



Universidade do Minho

Ano Letivo de 2018/2019



Extração de Conhecimento

Tomada de decisão do aluno

Luís Gomes (A78701), Joel Rodrigues (A79068) e Elisa Valente (A79093).

Conteúdos

- Descrição e análise do conjunto de dados;
- Visualização dos dados;
- Pré-preparação dos dados;
- Treino e validação do modelo;
- Seleção do modelo;
- Hiperparameterização (HP);
- Conclusão.

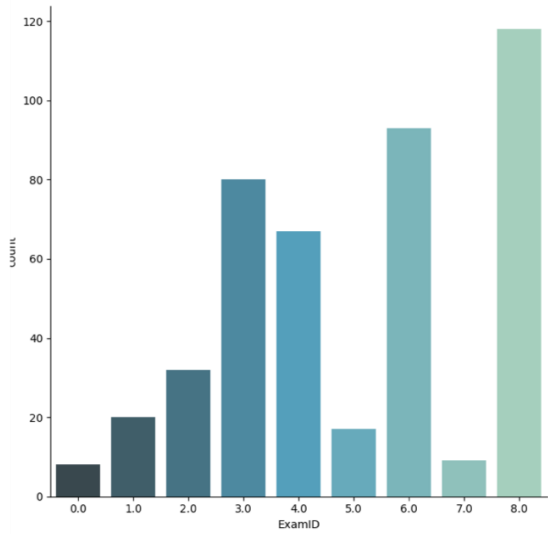
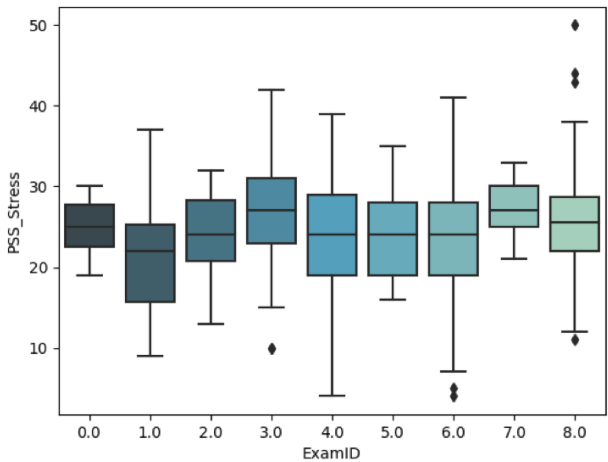
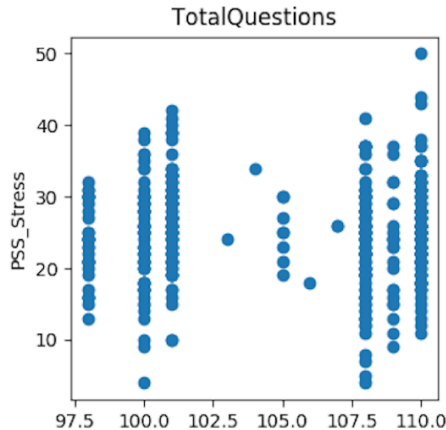
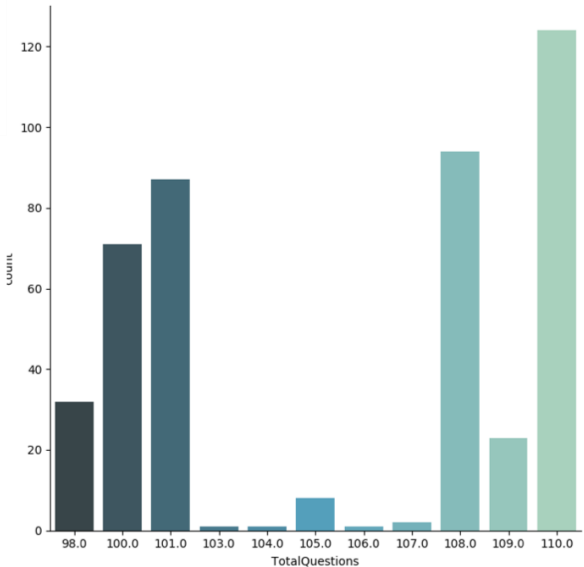
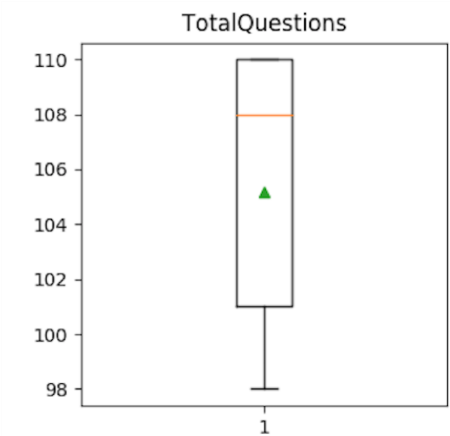


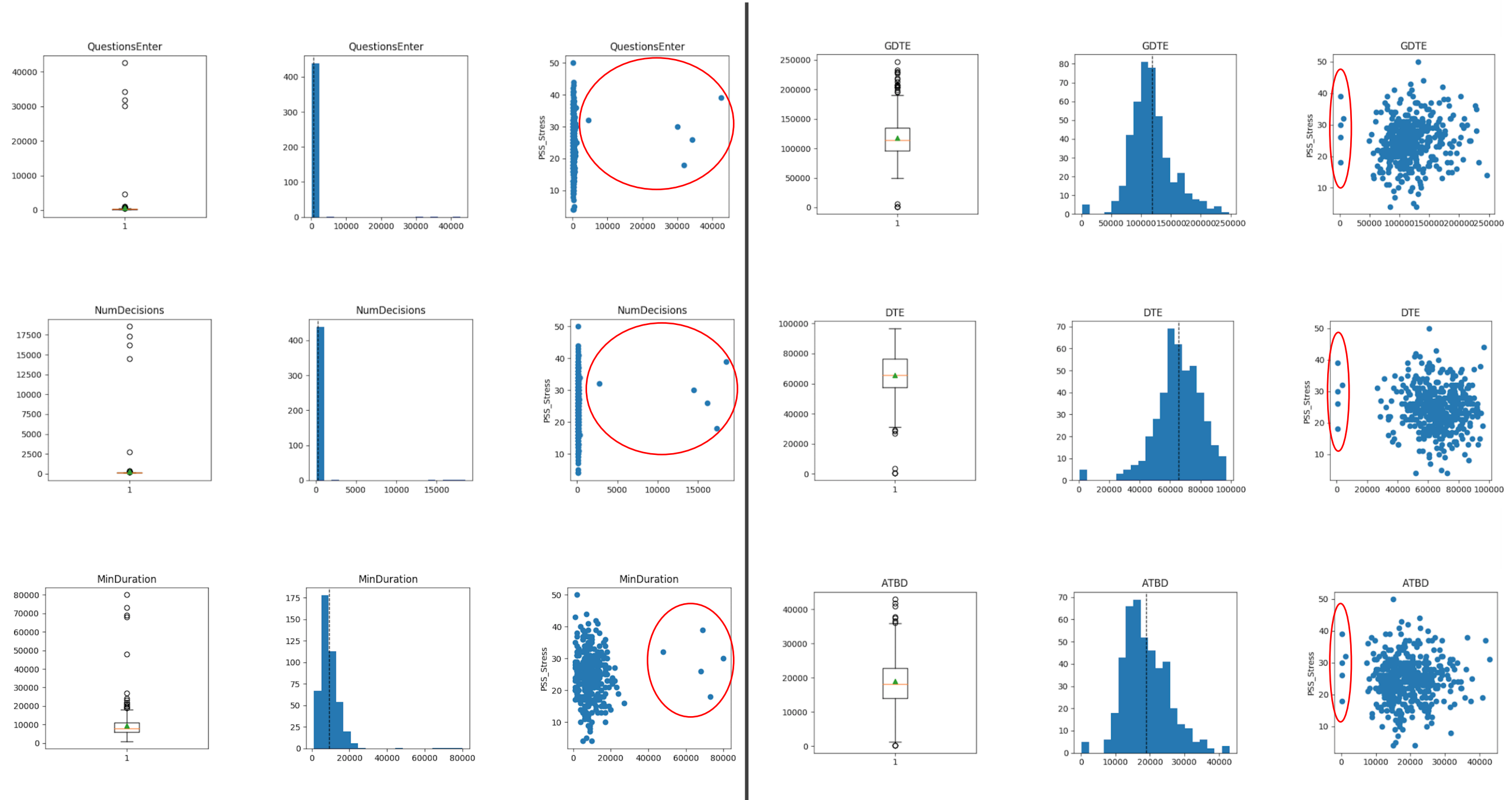
Descrição e análise do conjunto de dados

ExamID	FinalGrade	StudyID	TotalQuestions	ATBQ	ATBD
MTBD	DTE	GDTE	MaxDuration	MinDuration	NumDecisions
QuestionsEnter	DMR	CDMR	TotalDuration	VTBD	PSS_Stress

	Var. Categórica
	Var. Quantitativa Contínua
	Var. Quantitativa Discreta

Visualização dos dados





Seleção dos dados

ExamID	FinalGrade	StudyID	TotalQuestions	ATBQ	ATBD
MTBD	DTE	GDTE	MaxDuration	MinDuration	NumDecisions
QuestionsEnter	DMR	CDMR	TotalDuration	VTBD	PSS_Stress

	Não utilizados
	Preditores
	Variável de resposta

Filtragem de dados em falta

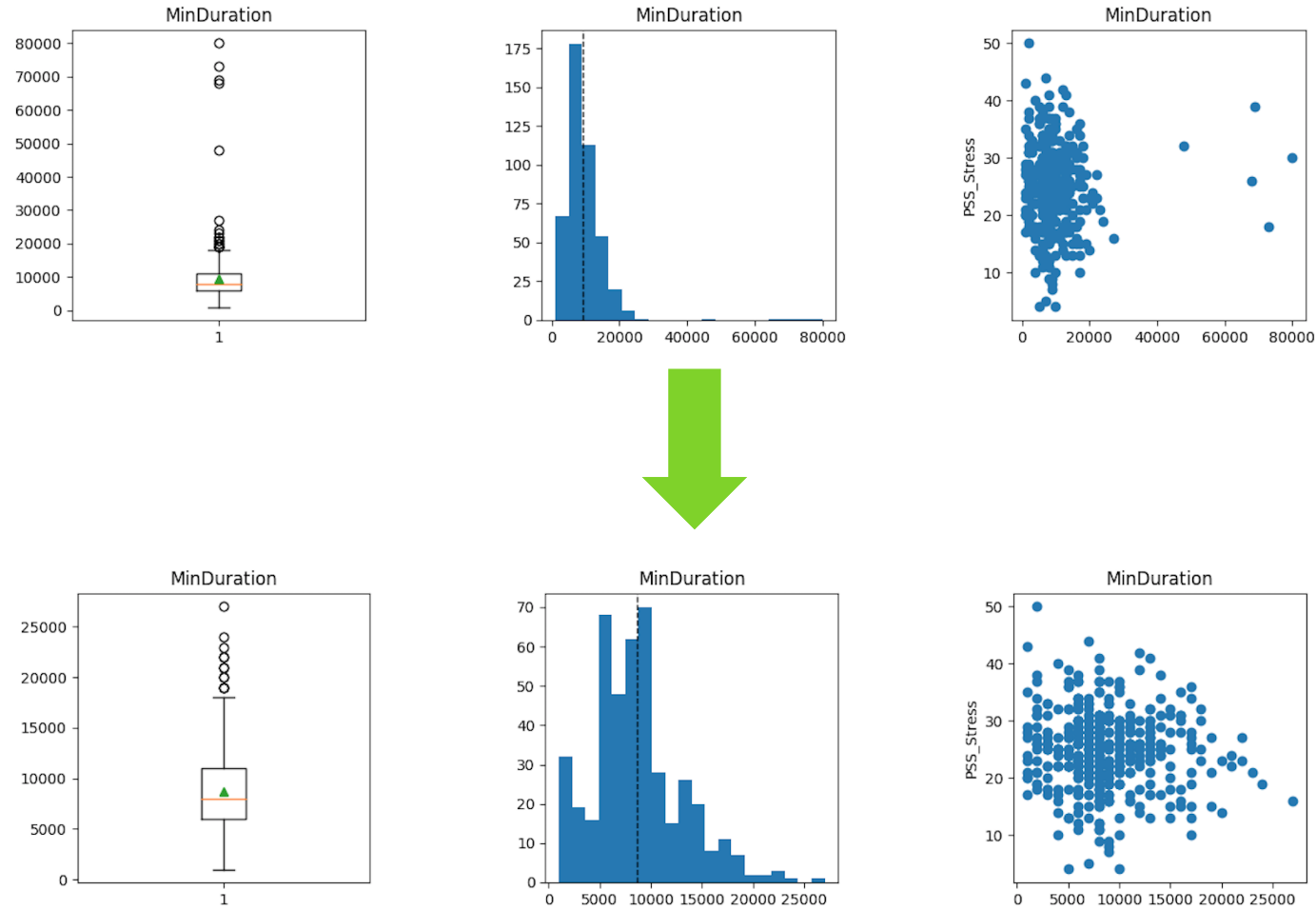
ExamID	0.000000
FinalGrade	0.000000
PSS_Stress	0.000000
StudyID	0.000000
TotalQuestions	0.000000
ATBQ	0.000000
ATBD	0.000000
DTE	0.000000
GDTE	0.000000
MaxDuration	0.000000
MTBD	0.009009
MinDuration	0.000000
NumDecisions	0.000000
QuestionsEnter	0.000000
DMR	0.000000
CDMR	0.000000
TotalDuration	0.000000
VTBD	0.000000

Filtragem de dados em falta

Dataset_DecisionPSS

ExamID	FinalGrade	PSS_Stress	StudyID	TotalQuestions	ATBQ	ATBD	DTE	GDTE	MaxDuration	MTBD	MinDuration	NumDecisions	QuestionsEnter	DMR	CDMR	TotalDuration	VTBD
1969	0.780952380952381	21	2016.03.13	105	58066.66667	14146.17169	51235.29412	70080.45977	330000.0	5000.0	7000.0	119	298	0.27869	0.73109	6097000.0	743929746.93789
1969	0.7333333333333333	27	2016.03.132	105	84676.19048	17746.50699	62174.82517	103383.72093	316000.0	10000.0	13000.0	143	327	0.28657	0.6014	8891000.0	655369612.77445
1969	0.6476190476190476	25	2016.03.2	105	54123.80952	20516.24549	46203.25203	75773.33333	403000.0	7000.0	2000.0	123	149	0.44565	0.60976	5683000.0	1428569481.50474
1969	0.7238095238095238	30	2016.03.1	105	73152.38095	15611.78862	60480.31496	90364.70588	490000.0	6000.0	9000.0	127	336	0.26132	0.66929	7681000.0	873741042.00818
1969	0.8285714285714286	19	2016.03.118	105	86352.38095	20560.0907	62531.03448	85537.73585	749000.0	6000.0	7000.0	145	281	0.33721	0.73103	9067000.0	1205215130.90085
1969	0.6190476190476191	30	2016.03.114	105	80466.66667	19926.88679	61224.63768	112653.33333	899000.0	8000.0	11000.0	138	279	0.32701	0.54348	8449000.0	1807959180.82876
1969	0.5142857142857142	25	2016.03.18	105	96733.33333	7991.3454	61186.74699	125395.06173	384000.0	2000.0	14000.0	166	1076	0.13091	0.48795	10157000.0	335986539.21211
1969	0.7714285714285715	23	2016.03.16	105	68790.47619	21561.19403	66266.05505	89172.83951	436000.0	6000.0	11000.0	109	226	0.32733	0.74312	7223000.0	1854121244.079
2311	0.49074074074074076	26	2015.01.118	108	102907.40741	192.41024	687.5348	1136.86579	186000.0	NaN	68000.0	16165	34302	0.29843	0.60476	11114000.0	291919.33393
2038	0.5596330275229358	32	2016.03PA.141	109	96706.42202	1246.42308	3827.5236	6065.01726	196000.0	1000.0	48000.0	2754	4603	0.3277	0.63108	10541000.0	5434301.59629
2167	0.4158415841584158	39	2016.01.49	101	100603.9604	154.32633	547.67423	1069.80417	148000.0	NaN	69000.0	18553	42626	0.28345	0.51194	10161000.0	146003.79928
2262	0.37	18	2016.01.42	100	103790.0	164.37553	599.21483	1187.66449	144000.0	NaN	73000.0	17321	31838	0.32327	0.50453	10379000.0	201500.55158
2055	0.4387755102040816	30	2016.03.8	98	102377.55102	209.5447	693.89308	1130.09687	138000.0	NaN	80000.0	14459	30123	0.30202	0.61401	10033000.0	171487.25066
2038	0.7431192660550459	20	2016.03PA.147	109	67688.07339	14875.0	42895.34884	71631.06796	336000.0	9000.0	9000.0	172	257	0.35102	0.59884	7378000.0	456521717.17172
2038	0.6788990825688074	24	2016.03PA.146	109	78853.21101	23104.83871	69314.51613	107437.5	469000.0	8000.0	5000.0	124	234	0.33423	0.64516	8595000.0	1777824558.73402
2038	0.5321100917431193	36	2016.03PA.145	109	87137.61468	20966.88742	67842.85714	141761.19403	277000.0	7000.0	10000.0	140	286	0.30905	0.47857	9498000.0	977098458.65323

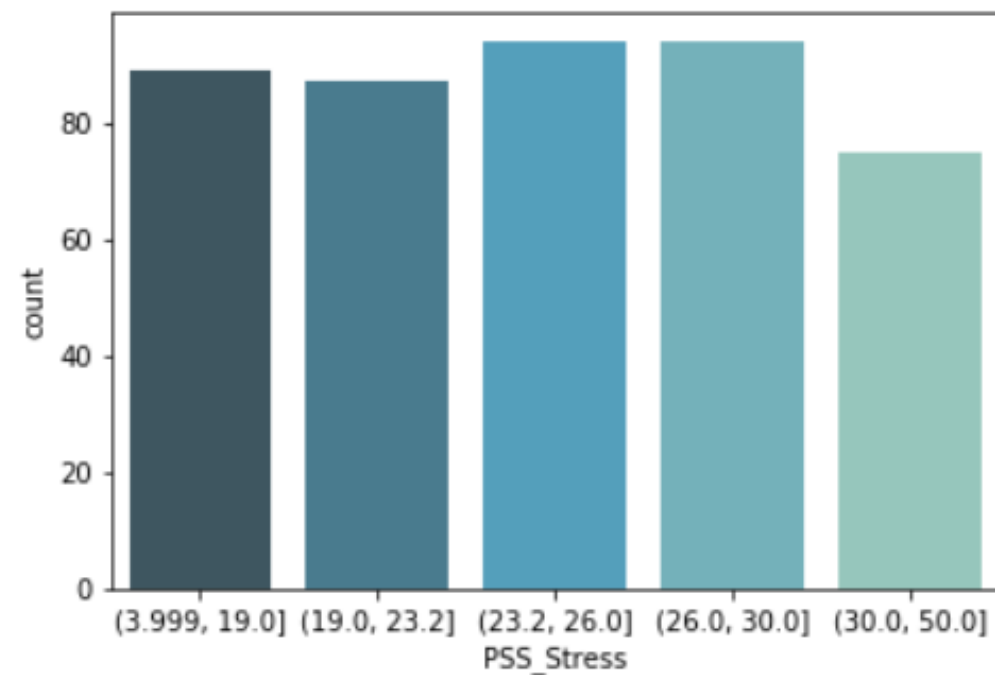
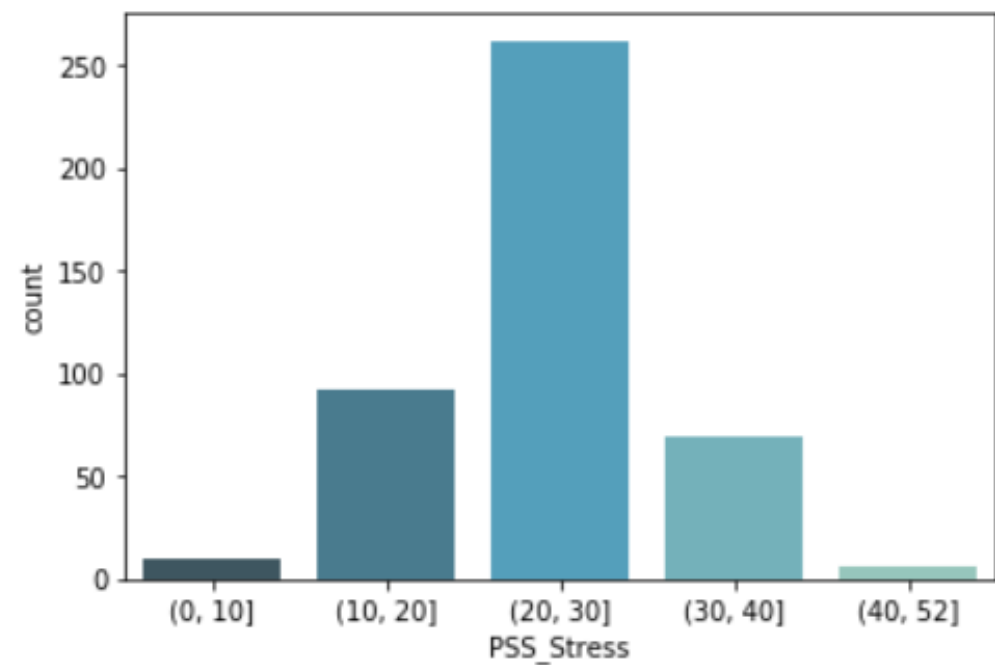
Filtragem de dados em falta



Transformações

Normalização	Padronização	Discretização e codificação de etiquetas
[0,1]	Média 0 e variância 1	Numérico para nominal (5 categorias)

Transformações



Treino e validação do modelo

Modelos Usados				
SVM	GaussianNB	LR	RNA	KNN

Preditores

- Normalizados
- Padronizados

Variável de resposta

- Discretizada
 - Padronizada
 - Normalizada

Classificação				
SVC	RNA_CLF	LR	KNN	GNB

SVM **0.597** t_disc f_norm

LR **0.599** t_disc f_norm

RNA_CLF **0.597** t_disc f_norm

SVM **0.597** t_disc f_stand

LR **0.588** t_disc f_stand

RNA_CLF **0.57** t_disc f_stand

KNN **0.526** t_disc f_norm

GNB **0.378** t_stand_norm f_stand

Seleção do modelo

Modelo de classificação

Regressão Logística

(variável de resposta discretizada e preditores normalizados)

Kfold=10

LR 0.599 t_disc f_norm

Treino 80% e Teste 20%

LR 0.693 t_disc f_norm

Hiperparameterizaçao

Grid search

```
gridSearch.best_params_ {'C': 1e-06, 'penalty': 'l2'}  
lr.C=0.001 ✓  
lr.penalty='l2' ✓  
x_train,x_test,y_train,y_test = train_test_split(features_normalized,target_discretized,test_size=0.2) ✓  
lr.fit(X=x_train,y=y_train)  


LogisticRegression(C=0.001, class_weight=None, dual=False, fit_intercept=True,  
    intercept_scaling=1, max_iter=100, multi_class='warn',  
    n_jobs=None, penalty='l2', random_state=None, solver='warn',  
    tol=0.0001, verbose=0, warm_start=False)



✕  
📄

  
lr.score(x_test,y_test) 0.7045454545454546
```

Conclusão





Universidade do Minho

Ano Letivo de 2018/2019



Extração de Conhecimento

Tomada de decisão do aluno

Luís Gomes (A78701), Joel Rodrigues (A79068) e Elisa Valente (A79093).