717   0  45   0]
 [   1  19  46 123  53   1   0 5767   6  81]
 [  33  26  45 119  34  40  50  30 5150 168]
 [  17  10  22 114  84   7   8 129  31 5391]]
```

Classificador: Logistic Regression | Acurácia: 0.8960 | Tempo: 0.44 segs.

Matriz de Confusão para Logistic Regression:

```
[[5301   5  24  36  16   6 112   7  52   1]
 [   1 5318 130 340 265  68 269  67  81 116]
 [  21  19 5559 108  11   1  38  87  42   2]
 [   5   5  40 5573  19  42   2  83  17  33]
 [  34   6  25   1 5299   2 126  41   7 181]
 [   7  10  29 633  96 4565  57  26  69  47]
 [  79  28   1   2  21 116 5496   0 115   0]
 [   0  42  36 126 168   2   0 5513  15 195]
 [ 105  47  54  66 100  50  72  74 4814 313]
 [  56  25   8 159 282   0   5 159  12 5107]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9245 | Tempo: 0.20 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```

[[5358      8    12    15    22     0    52    17    74     2]
 [   0 5991   225    94     9    31    42   209    29    25]
 [   21    39 5594    12     1     0     5   190    25     1]
 [    1    14    30 5422     1    19     2   293    23    14]
 [   25    92    44     0 5174     0    83     6    30   268]
 [    9    12     7   312     4 5013    54    26    65    37]
 [   79    46    36    19    60    41 5452     0   122     3]
 [    0    67    45     4    51     1     0 5882    22    25]
 [   83    61    37     4    51    32    63    58 4913   393]
 [   34    34     9    98    66     7    17    98    28 5422]]

```

Classificador: GaussianNB | Acurácia: 0.8864 | Tempo: 0.70 segs.

Matriz de Confusão para GaussianNB:

```

[[5211     1    11    34     2     1    38     0   261     1]
 [    1 5097   619   301    81    21   105   311    77    42]
 [    9    23 5295   442     3     1     8    55    52     0]
 [    2     1   209 5390     2    31     0   136    26    22]
 [   13     2    57    15 5238     0    37    46    88   226]
 [    9     7    32   104    32 4952    44     2   170   187]
 [   73     7    96     6    17    95 5268     0   296     0]
 [    2    48   182   435    20     1     1 5312    54    42]
 [  182     4    52   193    20    10    38    13 5096    87]
 [   25     2    63   159   206     5     0    51   178 5124]]

```

Classificador: BernoulliNB | Acurácia: 0.8549 | Tempo: 0.20 segs.

Matriz de Confusão para BernoulliNB:

```

[[4598     1   305    29   179    74   225    10   129    10]
 [  226 4292   954    11    27    28   186    73   287   571]
 [    8    20 5660    44     7     0    50    73    21     5]
 [    3     4   214 5351     7    50     0   142    13    35]
 [   68     1   104     0 5007     1   128    16    49   348]
 [    3    17    18   252    68 5042    40    11    46    42]
 [  112    34    25     0    87    78 5396     0   126     0]
 [    0    25   222    84   138     6     0 5526    30    66]
 [  334    20   195    12   224    58   100    27 4466   259]
 [   87    16    29    69   531    56    15   105   105 4800]]

```

Classificador: MultinomialNB | Acurácia: 0.8321 | Tempo: 0.06 segs.

Matriz de Confusão para MultinomialNB:

```

[[4928     4    37   125    81     2   333     7    30    13]
 [   15 3632   874    95   244    17   849   175   169   585]
 [   13    28 5596    71     4     2    35   116    21     2]
 [    8     9    57 5484     3    35     2   177    14    30]
 [   34    12    12     0 5270     0   146    18    23   207]
 [    4    19    12 1644    52 3572   111    22    42    61]
 [   40    69     0     0    36   117 5396     0   200     0]

```



```
[ 0 81 70 91 114 9 1 5617 31 83]
[ 429 44 88 19 130 25 207 62 4180 511]
[ 49 38 14 108 311 3 11 103 52 5124]]
```

Tamanho da base de treinamento: 13000

Classificador: KNN | Acurácia: 0.9387 | Tempo: 2.82 segs.

Matriz de Confusão para KNN:

```
[[5469 3 4 16 6 2 23 1 34 2]
[ 0 6114 163 123 56 5 38 65 34 57]
[ 11 12 5603 154 4 3 19 54 26 2]
[ 4 3 25 5633 2 58 1 53 23 17]
[ 12 13 16 5 5310 2 130 23 13 198]
[ 10 5 9 428 4 4898 40 13 81 51]
[ 32 10 5 2 4 48 5711 0 46 0]
[ 1 20 46 119 53 1 0 5772 6 79]
[ 34 30 39 118 36 40 53 28 5152 165]
[ 13 12 24 112 84 8 8 129 35 5388]]
```

Classificador: Logistic Regression | Acurácia: 0.8991 | Tempo: 0.51 segs.

Matriz de Confusão para Logistic Regression:

```
[[5305 5 22 34 17 6 108 7 55 1]
[ 2 5424 127 297 236 80 255 70 73 91]
[ 21 18 5553 103 12 1 40 95 43 2]
[ 5 4 39 5554 21 44 2 100 19 31]
[ 34 8 27 1 5298 2 121 42 6 183]
[ 7 11 28 592 98 4608 54 29 70 42]
[ 78 28 0 1 24 120 5494 0 113 0]
[ 0 42 35 125 172 3 0 5534 13 173]
[ 103 45 53 55 100 52 71 74 4860 282]
[ 56 23 8 156 292 0 5 164 12 5097]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9242 | Tempo: 0.23 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5351 8 13 15 22 0 57 17 76 1]
[ 0 5996 233 94 8 35 40 203 22 24]
[ 20 41 5591 12 1 0 7 191 24 1]
[ 1 16 29 5412 1 17 2 305 21 15]
[ 19 98 43 0 5168 0 85 8 26 275]
[ 8 12 7 306 5 5018 54 25 67 37]
[ 80 49 36 23 65 45 5434 0 123 3]
[ 0 72 45 3 53 1 0 5880 22 21]
[ 80 65 37 4 50 32 60 55 4921 391]
[ 28 34 11 98 62 8 17 99 25 5431]]
```

Classificador: GaussianNB | Acurácia: 0.8887 | Tempo: 0.72 segs.

Matriz de Confusão para GaussianNB:

```
[[5198 1 11 32 4 1 46 0 266 1]
```

```
[ 1 5192 564 266 87 21 101 313 73 37]
[ 8 24 5291 436 6 1 11 56 55 0]
[ 2 1 227 5368 2 30 0 135 31 23]
[ 11 3 51 12 5243 0 39 45 91 227]
[ 7 6 32 103 32 4963 48 2 161 185]
[ 67 7 88 4 16 96 5330 0 250 0]
[ 3 49 182 441 21 0 1 5306 56 38]
[ 179 4 56 171 21 12 41 13 5108 90]
[ 23 3 69 150 204 5 0 59 180 5120]]
```

Classificador: BernoulliNB | Acurácia: 0.8539 | Tempo: 0.19 segs.

Matriz de Confusão para BernoulliNB:

```
[[4594 1 279 28 186 81 240 12 127 12]
 [ 227 4276 922 12 28 28 184 74 317 587]
 [ 6 20 5646 46 7 0 58 77 23 5]
 [ 3 4 212 5336 9 55 0 155 13 32]
 [ 66 1 103 0 5018 1 126 19 48 340]
 [ 3 17 17 238 70 5063 40 11 45 35]
 [ 112 34 23 0 85 80 5399 0 125 0]
 [ 0 24 218 79 140 6 0 5543 27 60]
 [ 326 21 196 11 241 60 111 28 4437 264]
 [ 85 15 26 70 563 56 17 114 103 4764]]
```

Classificador: MultinomialNB | Acurácia: 0.8321 | Tempo: 0.06 segs.

Matriz de Confusão para MultinomialNB:

```
[[4922 4 42 123 86 2 332 7 32 10]
 [ 16 3613 861 88 259 22 862 186 178 570]
 [ 13 28 5584 73 5 2 41 119 21 2]
 [ 8 9 56 5477 4 36 2 184 15 28]
 [ 28 12 13 0 5274 0 146 19 23 207]
 [ 5 19 12 1619 55 3605 110 22 39 53]
 [ 37 66 0 0 35 125 5397 0 198 0]
 [ 0 80 72 89 116 9 1 5632 32 66]
 [ 419 43 85 16 134 27 209 63 4197 502]
 [ 44 38 14 110 337 4 9 104 52 5101]]
```

Tamanho da base de treinamento: 14000

Classificador: KNN | Acurácia: 0.9382 | Tempo: 2.94 segs.

Matriz de Confusão para KNN:

```
[[5469 3 3 17 6 2 24 2 32 2]
 [ 0 6100 173 118 52 5 46 69 37 55]
 [ 10 12 5613 148 4 3 16 54 26 2]
 [ 4 4 20 5638 2 57 2 53 22 17]
 [ 11 12 19 5 5302 2 130 24 16 201]
 [ 10 5 10 423 3 4900 41 15 83 49]
 [ 31 9 4 4 5 47 5711 0 47 0]]
```

```
[ 1 21 45 119 56 1 1 5766 7 80]
[ 39 31 41 129 35 37 53 28 5137 165]
[ 11 13 21 114 80 7 8 140 35 5384]]
```

Classificador: Logistic Regression | Acurácia: 0.8980 | Tempo: 0.68 segs.

Matriz de Confusão para Logistic Regression:

```
[[5297 5 22 36 18 6 112 7 56 1]
 [ 2 5369 134 296 254 92 260 72 72 104]
 [ 19 18 5557 94 11 1 47 88 51 2]
 [ 5 5 41 5548 23 42 2 104 17 32]
 [ 34 8 27 1 5296 2 113 44 8 189]
 [ 7 10 26 578 101 4616 55 36 70 40]
 [ 76 25 0 1 28 124 5502 0 102 0]
 [ 0 42 34 124 168 3 0 5528 17 181]
 [ 109 44 53 58 102 49 73 74 4846 287]
 [ 54 24 9 154 276 0 4 171 14 5107]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9220 | Tempo: 0.27 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5339 9 13 17 23 0 60 17 80 2]
 [ 0 5964 247 95 8 36 46 213 21 25]
 [ 20 39 5590 12 1 0 10 189 27 0]
 [ 1 16 29 5398 0 17 2 319 23 14]
 [ 21 99 48 0 5137 0 84 9 32 292]
 [ 8 11 10 307 6 5010 56 26 67 38]
 [ 82 50 36 23 69 44 5435 0 116 3]
 [ 0 71 45 3 49 1 0 5883 25 20]
 [ 85 67 36 4 48 33 69 59 4900 394]
 [ 30 33 12 104 62 7 17 105 25 5418]]
```

Classificador: GaussianNB | Acurácia: 0.8855 | Tempo: 0.69 segs.

Matriz de Confusão para GaussianNB:

```
[[5194 1 10 41 5 1 46 0 261 1]
 [ 1 5147 592 276 86 23 89 313 90 38]
 [ 9 23 5216 513 7 1 12 54 53 0]
 [ 2 1 234 5358 2 30 0 140 29 23]
 [ 8 2 50 15 5246 0 38 45 87 231]
 [ 7 6 35 103 31 4959 45 2 164 187]
 [ 67 7 92 5 20 99 5309 0 259 0]
 [ 3 48 179 441 22 1 1 5307 57 38]
 [ 183 4 51 197 22 13 43 13 5082 87]
 [ 22 3 72 154 199 5 0 63 184 5111]]
```

Classificador: BernoulliNB | Acurácia: 0.8525 | Tempo: 0.19 segs.

Matriz de Confusão para BernoulliNB:

```
[[4598 1 265 35 186 85 244 14 120 12]
 [ 286 4196 933 10 26 29 193 73 312 597]]
```

```
[ 7 20 5646 45 7 0 61 74 22 6]
[ 3 4 216 5334 8 56 0 154 12 32]
[ 65 1 108 0 5004 1 124 19 49 351]
[ 3 17 18 243 71 5055 40 11 44 37]
[ 110 34 20 0 92 82 5396 0 124 0]
[ 0 24 213 76 137 7 0 5559 28 53]
[ 340 21 196 11 238 60 111 30 4428 260]
[ 83 16 31 70 540 54 17 119 105 4778]]
```

Classificador: MultinomialNB | Acurácia: 0.8316 | Tempo: 0.06 segs.

Matriz de Confusão para MultinomialNB:

```
[[4914 4 39 132 90 2 327 8 32 12]
[ 17 3592 853 84 238 23 881 190 191 586]
[ 13 24 5587 74 5 2 44 117 20 2]
[ 7 9 58 5465 4 36 2 196 14 28]
[ 33 11 13 0 5258 0 146 21 23 217]
[ 4 19 12 1612 54 3605 109 23 41 60]
[ 37 66 1 0 35 125 5398 0 196 0]
[ 0 81 73 87 119 9 1 5630 34 63]
[ 425 41 87 20 129 25 214 65 4197 492]
[ 42 38 14 110 313 4 9 106 55 5122]]
```

Tamanho da base de treinamento: 15000

Classificador: KNN | Acurácia: 0.9374 | Tempo: 3.32 segs.

Matriz de Confusão para KNN:

```
[[5470 3 2 16 6 2 25 2 32 2]
[ 0 6093 174 118 53 5 46 69 40 57]
[ 11 13 5614 147 3 3 15 56 24 2]
[ 4 4 23 5629 2 58 2 57 22 18]
[ 12 10 19 4 5299 2 130 26 18 202]
[ 9 5 9 416 7 4898 44 15 86 50]
[ 34 9 3 4 5 43 5705 0 55 0]
[ 1 19 44 118 56 1 1 5757 7 93]
[ 41 30 38 132 36 37 52 29 5135 165]
[ 11 9 19 103 107 6 8 139 35 5376]]
```

Classificador: Logistic Regression | Acurácia: 0.8978 | Tempo: 0.59 segs.

Matriz de Confusão para Logistic Regression:

```
[[5307 5 22 34 17 6 109 6 53 1]
[ 2 5348 143 297 262 88 255 68 80 112]
[ 21 18 5550 90 10 1 49 97 50 2]
[ 5 5 42 5554 22 40 2 102 17 30]
[ 34 6 28 1 5296 2 120 42 7 186]
[ 7 10 26 577 105 4608 57 35 70 44]
[ 81 22 0 1 29 118 5482 0 125 0]
[ 0 42 34 120 177 3 0 5528 14 179]]
```

```

[ 115  42  52  60 101  45  70  72 4864 274]
[  54  24   9 154 276   0   4 162  13 5117]]
Classificador: Linear Discriminant Analysis | Acurácia: 0.9208 | Tempo: 0.26 segs.
Matriz de Confusão para Linear Discriminant Analysis:
[[5335   9  14  18  24   0  59  17  81   3]
 [   0 5945 265  91   7  36  47 216  20  28]
 [  21  41 5581  12   1   0  13 192  27   0]
 [   1  18  30 5397   0  17   2 314  24  16]
 [  21  97  46   0 5146   0  85   6  34 287]
 [   8  12  10 314   5 4996  58  28  71  37]
 [  83  47  35  25  70  46 5422   0 128   2]
 [   0  73  43   3  52   1   0 5871  29  25]
 [  90  62  35   6  51  32  64  63 4888 404]
 [  32  32  11 102  61   7  18 100  28 5422]]
Classificador: GaussianNB | Acurácia: 0.8844 | Tempo: 0.73 segs.
Matriz de Confusão para GaussianNB:
[[5238   1  10  42   4   1  46   0 217   1]
 [   1 5091 620 276  84  20  90 334  92  47]
 [   9  23 5213 511   7   1  12  61  51   0]
 [   2   1 240 5352   2  29   0 143  27  23]
 [  12   2  45  13 5230   1  39  49  91 240]
 [   7   6  34 102  26 4961  48   3 158 194]
 [  71   7  91   5  19  97 5294   0 274   0]
 [   2  52 175 455  23   0   1 5296  52  41]
 [ 185   4  61 194  21  12  44  14 5069  91]
 [  38   3  72 149 186   5   0  62 174 5124]]
Classificador: BernoulliNB | Acurácia: 0.8495 | Tempo: 0.21 segs.
Matriz de Confusão para BernoulliNB:
[[4603   1 243  42 185  85 253  13 122  13]
 [ 346 4081 918  10  25  29 204  75 313 654]
 [   7  20 5630  43   7   0  77  75  22   7]
 [   3   4 215 5337   8  55   0 152  12  33]
 [  63   1 108   0 4999   1 129  20  49 352]
 [   3  17  18 259  73 5032  40  11  45  41]
 [ 110  33  19   0  89  80 5395   0 132   0]
 [   0  23 210  75 140   6   0 5559  29  55]
 [ 357  20 191  11 239  57 116  30 4399 275]
 [  82  16  27  70 537  54  18 120 103 4786]]
Classificador: MultinomialNB | Acurácia: 0.8303 | Tempo: 0.06 segs.
Matriz de Confusão para MultinomialNB:
[[4919   4  40 129  88   2 327   8  33  10]
 [  17 3560 847  83 231  21 886 186 212 612]
 [  13  24 5587  71   6   2  47 117  20   1]

```

```
[ 7 9 59 5465 2 37 3 194 14 29]
[ 29 11 13 0 5254 0 151 20 26 218]
[ 5 19 12 1611 59 3597 112 24 40 60]
[ 37 64 1 0 35 124 5393 0 204 0]
[ 0 80 69 90 119 9 1 5630 34 65]
[ 431 41 87 21 130 25 221 63 4167 509]
[ 40 38 13 110 309 4 11 107 57 5124]]
```

Tamanho da base de treinamento: 16000

Classificador: KNN | Acurácia: 0.9381 | Tempo: 3.56 segs.

Matriz de Confusão para KNN:

```
[[5460 3 2 16 6 2 25 2 42 2]
[ 1 6094 171 131 57 4 43 67 44 43]
[ 12 12 5608 154 4 3 15 54 24 2]
[ 4 3 21 5635 2 54 2 58 22 18]
[ 11 9 17 3 5314 1 137 28 22 180]
[ 8 5 9 390 7 4907 41 15 101 56]
[ 31 9 3 4 4 44 5695 0 68 0]
[ 1 17 44 118 46 2 1 5779 5 84]
[ 39 30 38 127 34 34 49 28 5148 168]
[ 11 8 17 102 111 7 8 140 32 5377]]
```

Classificador: Logistic Regression | Acurácia: 0.9011 | Tempo: 0.67 segs.

Matriz de Confusão para Logistic Regression:

```
[[5307 5 21 33 16 6 105 6 60 1]
[ 1 5358 142 294 251 79 248 83 94 105]
[ 19 18 5552 92 10 1 45 96 53 2]
[ 4 5 40 5549 21 39 1 108 20 32]
[ 35 7 29 1 5298 1 116 43 8 184]
[ 7 10 24 533 97 4653 54 31 74 56]
[ 82 24 0 1 24 111 5488 0 128 0]
[ 0 42 34 109 161 2 0 5598 16 135]
[ 92 43 49 57 93 45 66 72 4901 277]
[ 52 24 9 144 262 0 4 166 13 5139]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9211 | Tempo: 0.27 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5329 9 14 16 24 0 60 18 88 2]
[ 0 5940 258 86 8 32 43 230 30 28]
[ 20 42 5586 12 1 0 12 187 27 1]
[ 1 18 30 5407 0 17 2 304 22 18]
[ 21 95 45 0 5141 0 89 8 31 292]
[ 9 12 9 314 5 4995 55 28 71 41]
[ 84 45 36 26 70 42 5412 0 141 2]
[ 0 66 45 3 48 1 0 5885 25 24]
[ 82 64 35 5 52 29 61 63 4893 411]]
```

```

[ 32  33  11  95  61  6  18 102  27 5428]]
Classificador: GaussianNB | Acurácia: 0.8859 | Tempo: 0.74 segs.
Matriz de Confusão para GaussianNB:
[[5213  1  10  36  4  1  43  0 251  1]
 [ 2 5115 600 258 86 22 90 347 91 44]
 [ 9  25 5226 503  6  1 10  61 47  0]
 [ 2  1 232 5356  2 30  0 146 28 22]
 [13  2  47  13 5253  1 36  50 85 222]
 [ 7  6  33 102 23 4949 45  3 166 205]
 [72  7  90  5 20  97 5303  0 264  0]
 [ 2 49 177 414 22  1  1 5338 51 42]
 [180  4  50 200 24  8 41  14 5087 87]
 [ 34  3  63 155 203  5  0  61 175 5114]]
Classificador: BernoulliNB | Acurácia: 0.8520 | Tempo: 0.20 segs.
Matriz de Confusão para BernoulliNB:
[[4597  1 265  35 181  82 243  13 131  12]
 [289 4147 952  11  26  29 193  73 332 603]
 [ 6  20 5644  48  7  0  60  74  23  6]
 [ 3  4 214 5332  7 53  0 157  12 37]
 [60  2 110  0 5001  1 127  18 52 351]
 [ 3 16  19 255  70 5039 39  10 45 43]
 [109 33 23  0  85  79 5391  0 138  0]
 [ 0 23 204 70 130  6  0 5580 29 55]
 [341 20 192 11 231 57 109 31 4439 264]
 [ 81 15 29 69 531 53 17 119 105 4794]]
Classificador: MultinomialNB | Acurácia: 0.8321 | Tempo: 0.06 segs.
Matriz de Confusão para MultinomialNB:
[[4917  4  41 129  84  2 328  8  36  11]
 [ 17 3584 908  83 225 21 849 223 194 551]
 [13  25 5589  77  6  2  39 115 21  1]
 [ 7  9  56 5469  1 37  3 195 14 28]
 [27 11 13  0 5260  0 152 25 25 209]
 [ 4 19 12 1632 56 3577 110 24 41 64]
 [37 64  1  0 33 122 5391  0 210  0]
 [ 0 78 72 83 112  9  1 5651 31 60]
 [413 41 88 22 126 25 199 66 4232 483]
 [40 38 14 110 307  2 10 107 56 5129]]
Tamanho da base de treinamento: 17000
Classificador: KNN | Acurácia: 0.9374 | Tempo: 3.68 segs.
Matriz de Confusão para KNN:
[[5452  3  2  14  6  2 28  2 49  2]
 [ 1 6089 169 124 57  9 42 63 48 53]
 [14 11 5587 156  3  3 16 69 27 2]

```

```
[ 4 2 20 5629 1 54 2 67 24 16]
[ 12 10 19 3 5321 0 133 29 27 168]
[ 9 4 9 397 8 4900 46 14 99 53]
[ 31 9 3 3 4 41 5699 0 68 0]
[ 1 14 37 131 44 2 1 5775 6 86]
[ 37 30 39 127 32 30 53 30 5147 170]
[ 9 8 16 91 118 9 9 131 48 5374]]
```

Classificador: Logistic Regression | Acurácia: 0.9048 | Tempo: 0.73 segs.

Matriz de Confusão para Logistic Regression:

```
[[5331 5 20 22 16 6 96 6 57 1]
[ 0 5416 131 274 270 75 221 98 93 77]
[ 19 18 5551 92 11 1 44 99 52 1]
[ 4 3 40 5551 22 39 1 106 21 32]
[ 35 7 28 1 5313 1 114 42 6 175]
[ 6 10 26 512 98 4675 50 29 77 56]
[ 84 24 0 1 29 96 5509 0 115 0]
[ 0 41 32 106 167 2 0 5628 16 105]
[ 88 42 47 49 95 45 58 66 4963 242]
[ 57 23 8 140 277 0 4 162 17 5125]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9231 | Tempo: 0.29 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5338 11 11 15 22 0 57 17 87 2]
[ 0 5952 235 94 8 35 42 233 29 27]
[ 21 44 5568 12 1 0 13 201 27 1]
[ 1 15 29 5425 1 17 1 289 23 18]
[ 22 79 42 0 5179 0 88 7 31 274]
[ 10 10 8 308 6 5000 55 28 73 41]
[ 84 45 37 24 70 38 5420 0 138 2]
[ 0 59 48 3 50 1 0 5889 23 24]
[ 79 59 36 6 57 29 58 57 4935 379]
[ 33 31 9 86 69 7 17 100 28 5433]]
```

Classificador: GaussianNB | Acurácia: 0.8864 | Tempo: 0.70 segs.

Matriz de Confusão para GaussianNB:

```
[[5221 1 9 35 4 1 41 0 247 1]
[ 1 5084 618 238 90 23 95 384 85 37]
[ 9 24 5226 496 6 1 9 66 51 0]
[ 2 1 228 5352 2 32 0 147 33 22]
[ 13 2 48 10 5263 1 35 50 86 214]
[ 8 6 30 101 39 4943 47 2 179 184]
[ 78 7 89 5 18 87 5312 0 262 0]
[ 1 42 155 394 22 1 1 5383 60 38]
[ 178 4 54 187 26 6 43 14 5096 87]
[ 35 3 62 147 212 4 0 60 184 5106]]
```


Classificador: BernoulliNB | Acurácia: 0.8528 | Tempo: 0.19 segs.

Matriz de Confusão para BernoulliNB:

```
[[4592    1  267   32  181   85  237   12  143   10]
 [ 280 4160  941   11   28   27  194   77  362  575]
 [   7   20 5645   46    7    0   57   78   24    4]
 [   3    4  213 5334    7   54    0  155   12   37]
 [  61    1  107    0 5013    1  124   17   54  344]
 [   3   16   20  244   77 5045   40   10   47   37]
 [ 109   34   23    0   90   76 5389    0  137    0]
 [   0   23  203   72  137    6    0 5568   30   58]
 [ 324   20  186   10  232   52  101   30 4501  239]
 [  82   15   29   67  560   54   16  116  108 4766]]
```

Classificador: MultinomialNB | Acurácia: 0.8343 | Tempo: 0.06 segs.

Matriz de Confusão para MultinomialNB:

```
[[4925    4   43  121   85    2  325    7   37   11]
 [  17 3600  944   82  267   22  805   253  189  476]
 [  13   25 5589   76    6    2   37  118   21    1]
 [   7    9   57 5472    3   38    2  191   14   26]
 [  29   11   13    0 5278    0  144   23   25  199]
 [   4   19   12 1604   61 3605  108   22   42   62]
 [  38   66    1    0   36  123 5393    0  201    0]
 [   0   78   66   88  115   10    1 5648   33   58]
 [ 395   41   85   22  127   24  182   65 4315  439]
 [  41   38   14  109  338    2    8  105   55 5103]]
```

Tamanho da base de treinamento: 18000

Classificador: KNN | Acurácia: 0.9369 | Tempo: 4.11 segs.

Matriz de Confusão para KNN:

```
[[5454    3    1   11    6    2   30    2   48    3]
 [   1 6052  177  122   64    8   36   76   58   61]
 [  15   13 5586  156    3    1   16   65   29    4]
 [   4    1   20 5625    1   56    2   63   25   22]
 [  13    9   23    2 5308    0  137   29   24  177]
 [   9    3    9  378    4 4915   45   12   88   76]
 [  25    9    5    2    4   44 5706    0   63    0]
 [   1   14   32  128   48    2    1 5762    5  104]
 [  38   24   37  110   33   34   58   27 5152  182]
 [   9    8   17   95  113    8   10  123   46 5384]]
```

Classificador: Logistic Regression | Acurácia: 0.9083 | Tempo: 0.79 segs.

Matriz de Confusão para Logistic Regression:

```
[[5359    5   21   15   16    6   81    6   50    1]
 [   1 5536  124  274  206   73  203   82   79   77]
 [  21   18 5569   90   11    1   38   92   47    1]
 [   4    3   39 5574   15   40    1   89   20   34]
```

```
[ 36 8 29 1 5316 1 110 39 6 176]
[ 6 10 23 513 92 4690 51 22 74 58]
[ 85 24 0 1 23 93 5524 0 108 0]
[ 0 41 34 112 169 2 0 5606 16 117]
[ 89 42 49 54 93 44 60 63 4949 252]
[ 55 23 9 137 267 0 4 156 14 5148]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9251 | Tempo: 0.29 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5344 11 11 15 20 0 55 17 84 3]
[ 0 5994 216 93 9 29 39 218 30 27]
[ 21 39 5588 12 1 0 8 190 28 1]
[ 1 14 30 5444 1 18 2 270 20 19]
[ 20 83 46 0 5185 0 87 7 28 266]
[ 8 10 8 317 6 4999 55 27 66 43]
[ 81 47 37 23 68 35 5442 0 123 2]
[ 0 60 49 4 51 1 0 5883 25 24]
[ 80 58 39 5 54 29 60 56 4935 379]
[ 34 31 9 82 68 7 18 96 31 5437]]
```

Classificador: GaussianNB | Acurácia: 0.8881 | Tempo: 0.68 segs.

Matriz de Confusão para GaussianNB:

```
[[5220 1 9 33 2 1 45 0 248 1]
[ 1 5170 561 238 86 22 98 363 79 37]
[ 9 24 5244 479 6 1 11 61 52 1]
[ 2 1 219 5369 2 32 0 136 33 25]
[ 14 2 50 11 5247 1 33 48 93 223]
[ 9 6 26 103 34 4945 46 2 174 194]
[ 77 7 90 7 17 88 5304 0 268 0]
[ 1 44 159 412 22 1 1 5353 61 43]
[ 174 3 52 179 24 6 42 14 5111 90]
[ 30 3 60 145 204 5 0 56 192 5118]]
```

Classificador: BernoulliNB | Acurácia: 0.8565 | Tempo: 0.19 segs.

Matriz de Confusão para BernoulliNB:

```
[[4599 1 274 29 174 85 237 11 138 12]
[ 203 4360 895 11 27 29 190 75 292 573]
[ 7 20 5649 46 7 0 57 73 24 5]
[ 3 4 214 5343 6 51 0 148 13 37]
[ 60 2 108 0 5010 1 123 17 55 346]
[ 3 16 21 243 71 5047 39 10 46 43]
[ 111 33 24 0 85 76 5395 0 134 0]
[ 0 24 214 78 135 6 0 5548 35 57]
[ 326 20 189 9 228 52 97 28 4489 257]
[ 83 15 29 65 529 56 16 115 113 4792]]
```

Classificador: MultinomialNB | Acurácia: 0.8352 | Tempo: 0.06 segs.

Matriz de Confusão para MultinomialNB:

```
[[4939    4   44  117   83    2  320    7   32   12]
 [  14 3631  971   85  239   21  771  229  172  522]
 [   13   25 5592   77    5    2   37  114   21    2]
 [    7    9   56 5483    2   37    2  180   14   29]
 [   27   10   14    0 5272    0  140   25   26  208]
 [    4   19   12 1594   56 3614  106   21   42   71]
 [   38   64    1    0   33  121 5392    0  209    0]
 [    0   78   70   92  110   10    1 5641   36   59]
 [  401   41   90   22  125   25  180   61 4291  459]
 [   40   38   14  107  309    4    8  102   62 5129]]
```

Tamanho da base de treinamento: 19000

Classificador: KNN | Acurácia: 0.9378 | Tempo: 3.79 segs.

Matriz de Confusão para KNN:

```
[[5459    3    1   12    6    2   28    2   44    3]
 [    1 6074  161  114   61    7   34   79   62   62]
 [   14   13 5602  149    3    1   16   58   28    4]
 [    4    1   21 5624    0   58    1   62   25   23]
 [   15    9   21    2 5300    0  135   28   23  189]
 [    9    3    9  419    3 4879   45   10   86   76]
 [   26    9    5    2    5   44 5705    0   62    0]
 [    1   17   34  124   49    2    1 5762    5  102]
 [   36   23   34  105   32   30   56   25 5183  171]
 [   10    8   15   94   91    7    7  126   47 5408]]
```

Classificador: Logistic Regression | Acurácia: 0.9093 | Tempo: 0.83 segs.

Matriz de Confusão para Logistic Regression:

```
[[5375    5   19   13   15    4   73    6   49    1]
 [    1 5536  120  280  225   71  193   82   73   74]
 [   21   18 5585   88   11    1   36   84   43    1]
 [    4    3   40 5581   15   40    1   85   20   30]
 [   36    7   28    1 5318    1  110   41    7  173]
 [    6   12   23  517   87 4689   51   22   75   57]
 [   88   26    0    1   23   91 5515    0  114    0]
 [    0   40   37  119  173    2    0 5601   15  110]
 [   83   41   50   51   90   45   59   61 4982  233]
 [   57   22    8  142  264    0    4  154   17 5145]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9264 | Tempo: 0.30 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5355   10   11   16   20    0   51   17   78    2]
 [    0 6010  216   87    8   22   39  216   31   26]
 [   20   39 5602   12    1    0    6  180   27    1]
 [    1   13   29 5456    1   18    1  259   22   19]
 [   25   81   41    0 5184    0   87    7   30  267]]
```

```
[ 9 10 6 322 6 4999 53 24 67 43]
[ 82 48 38 20 64 36 5442 0 125 3]
[ 0 59 48 4 55 1 0 5881 26 23]
[ 82 55 39 4 53 25 57 55 4965 360]
[ 34 31 9 90 62 7 17 96 30 5437]]
```

Classificador: GaussianNB | Acurácia: 0.8885 | Tempo: 0.73 segs.

Matriz de Confusão para GaussianNB:

```
[[5224 1 10 32 2 1 43 0 246 1]
[ 1 5179 576 232 85 23 88 356 80 35]
[ 9 24 5269 459 6 1 10 56 53 1]
[ 2 1 206 5390 1 32 0 134 31 22]
[ 15 2 52 12 5251 1 32 47 89 221]
[ 9 6 28 106 32 4942 46 2 170 198]
[ 79 7 90 7 16 88 5297 0 274 0]
[ 1 45 171 421 22 1 1 5333 61 41]
[ 172 3 50 185 24 7 38 13 5114 89]
[ 26 5 52 158 219 4 0 54 186 5109]]
```

Classificador: BernoulliNB | Acurácia: 0.8584 | Tempo: 0.20 segs.

Matriz de Confusão para BernoulliNB:

```
[[4600 1 275 27 170 83 241 11 139 13]
[ 172 4448 885 11 26 29 186 74 291 533]
[ 6 20 5660 41 7 0 55 71 23 5]
[ 3 4 212 5350 6 50 0 143 13 38]
[ 63 2 106 0 5005 1 123 17 56 349]
[ 3 16 19 255 67 5040 38 10 46 45]
[ 110 34 24 0 81 77 5392 0 140 0]
[ 0 24 223 82 138 6 0 5532 33 59]
[ 314 20 191 10 230 56 94 27 4510 243]
[ 83 16 27 69 513 55 13 110 122 4805]]
```

Classificador: MultinomialNB | Acurácia: 0.8370 | Tempo: 0.06 segs.

Matriz de Confusão para MultinomialNB:

```
[[4958 3 41 116 80 2 311 7 30 12]
[ 14 3653 985 86 240 18 747 224 166 522]
[ 13 25 5601 79 5 2 35 106 20 2]
[ 7 9 55 5498 2 34 2 170 14 28]
[ 27 9 13 0 5275 0 138 24 26 210]
[ 4 19 12 1602 54 3613 104 20 42 69]
[ 39 66 1 0 31 120 5392 0 209 0]
[ 0 79 72 99 110 10 1 5625 35 66]
[ 392 41 90 21 118 25 171 59 4331 447]
[ 40 37 14 109 290 4 8 103 66 5142]]
```

Tamanho da base de treinamento: 20000

Classificador: KNN | Acurácia: 0.9400 | Tempo: 4.38 segs.

Matriz de Confusão para KNN:

```
[[5466    3    1    12    4    2    25    2    42    3]
 [   0 6113  157  121   55   10   32   74   38   55]
 [  14   14 5616  143    4    1   16   50   27    3]
 [   4    1   22 5628    0   62    1   54   25   22]
 [  14    7   20    1 5292   21  113   29   14  211]
 [  10    3    8  377    2 4939   38   10   81   71]
 [  26    9    5    2    4   44 5723    0   45    0]
 [   1   16   45  123   47    2    1 5762    5   95]
 [  34   22   33  114   34   29   54   25 5176  174]
 [  11    8   15   93   91   11    7  119   46 5412]]
```

Classificador: Logistic Regression | Acurácia: 0.9114 | Tempo: 0.82 segs.

Matriz de Confusão para Logistic Regression:

```
[[5377    5   18   12   15    4   71    6   51    1]
 [   1 5581  121  269  205   78  182   72   72   74]
 [  22   19 5585   90   12    1   33   82   43    1]
 [   4    3   40 5591   15   39    1   75   21   30]
 [  37    8   28    0 5311    2  106   41    9  180]
 [   6   12   27  494   80 4724   49   22   73   52]
 [  86   26    0    0   20   96 5519    0  111    0]
 [   0   40   40  121  168    2    0 5603   15  108]
 [  81   43   47   53   84   46   57   55 5000  229]
 [  56   23    8  143  252    0    4  149   17 5161]]
```

Classificador: Linear Discriminant Analysis | Acurácia: 0.9279 | Tempo: 0.32 segs.

Matriz de Confusão para Linear Discriminant Analysis:

```
[[5358   10   11   15   19    0   47   17   80    3]
 [   0 6027  222   85    9   22   38  199   31   22]
 [  22   41 5605   12    1    0    4  175   27    1]
 [   1   12   29 5470    1   19    1  247   23   16]
 [  20   71   42    0 5208    0   86    5   29  261]
 [   9   11    6  314    4 5015   50   24   67   39]
 [  77   49   37   15   56   36 5460    0  125    3]
 [   0   58   47    6   58    1    0 5882   22   23]
 [  80   59   38    5   51   29   54   57 4961  361]
 [  34   31    9   91   69    7   16   98   29 5429]]
```

Classificador: GaussianNB | Acurácia: 0.8891 | Tempo: 0.67 segs.

Matriz de Confusão para GaussianNB:

```
[[5220    1   11   32    2    1   41    0  251    1]
 [   1 5184  583  238   86   22   85  340   80   36]
 [   9   24 5289  447    4    1    8   52   53    1]
 [   2    1  212 5390    1   33    0  127   31   22]
 [  14    2   44   12 5273    0   32   44   90  211]
 [   9    6   29  103   31 4958   46    2  169  186]
```

```
[ 78  7  89  8  15  90 5286  0  285  0]
[  1  47 175 426  21  1  1 5323  60  42]
[ 175  5  53 182  23  7  38  13 5112  87]
[  25  5  62 151 221  4  0  55 184 5106]]
```

Classificador: BernoulliNB | Acurácia: 0.8609 | Tempo: 0.20 segs.

Matriz de Confusão para BernoulliNB:

```
[[4598  1 286  28 169  81 236  11 139  11]
 [ 118 4566 852  10  28  29 184  72 273 523]
 [  6  21 5665  43  7  0  51  68  23  4]
 [  3  4 211 5356  9  49  0 139  12  36]
 [  63  1 104  0 5024  1 118  16  54 341]
 [  3 16  20 248  67 5052  38  9  45  41]
 [ 110 34  25  0  77  79 5394  0 139  0]
 [  0 25 223  90 137  5  0 5523  34  60]
 [ 304 21 194  11 227  58  92  27 4522 239]
 [  83 18  25  68 525  58  13 108 128 4787]]
```

Classificador: MultinomialNB | Acurácia: 0.8387 | Tempo: 0.06 segs.

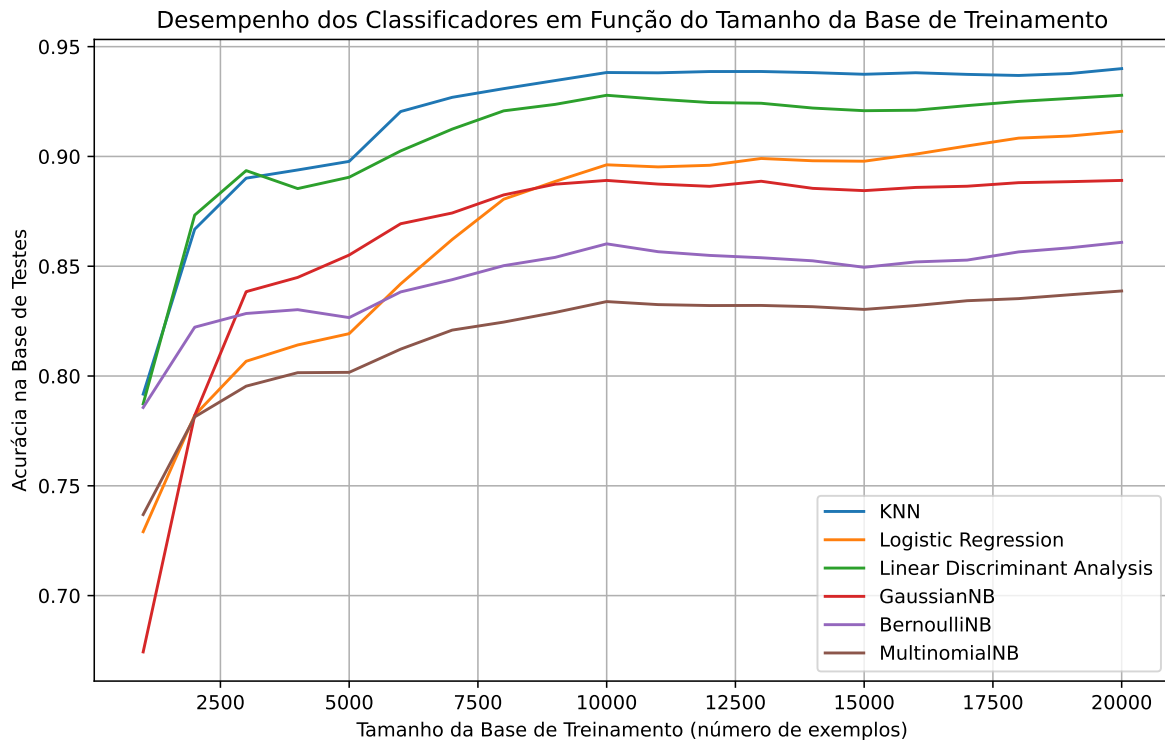
Matriz de Confusão para MultinomialNB:

```
[[4950  3  45 114  80  2 318  6  30  12]
 [  10 3691 997  86 263  19 702 205 150 532]
 [  13  25 5605  77  5  2  33 105  21  2]
 [  7  9  53 5512  2  34  2 159  14  27]
 [  25 11  13  0 5287  0 136  25  23 202]
 [  4 19  12 1587  48 3643  99  20  42  65]
 [  37 65  2  0  31 121 5392  0 210  0]
 [  0 82  67 114 107  10  1 5612  35  69]
 [ 380 42  92  23 113  26 163  59 4357 440]
 [  40 38  14 111 289  4  8 104  66 5139]]
```

Tabela de Tempos de Classificação (58.000 exemplos de teste):

Classificador	Tempo (segundos)

KNN	4.077047
Logistic Regression	0.042888
Linear Discriminant Analysis	0.041923
GaussianNB	0.637125
BernoulliNB	0.148777
MultinomialNB	0.043917



3 Em qual ponto o tamanho da base de treinamento deixa de ser relevante?

O desempenho dos classificadores começa a saturar (ou seja, a acurácia para de aumentar de maneira significativa) ao redor de 10.000 exemplos.

- KNN: A acurácia aumenta rapidamente até cerca de 5.000 exemplos, e a partir dos 10.000 exemplos, o desempenho estabiliza em torno de 0,93 a 0,94.
- Logistic Regression: O desempenho estabiliza por volta dos 10.000 exemplos, com uma leve melhoria até o final, chegando a aproximadamente 0,91.
- LDA: A partir de 10.000 exemplos, o classificador também mostra uma estabilização em torno de 0,92 a 0,93.
- GaussianNB, BernoulliNB e MultinomialNB: Esses classificadores apresentam um comportamento de estabilização mais cedo, entre 5.000 e 10.000 exemplos, sem mudanças significativas após 10.000 exemplos.

Portanto, para a maioria dos classificadores, 10.000 exemplos parece ser o ponto onde o aumento da base de treinamento passa a ter impacto marginal sobre o desempenho.

4 Qual é o classificador que tem o melhor desempenho com poucos dados (< 1000 exemplos)?

Com menos de 1.000 exemplos:

- KNN e LDA têm um desempenho inicial muito bom. O KNN apresenta um crescimento acentuado e ultrapassa a acurácia de 0,80 rapidamente.
- GaussianNB apresenta uma queda acentuada com poucos exemplos, enquanto BernoulliNB e MultinomialNB têm desempenho mais estável, mas inferior.

Com base na acurácia inicial, KNN parece ter o melhor desempenho com menos de 1.000 exemplos.

5 Qual é o classificador que tem melhor desempenho com todos os dados?

Com todos os 20.000 exemplos:

- KNN é o classificador com o melhor desempenho geral, atingindo uma acurácia ligeiramente superior a 0,94.
- LDA segue de perto, com acurácia em torno de 0,93.
- Logistic Regression chega a cerca de 0,91.
- Naive Bayes (GaussianNB, BernoulliNB, MultinomialNB) têm desempenho inferior, alcançando valores entre 0,83 e 0,89

6 Qual é o classificador mais rápido para classificar os 58.000 exemplos de teste?

O classificador mais rápido foi a Logistic Regression com 0,042 segundos, seguido de perto pelo MultinomialNB com 0,044 segundos.

7 Os erros são os mesmos para todos os classificadores quando todos eles utilizam toda a base de treinamento?

Não, os erros não são os mesmos para todos os classificadores, mesmo quando todos eles utilizam toda a base de treinamento. A seguir estão algumas considerações sobre o motivo disso:

- Algoritmos Diferentes, Mecanismos Diferentes: Cada algoritmo de classificação faz suposições diferentes sobre a estrutura dos dados, o que resulta em padrões de erro diferentes. Por exemplo:
 - KNN classifica novos pontos com base nos vizinhos mais próximos, o que pode causar erros em áreas com classes muito misturadas.
 - Naive Bayes assume que os atributos são condicionalmente independentes, o que pode levar a previsões erradas quando essa suposição não é verdadeira nos dados.
 - LDA maximiza a separação entre as classes de forma linear, o que pode ser menos eficaz em dados que exigem uma fronteira de decisão não linear.
- Distribuição de Erros em Matrizes de Confusão: A matriz de confusão fornece uma visão detalhada de onde os erros ocorrem, mostrando como os exemplos de uma classe são classificados incorretamente como outra classe. A matriz de confusão será diferente para cada classificador, pois:
 - KNN pode errar mais em classes com muitos vizinhos de outras classes próximas.
 - Logistic Regression pode errar se a relação entre as variáveis explicativas e a resposta não for linear.
 - Naive Bayes pode cometer erros sistemáticos em classes cujas distribuições de atributos não seguem o modelo estatístico assumido.
- Classificadores Probabilísticos vs Baseados em Distância: Classificadores como Naïve Bayes (probabilísticos) podem gerar erros sistemáticos em todas as classes se os atributos não forem independentes, enquanto KNN (baseado em distância) pode cometer erros mais localizados em áreas densas de dados.

8 Conclusão sobre as Matrizes de Confusão:

Os padrões de erro (erros de classificação entre as classes) variam significativamente entre os classificadores, mesmo quando todos utilizam a mesma quantidade de dados. Isso é uma consequência direta das diferentes abordagens e suposições que cada algoritmo utiliza para separar as classes. Assim, as matrizes de confusão serão diferentes entre os classificadores, refletindo suas diferentes fraquezas e vantagens em relação ao conjunto de dados específico