PROJETO AM 2022-2 FRANCISCO

Aprendizagem de Máquina
Equipe:
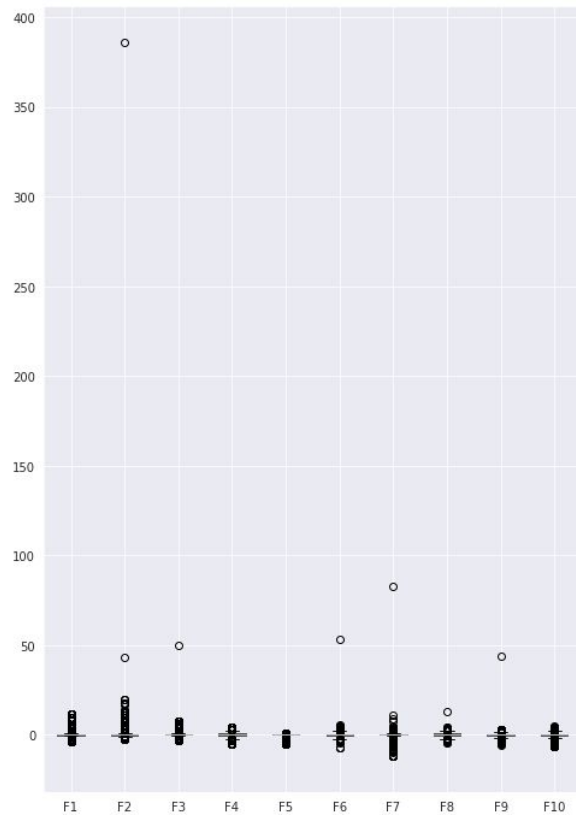- Danilo Vaz (dvma)
- Humberto Lopes (hlfs2)
- Marcelo Valois (mmv4)
- Matheus Albuquerque (mvca)

Dataset

- 10 atributos e uma classe
- Usa as 4 primeiras linhas
- 12 classes

cin.ufpe.br

Treinamento

- Divisão 80/20
- Standardization - Robust Scaler
- Cross Validation estratificado (5 folds)

Modelos

- KNN
- Naive Bayes
- Logistic Regression
- Decision Tree

cin.ufpe.br

Resultado KNN usando valores default

Treino:
- Average precision: 0.843951 (0.039310)
- Average recall: 0.776503 (0.042320)
- Average f1-score: 0.801923 (0.036729)
- Average accuracy: 0.825142 (0.008664)

Validação:
- Precision: 0.81
- Recall: 0.88
- F1-score: 0.84
- Accuracy: 0.84

Fone: + 55 81 2126.8430   Fax: + 55 81 2126.8438   E-mail: contato@cin.ufpe.br
Av. Jornalista Anibal Fernandes, s/n Cidade Universitária - Recife - CEP: 50740-560

cin.ufpe.br

Resultado Naive Bayes usando valores default

Treino:
- Average precision: 0.368469 (0.013381)
- Average recall: 0.455413 (0.024343)
- Average f1-score: 0.355641 (0.017897)
- Average accuracy: 0.306690 (0.023662)

Validação:
- Precision: 0.48
- Recall: 0.36
- F1-score: 0.36
- Accuracy: 0.28

Fone: + 55 81 2126.8430   Fax: + 55 81 2126.8438   E-mail: contato@cin.ufpe.br
Av. Jornalista Anibal Fernandes, s/n Cidade Universitária - Recife - CEP: 50740-560

cin.ufpe.br

Resultado Logistic Regression usando valores default

Treino:
- Average precision: 0.461227 (0.023740)
- Average recall: 0.390985 (0.015685)
- Average f1-score: 0.396923 (0.016273)
- Average accuracy: 0.565196 (0.007598)

Validação:
- Precision: 0.38
- Recall: 0.41
- F1-score: 0.38
- Accuracy: 0.55

Fone: + 55 81 2126.8430   Fax: + 55 81 2126.8438   E-mail: contato@cin.ufpe.br
Av. Jornalista Anibal Fernandes, s/n Cidade Universitária - Recife - CEP: 50740-560

cin.ufpe.br

Resultado Decision Tree usando valores default

Treino:
- Average precision: 0.904252 (0.031071)
- Average recall: 0.896720 (0.026768)
- Average f1-score: 0.899264 (0.029052)
- Average accuracy: 0.937678 (0.019316)

Validação:
- Precision: 0.91
- Recall: 0.92
- F1-score: 0.92
- Accuracy: 0.94

Fone: + 55 81 2126.8430   Fax: + 55 81 2126.8438   E-mail: contato@cin.ufpe.br
Av. Jornalista Anibal Fernandes, s/n Cidade Universitária - Recife - CEP: 50740-560

cin.ufpe.br

Busca pelos hiperparâmetros

- Grid Search
- Bayes Search

cin.ufpe.br

Hiperparâmetros KNN

- K = 3, definido no conjunto de validação
- ponderando pela distância
- Distância Euclidiana

Resultado:
- Precision: 0.87
- Recall: 0.90
- F1-score: 0.88
- Accuracy: 0.86

Fone: + 55 81 2126.8430   Fax: + 55 81 2126.8438   E-mail: contato@cin.ufpe.br
Av. Jornalista Anibal Fernandes, s/n Cidade Universitária - Recife - CEP: 50740-560

cin.ufpe.br

Hiperparâmetros Naive Bayes

- Smoothing ≅ 1.955e-09

Resultado:
- Precision: 0.48
- Recall: 0.36
- F1-score: 0.36
- Accuracy: 0.28

Fone: + 55 81 2126.8430   Fax: + 55 81 2126.8438   E-mail: contato@cin.ufpe.br
Av. Jornalista Anibal Fernandes, s/n Cidade Universitária - Recife - CEP: 50740-560

cin.ufpe.br

Hiperparâmetros Logistic Regression

- Penalty: l2
- Solver: lbfgs
- C ≅ 0.7616

Resultado:
- Precision: 0.38
- Recall: 0.41
- F1-score: 0.38
- Accuracy: 0.55

Fone: + 55 81 2126.8430   Fax: + 55 81 2126.8438   E-mail: contato@cin.ufpe.br
Av. Jornalista Anibal Fernandes, s/n Cidade Universitária - Recife - CEP: 50740-560

cin.ufpe.br

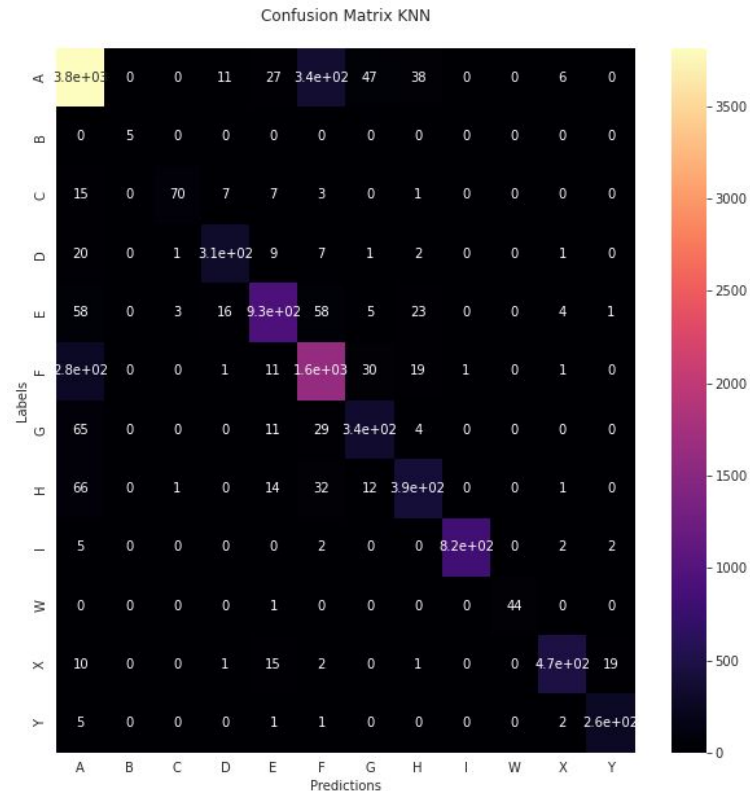# Hiperparâmetros Decision Tree e Resultado Validação

- Criterion: Entropy
- Depth: 30
- minimum samples leaf: 1
- minimum samples split: 1

Resultado:
- Precision: 0.97
- Recall: 0.98
- F1-score: 0.97
- Accuracy: 0.98

Fone: + 55 81 2126.8430   Fax: + 55 81 2126.8438   E-mail: contato@cin.ufpe.br
Av. Jornalista Anibal Fernandes, s/n Cidade Universitária - Recife - CEP: 50740-560

cin.ufpe.br

Resultado KNN no teste

	precision	recall	f1-score	support
A	0.88	0.89	0.88	4286
B	1.00	1.00	1.00	5
C	0.93	0.68	0.79	103
D	0.90	0.88	0.89	353
E	0.91	0.85	0.88	1095
F	0.77	0.83	0.80	1962
G	0.78	0.76	0.77	447
H	0.82	0.76	0.79	520
I	1.00	0.99	0.99	832
W	1.00	0.98	0.99	45
X	0.97	0.91	0.94	522
Y	0.92	0.97	0.94	267
accuracy			0.87	10437
macro avg	0.91	0.87	0.89	10437
weighted avg	0.87	0.87	0.87	10437

Confusion Matrix KNN

Fone: + 55 81 2126.8430   Fax: + 55 81 2126.8438   E-mail: contato@cin.ufpe.br
Av. Jornalista Anibal Fernandes, s/n Cidade Universitária - Recife - CEP: 50740-560

cin.ufpe.br

Resultado Naive Bayes no teste

	precision	recall	f1-score	support
A	0.50	0.06	0.11	4286
B	1.00	1.00	1.00	5
C	0.04	0.07	0.05	103
D	0.16	0.25	0.19	353
E	0.25	0.09	0.14	1095
F	0.28	0.46	0.35	1962
G	0.16	0.70	0.26	447
H	0.24	0.80	0.36	520
I	0.69	0.64	0.67	832
W	0.10	0.42	0.16	45
X	0.50	0.63	0.56	522
Y	0.50	0.47	0.48	267
accuracy			0.30	10437
macro avg	0.37	0.47	0.36	10437
weighted avg	0.40	0.30	0.26	10437

Confusion Matrix NB
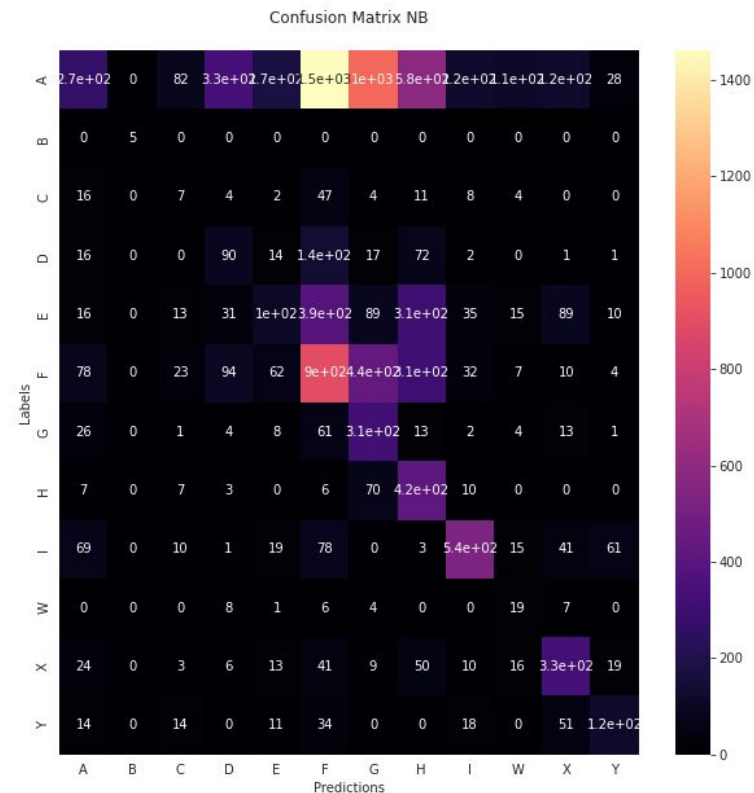
Resultado Logistic Regression no teste

	precision	recall	f1-score	support
A	0.53	0.94	0.68	4286
B	1.00	1.00	1.00	5
C	0.00	0.00	0.00	103
D	0.00	0.00	0.00	353
E	0.55	0.25	0.34	1095
F	0.38	0.10	0.16	1962
G	0.00	0.00	0.00	447
H	0.49	0.20	0.29	520
I	0.86	0.88	0.87	832
W	0.22	0.04	0.07	45
X	0.75	0.71	0.73	522
Y	0.69	0.64	0.67	267
accuracy			0.56	10437
macro avg	0.46	0.40	0.40	10437
weighted avg	0.50	0.56	0.48	10437

Confusion Matrix LR

Fone: + 55 81 2126.8430   Fax: + 55 81 2126.8438   E-mail: contato@cin.ufpe.br
Av. Jornalista Anibal Fernandes, s/n Cidade Universitária - Recife - CEP: 50740-560

cin.ufpe.br
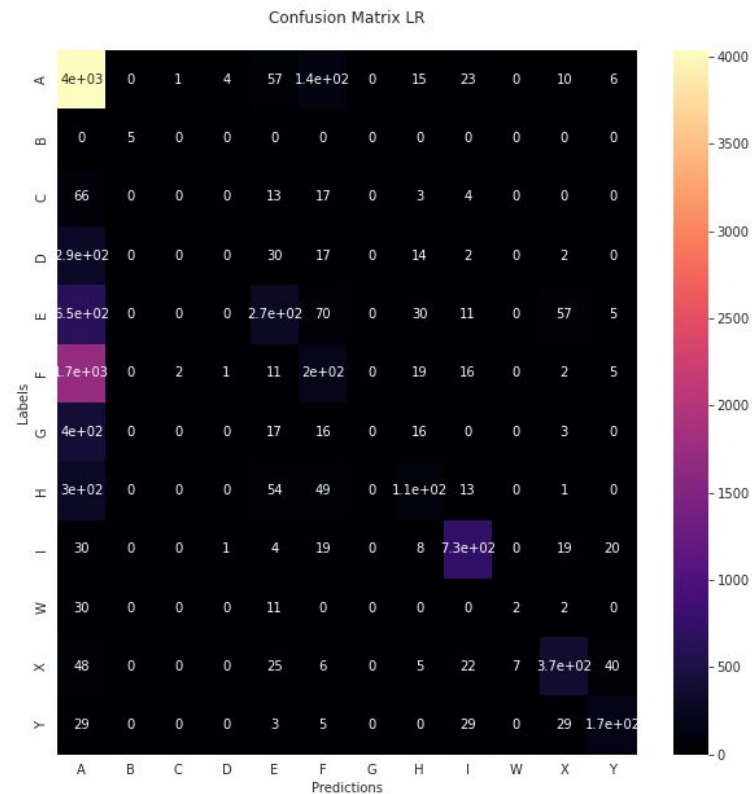
Resultado Decision Tree no teste

	precision	recall	f1-score	support
A	0.99	0.99	0.99	4286
B	1.00	1.00	1.00	5
C	0.95	0.94	0.95	103
D	1.00	0.97	0.98	353
E	0.96	0.97	0.96	1095
F	0.99	0.99	0.99	1962
G	0.98	0.96	0.97	447
H	0.94	0.97	0.95	520
I	1.00	0.99	0.99	832
W	0.94	0.98	0.96	45
X	0.93	0.96	0.95	522
Y	0.98	0.97	0.98	267
accuracy			0.98	10437
macro avg	0.97	0.97	0.97	10437
weighted avg	0.98	0.98	0.98	10437

Confusion Matrix Decision Tree