FATIGUE MONITORING & INJURY PREVENTION IN ELITE SWIMMERS

Hugo Veríssimo 124348

João Cardoso 50750

Foundations of Machine Learning 24/25

universidade de aveiro

CONTENTS

universidade de aveiro

PROJECT OVERVIEW

THE PROBLEM

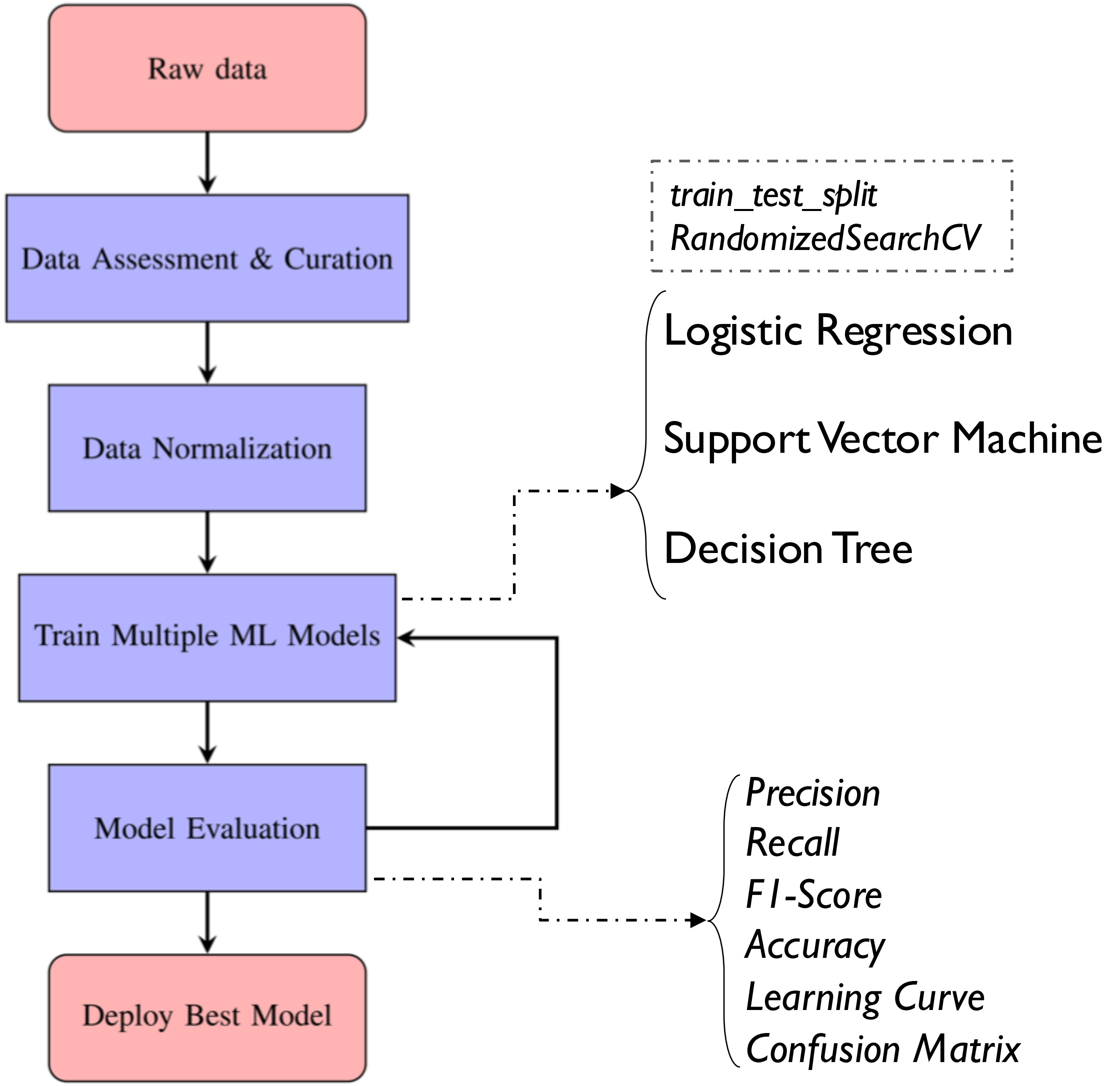
- Improving performance of elite athletes is the ultimate goal of any sports coach.
- Balance between workload and rest is crucial.
- Too high effort, may lead to injury, too low and some gains may be left on the table.

universidade de aveiro

STATE OF THE ART

- Fatigue and injury prediction using ML models, such as Logistic Regression, Random Forest, Support Vector Machine, among others, across different sports contexts.
- Feature engineering and selection is among the most debated topics.
- Overfitting problems.

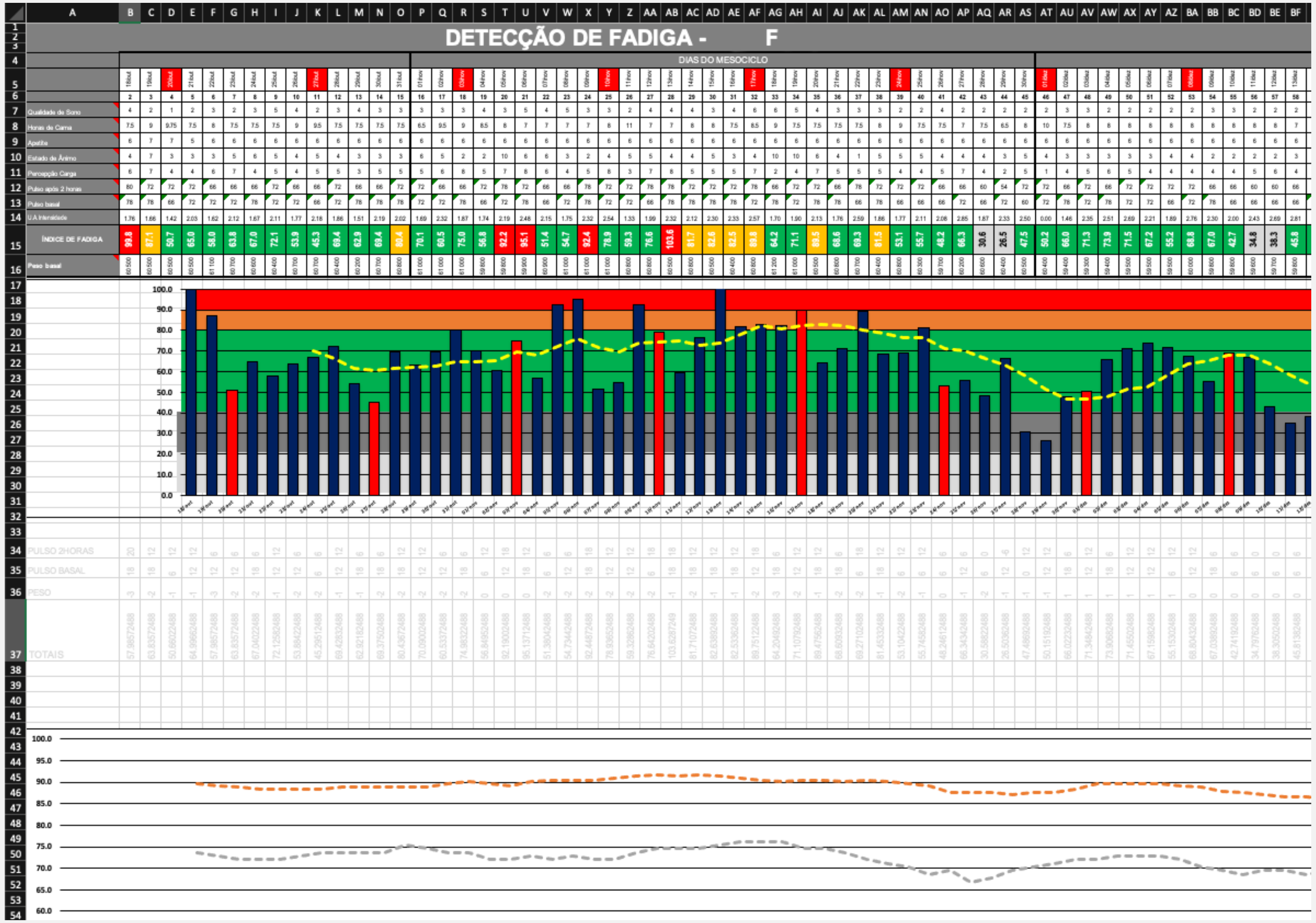
METHODOLOGY



DATA ANALYSIS

RAW DATA

- Swimming club CAPGE
- Season 2019/2020
- 7 athletes
- 11 different metrics
- Around 200 obs. per athlete



universidade de aveiro

STRUCTURED DATA

	Date	Appetite	StSpirit	pEffort	ual	FatigInd	DeltaWt (%)	DeltaHrtwo	Name	Sex_F	SleepInd
0	2019-10-18	6.0	4.0	6.0	1.76	99.809325	NaN	2.0	F_f	1	-3.50
1	2019-10-19	7.0	7.0	7.0	1.66	87.095325	0.000000	-6.0	F_f	1	-7.00
2	2019-10-20	7.0	3.0	4.0	1.42	50.660225	0.000000	6.0	F_f	1	-8.75
3	2019-10-21	5.0	3.0	4.0	2.03	64.996625	0.000000	0.0	F_f	1	-5.50
4	2019-10-22	6.0	3.0	6.0	1.62	57.985725	0.009917	-6.0	F_f	1	-5.00
...	...	...	...	...	...	...	...	...	...	...	...
1267	2020-05-12	6.0	5.0	6.0	2.00	115.433793	0.000000	12.0	I_f	1	-5.00
1268	2020-05-13	6.0	5.0	5.0	2.00	53.098793	0.000000	6.0	I_f	1	-6.50
1269	2020-05-14	6.0	5.0	5.0	2.00	53.423793	0.000000	6.0	I_f	1	-6.00
1270	2020-05-15	6.0	5.0	6.0	2.00	80.333793	0.000000	6.0	I_f	1	-5.00
1271	2020-05-16	6.0	5.0	5.0	2.00	65.123793	0.000000	0.0	I_f	1	-6.00

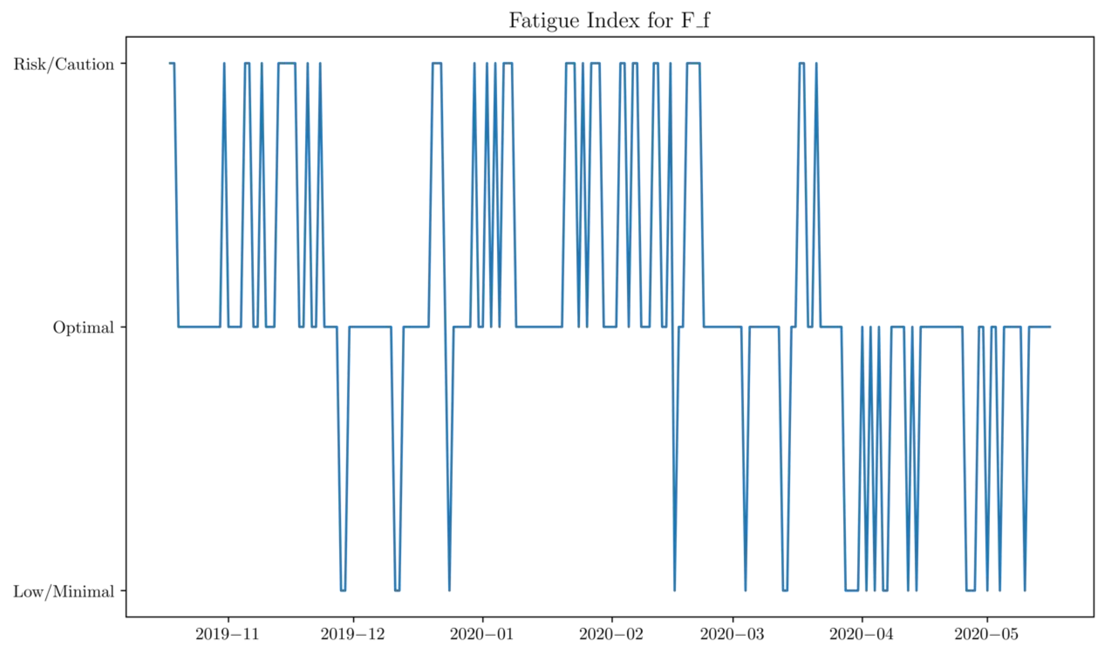
1272 rows × 11 columns

FEATURE ENGINEERING

- The figure, illustrating the periodicity of training loads, highlights the dynamic nature of fatigue.
- Using EWMA helps to manage this complexity.

$$\text{EWMA}_{\text{today}} = \text{Feature}_{\text{today}} \cdot \lambda_a + (1 - \lambda_a) \cdot \text{EWMA}_{\text{yesterday}}$$

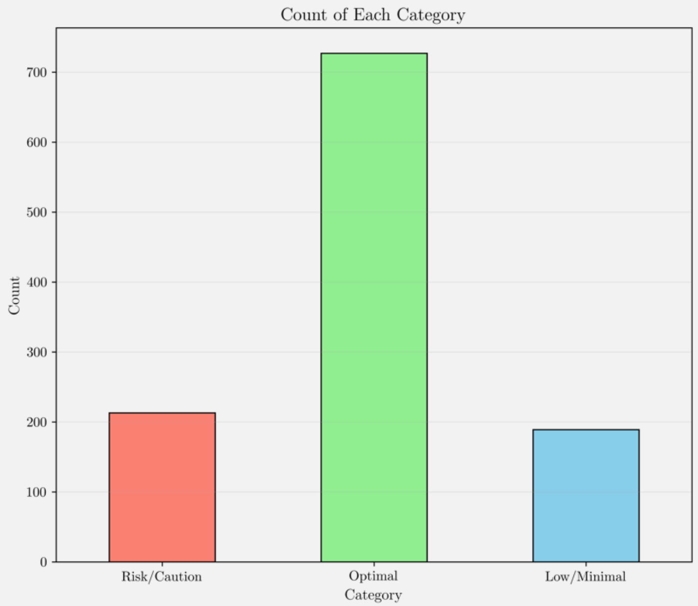
$$\lambda_a = \frac{2}{N+1}$$

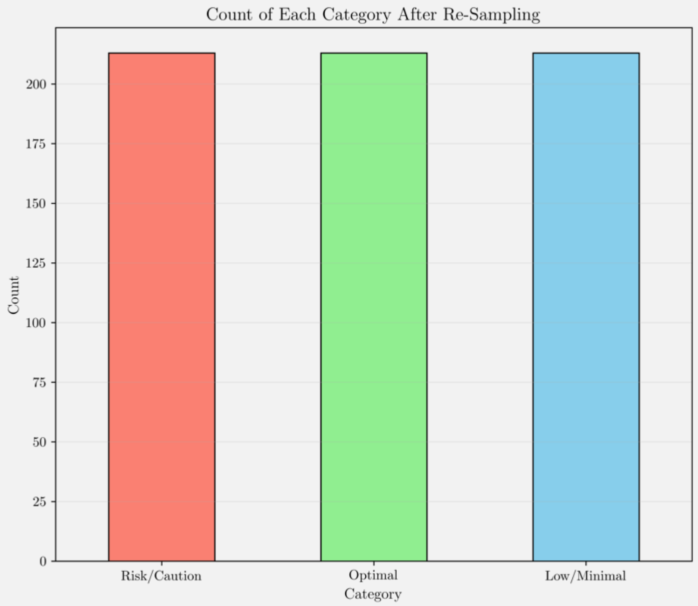


universidade de aveiro

OVERCOMING DATA CHALLENGES

Range	Initial Classes	Final Classes
≥ 90	Risk	Risk/Caution
≥ 80	Caution	
≥ 40	Optimal	Optimal
< 40	Low/Minimal	Low/Minimal



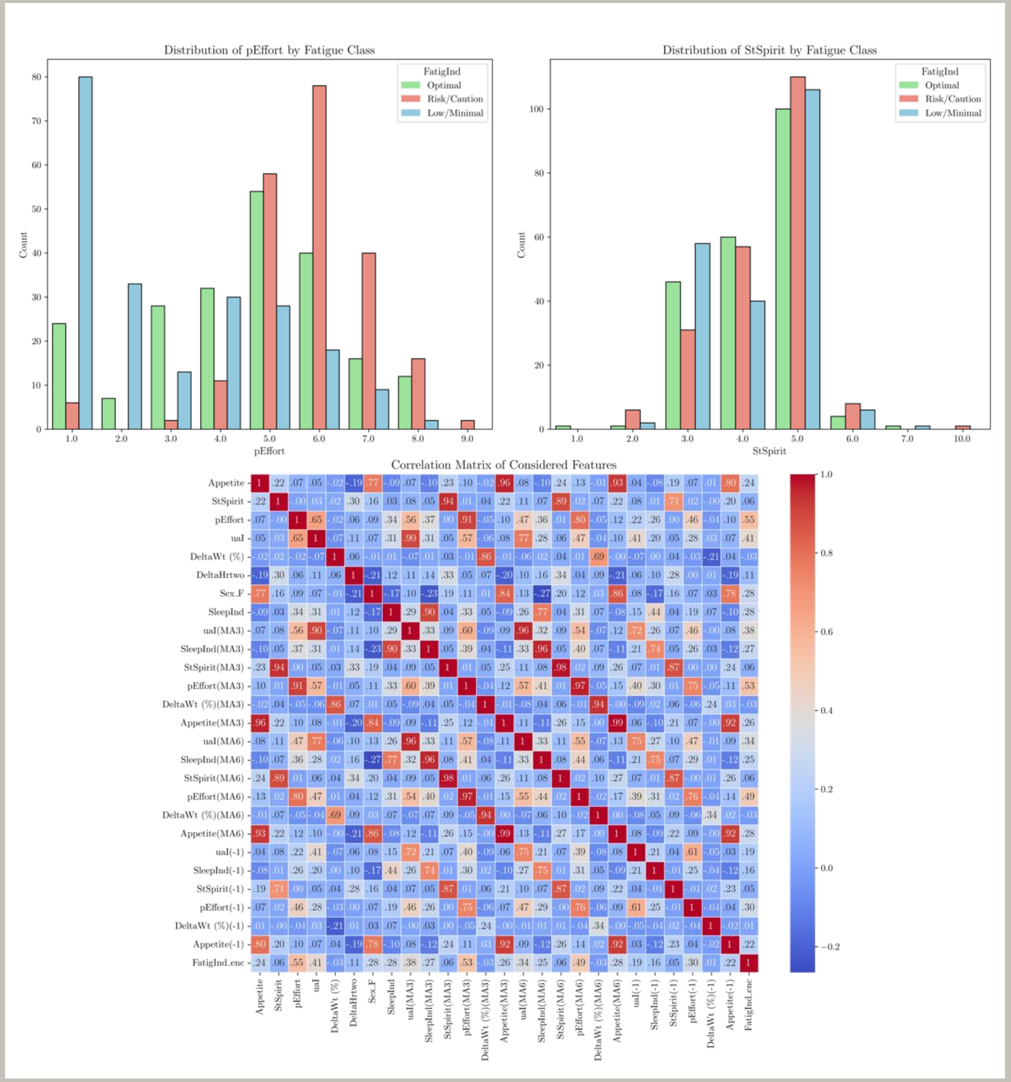


~~SMOTE and random undersampling~~

Random oversampling and undersampling

FEATURE SELECTION

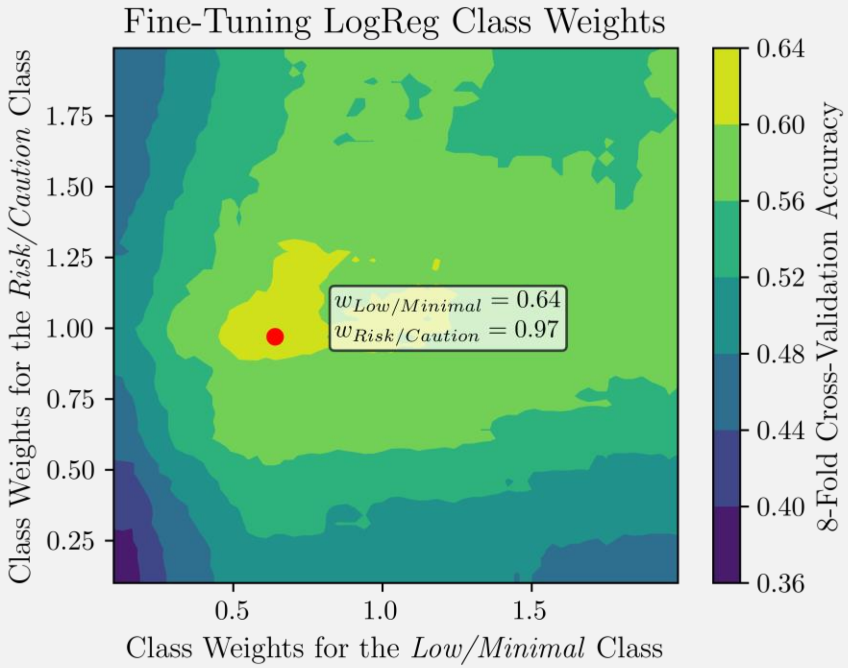
Feature	Description
Sex_F	Athlete's gender.
pEffort	Perceived effort from the workout.
uaI	Intensity from each workout.
SleepInd	Index based on quality of sleep and time in bed.
Appetite(MA6)	Appetite measure, averaged with EWMA(6).
pEffort(MA6)	*pEffort* averaged with EWMA(6).
uaI(MA6)	*uaI* averaged with EWMA(6).
SleepInd(MA6)	*SleepInd* averaged with EWMA(6).



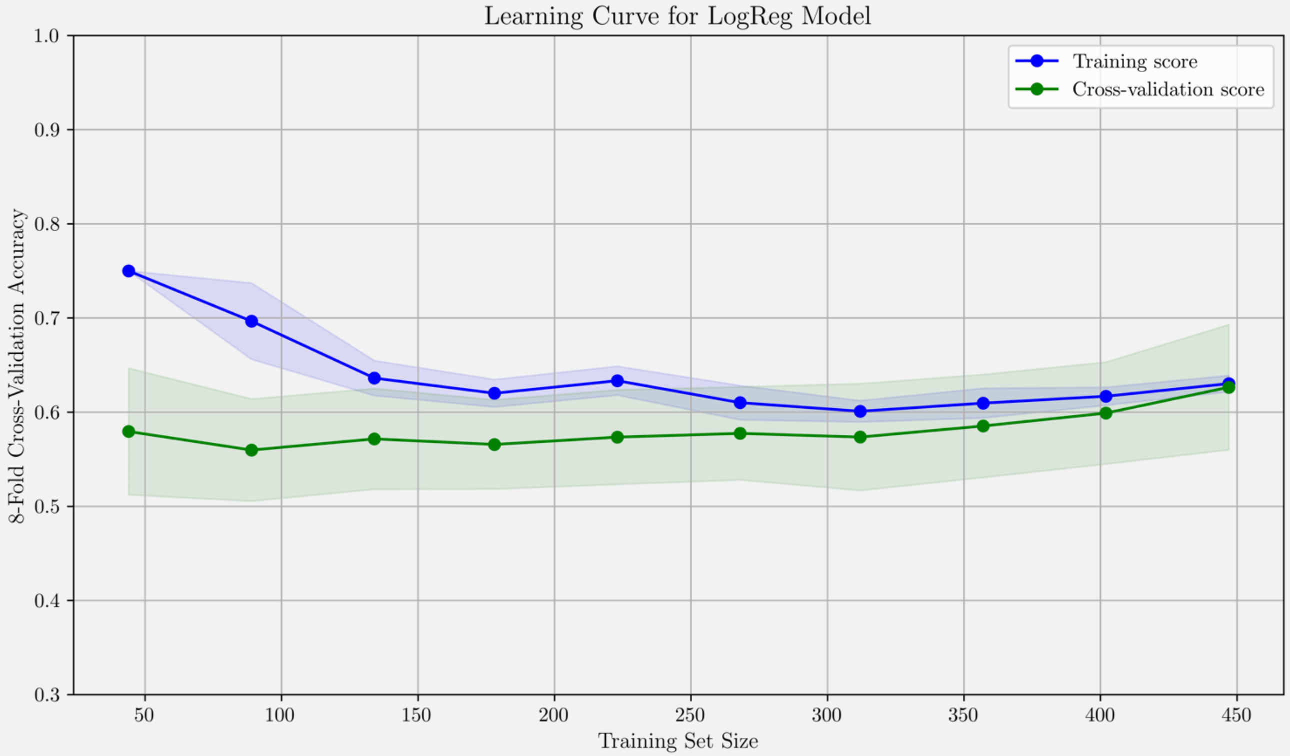
universidade de aveiro

MACHINE LEARNING MODELS

LOGISTIC REGRESSION

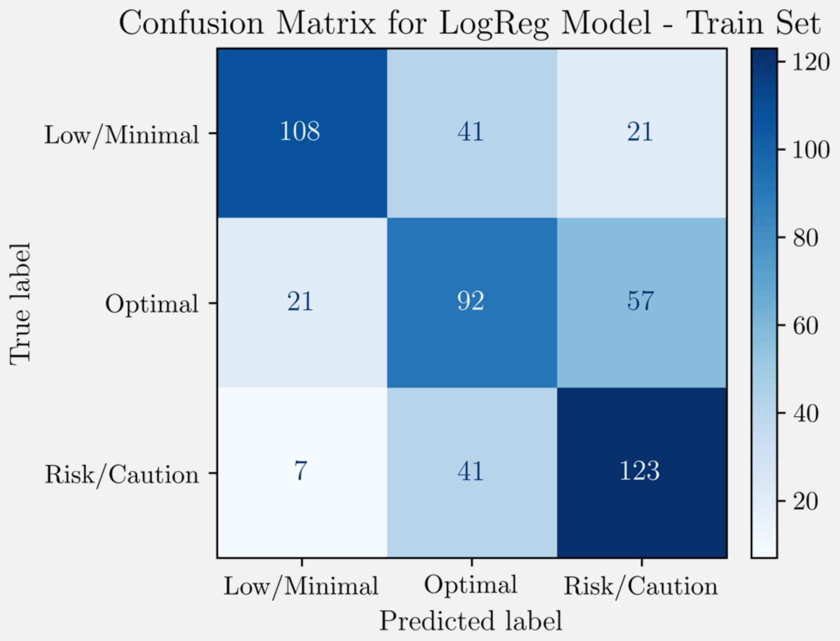


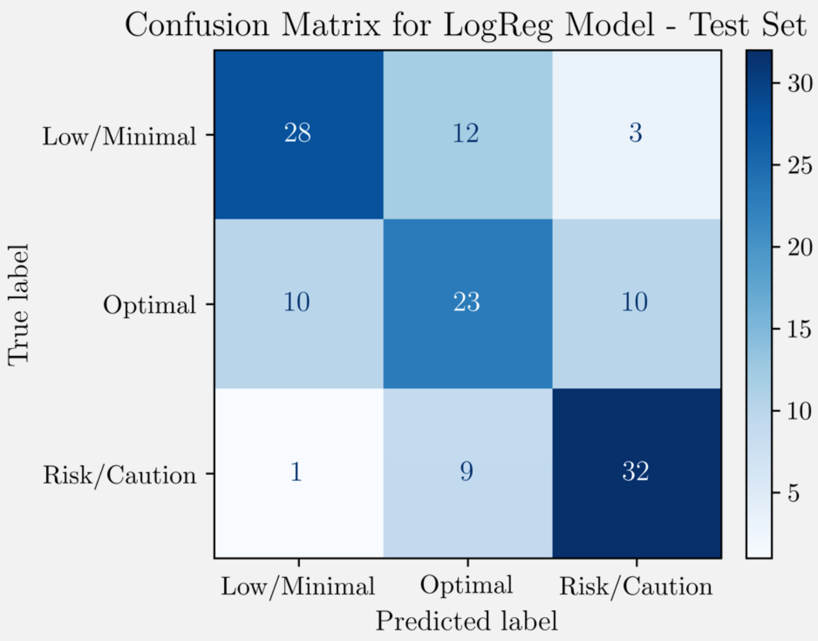
Hyperparameter	Possible Values	Best Value
C	$[0.01, 300]$	2.13
Regularization	$\{L1,\ L2,\ \text{none}\}$	$L1$



universidade de aveiro

LOGREG RESULTS





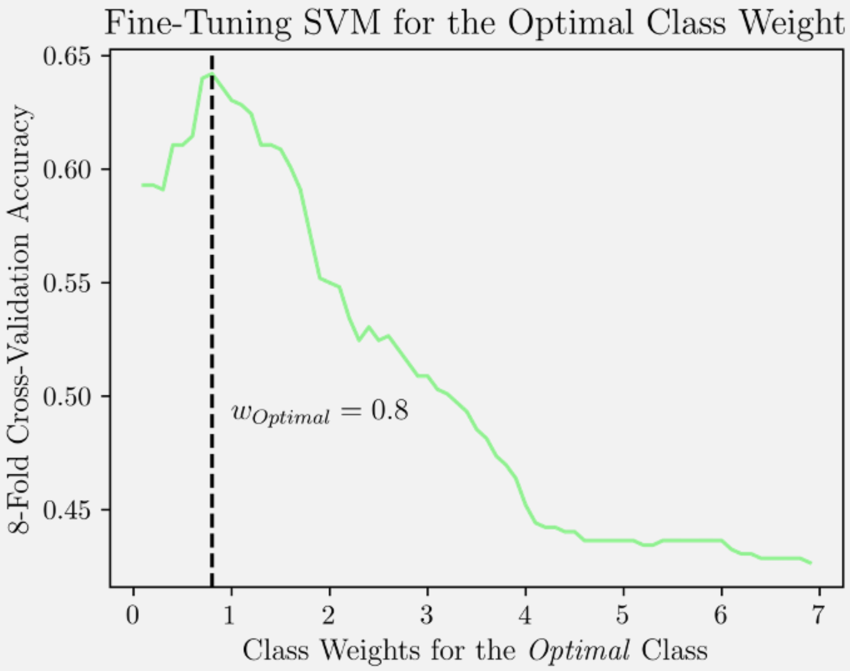
Classification Report: Train Set

Class	Precision	Recall	F1-Score	Support
Low/Minimal	0.79	0.64	0.71	170
Optimal	0.53	0.54	0.53	170
Risk/Caution	0.61	0.72	0.66	171
Accuracy			0.63	511
Macro avg	0.64	0.63	0.63	511
Weighted avg	0.64	0.63	0.63	511

Classification Report: Test Set

Class	Precision	Recall	F1-Score	Support
Low/Minimal	0.72	0.65	0.68	43
Optimal	0.52	0.52	0.53	43
Risk/Caution	0.71	0.76	0.74	42
Accuracy			0.65	128
Macro avg	0.65	0.65	0.65	128
Weighted avg	0.65	0.65	0.65	128

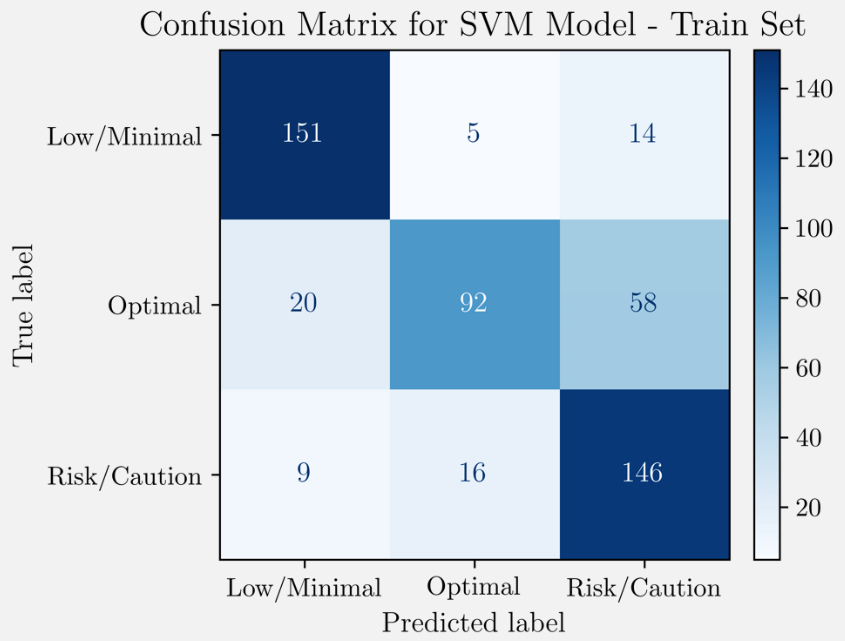
SUPPORT VECTOR MACHINE

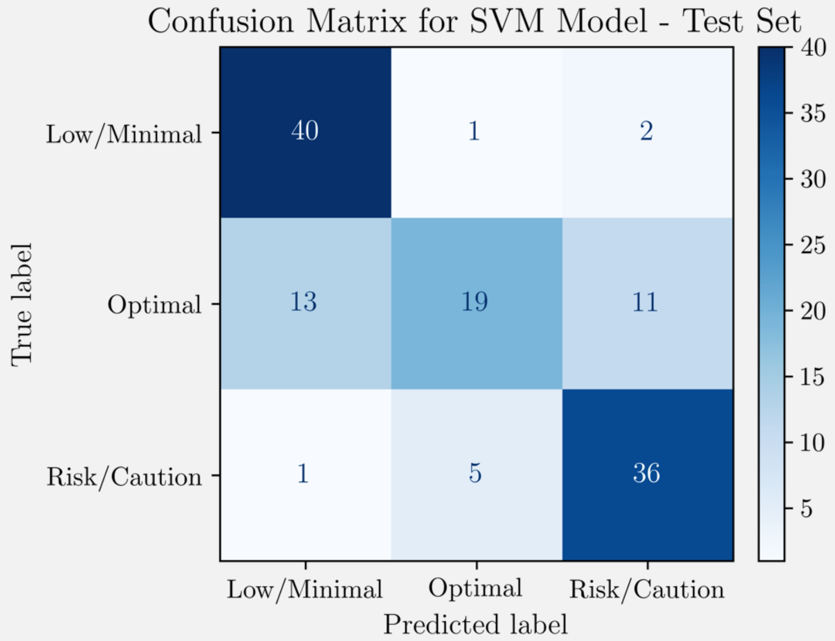


Hyperparameter	Possible Values	Best Value
C	$[0, 100]$	6.93
γ	{scale, auto, 0.1, 0.01, 0.001}	auto
Kernel	{linear, rbf, poly, sigmoid}	rbf
Degree	{1, 2, 3}	
Coef_0	$[-5, 5]$	

universidade de aveiro

SVM RESULTS





Classification Report: Train Set

Class	Precision	Recall	F1-Score	Support
Low/Minimal	0.84	0.89	0.86	170
Optimal	0.81	0.54	0.65	170
Risk/Caution	0.67	0.85	0.75	171
Accuracy			0.76	511
Macro avg	0.77	0.76	0.75	511
Weighted avg	0.77	0.76	0.75	511

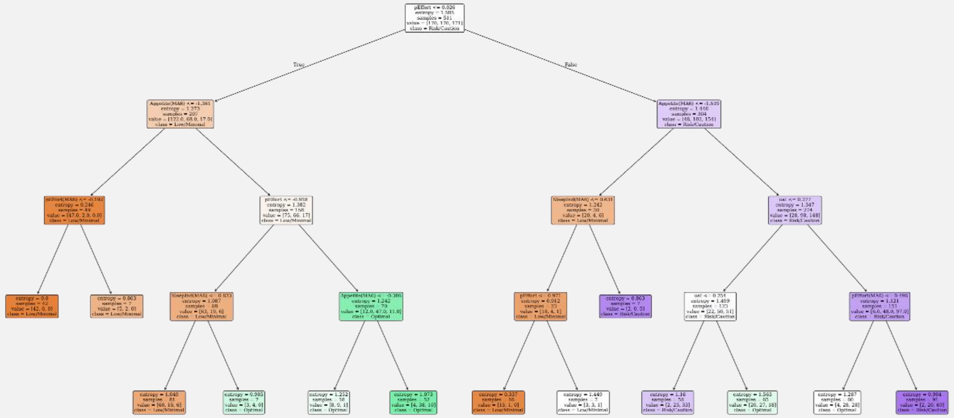
Classification Report: Test Set

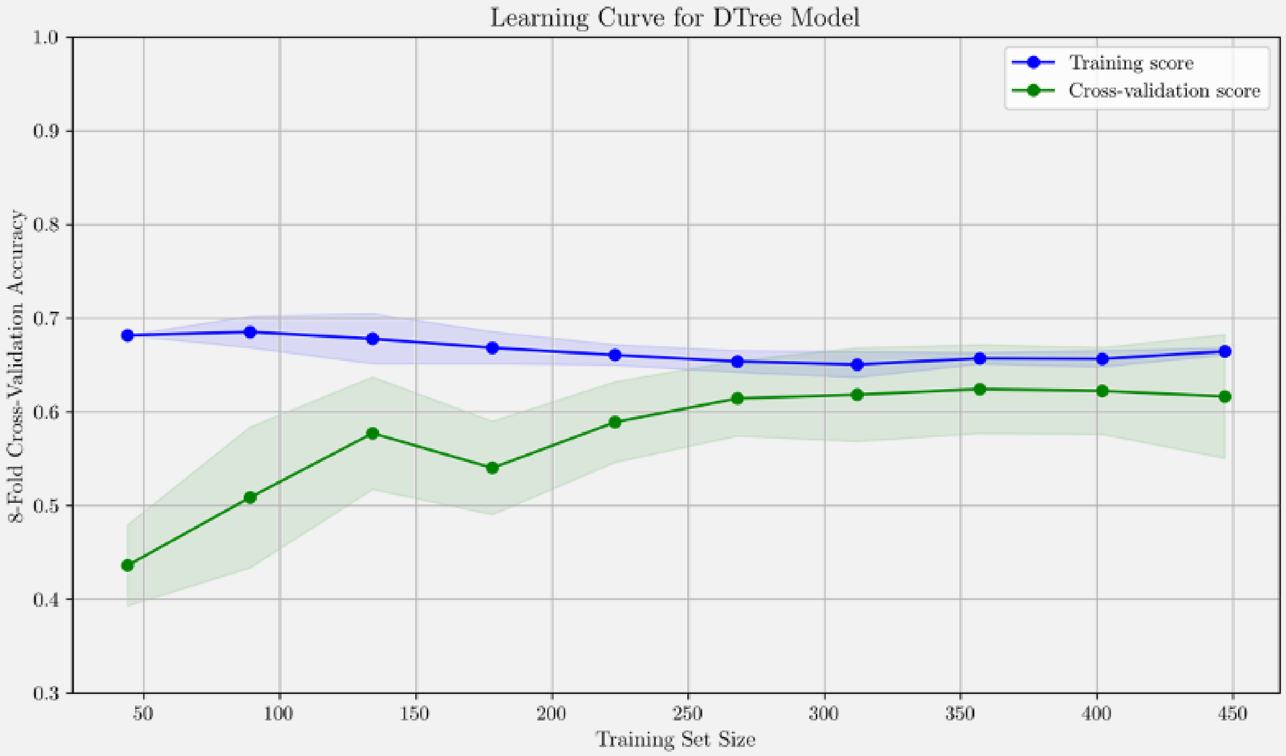
Class	Precision	Recall	F1-Score	Support
Low/Minimal	0.74	0.93	0.82	43
Optimal	0.76	0.44	0.56	43
Risk/Caution	0.73	0.86	0.79	42
Accuracy			0.74	128
Macro avg	0.75	0.74	0.72	128
Weighted avg	0.75	0.74	0.72	128

universidade de aveiro

DECISION TREE

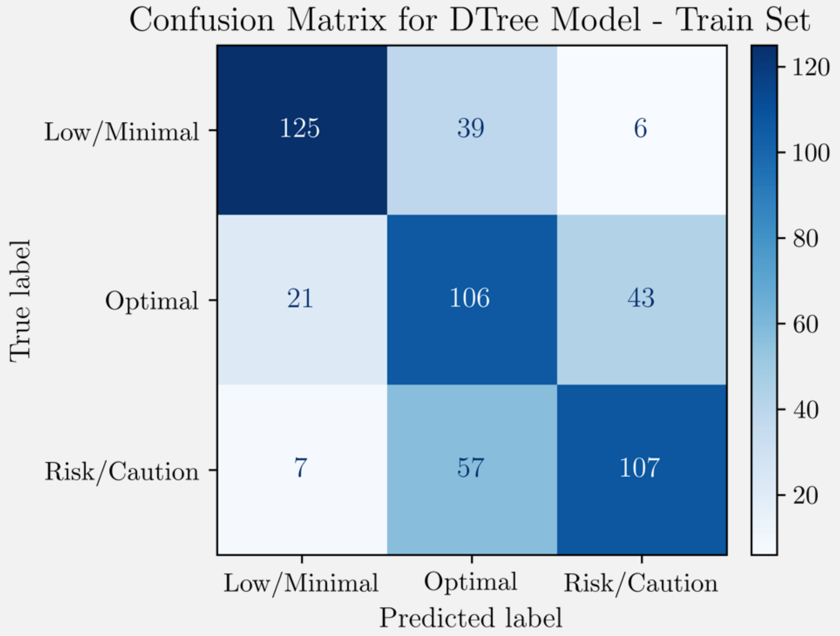
Hyperparameter	Possible Values	Best Value
Split Criterion	{gini, entropy}	entropy
Max Depth	$[2, 3, \ldots, 8]$	4
Min Samples to Split	$[5, 6, \ldots, 20]$	11
Min Samples per Leaf	$[3, 4, \ldots, 10]$	7

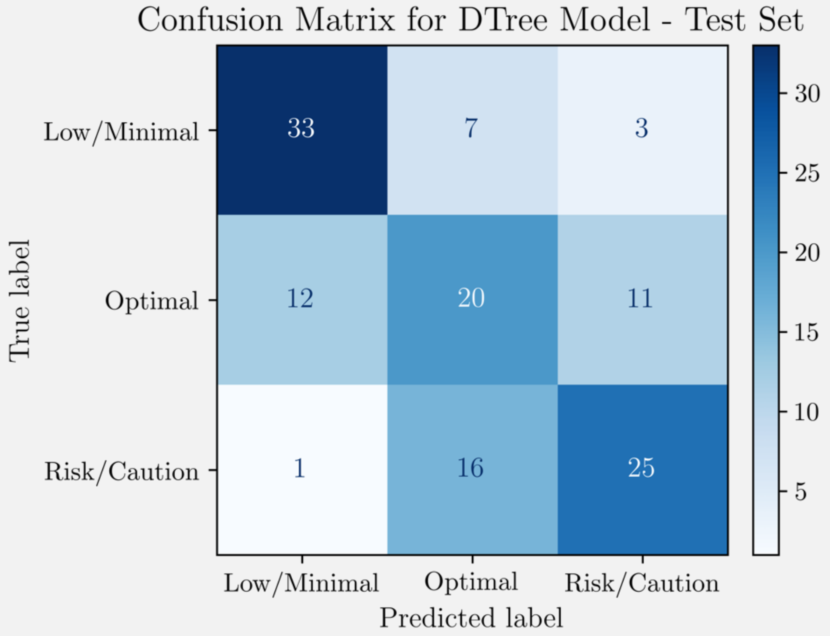




universidade de aveiro

DTREE RESULTS





Classification Report: Train Set

Class	Precision	Recall	F1-Score	Support
Low/Minimal	0.82	0.74	0.77	170
Optimal	0.52	0.62	0.57	170
Risk/Caution	0.69	0.63	0.65	171
Accuracy			0.66	511
Macro avg	0.68	0.66	0.67	511
Weighted avg	0.68	0.66	0.67	511

Classification Report: Test Set

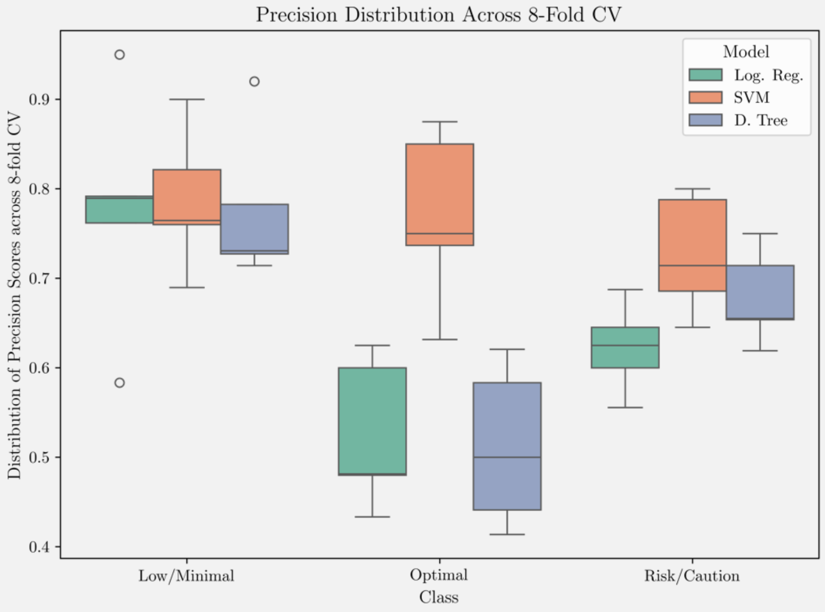
Class	Precision	Recall	F1-Score	Support
Low/Minimal	0.72	0.77	0.74	43
Optimal	0.47	0.47	0.47	43
Risk/Caution	0.64	0.60	0.62	43
Accuracy			0.61	128
Macro avg	0.61	0.61	0.61	128
Weighted avg	0.61	0.61	0.61	128

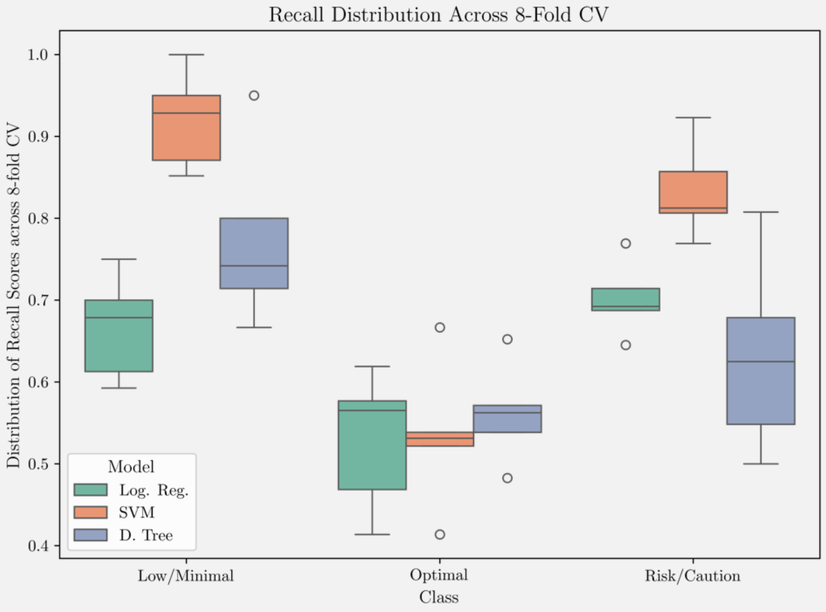
universidade de aveiro

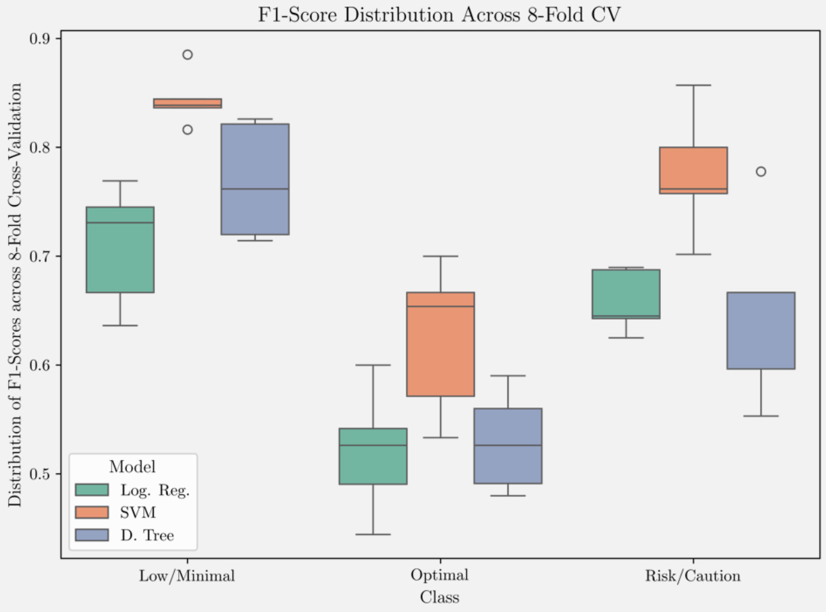
RESULTS ANALYSIS

universidade de aveiro

PERFORMANCE METRICS

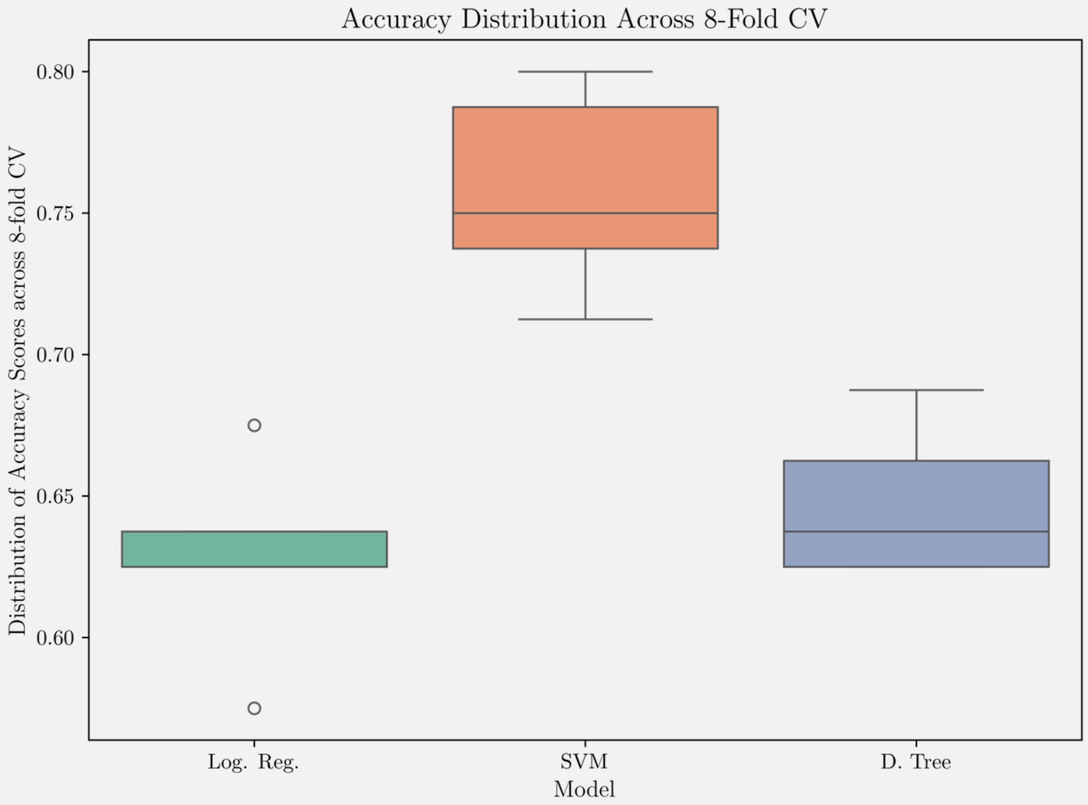






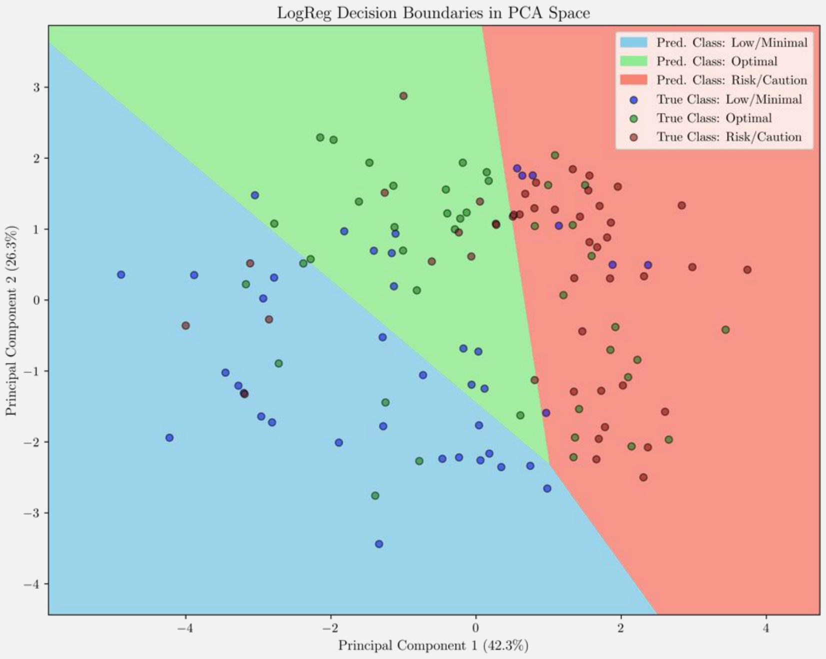
universidade de aveiro

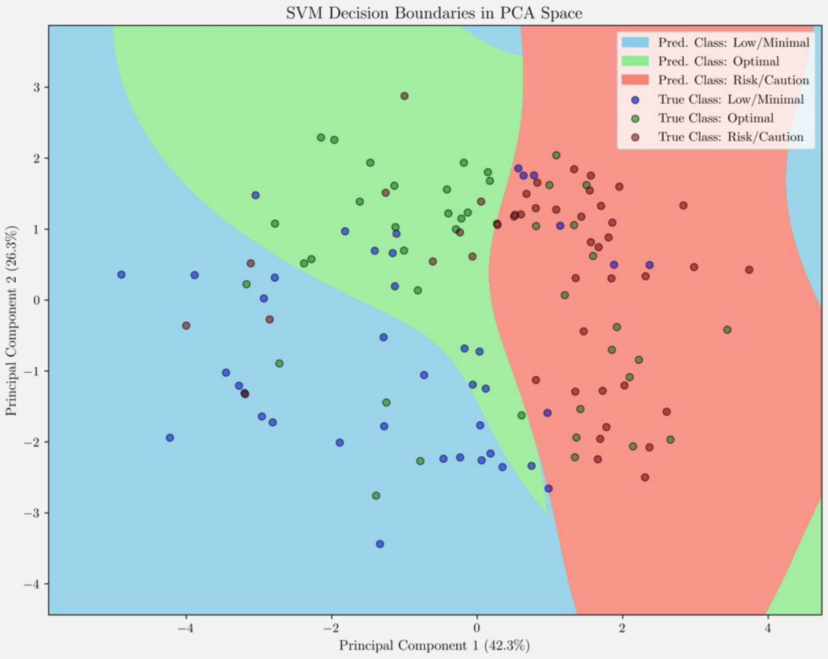
PERFORMANCE METRICS

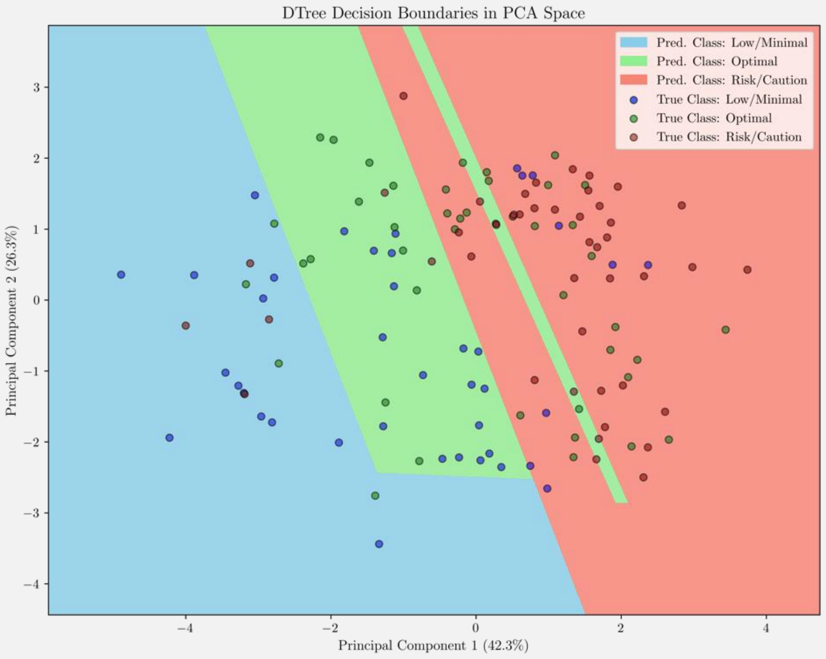


DECISION BOUNDARIES

PCA was applied to approximate the decision boundaries of each model within a 2D space, where 31% of the variance was lost due to dimensionality reduction, providing a rough visualization of the models' performance.





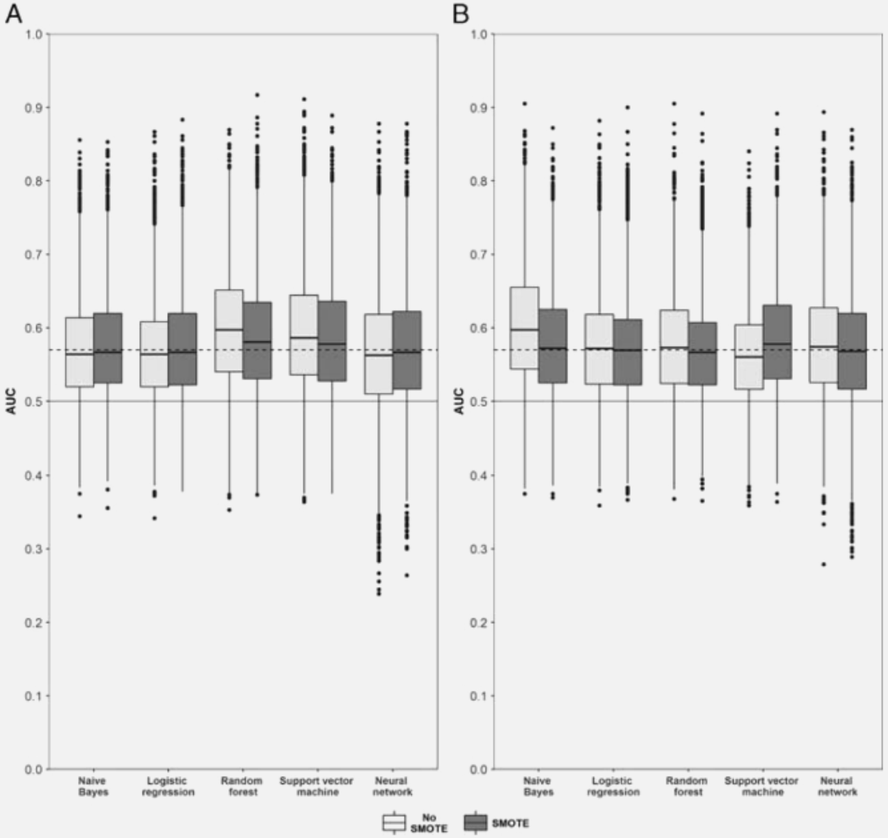


universidade de aveiro

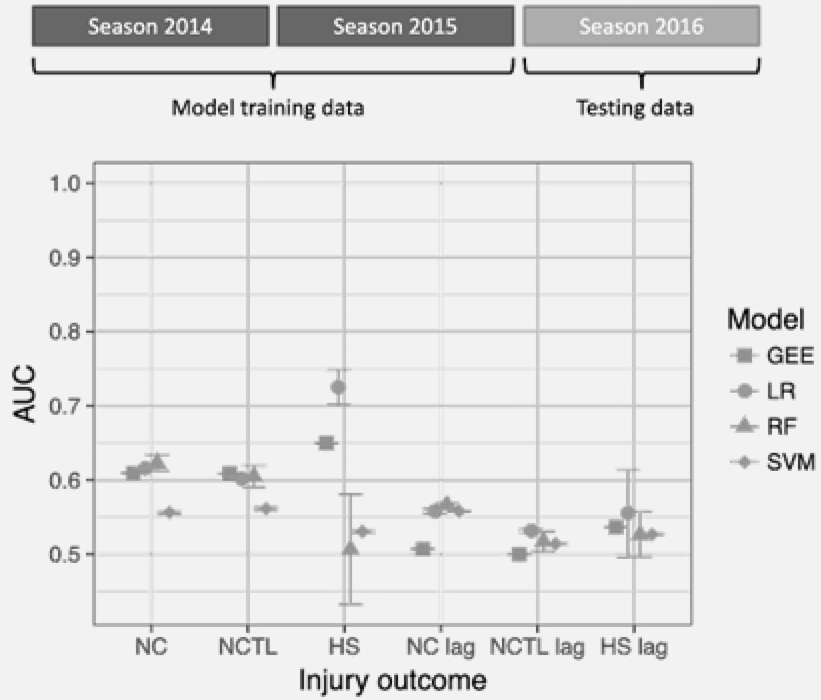
LITERATURE BENCHMARK

- Most literature focus on injury prediction
- Performance criteria not directly comparable: most of the problems are binary (injury or no injury)
- Models developed are well aligned with those in the literature.

"Predictive modeling of hamstring strain injuries in elite Australian footballers," 2017, Rudy et al.



"Predictive modelling of training loads and injury in Australian football", 2018, Carey et al.



universidade de aveiro

CONCLUSIONS

- Despite the reduced dataset, it was possible to implement and assess different machine learning models.
- Of the three models tested (LogReg, SVM, DTree), the support vector machine model stood out as being the most consistent across the board.
- The model's performance aligns closely with published results.
- The model will be implemented in Excel format and shared with the Head Coach Daniel Tavares.
- Future work: follow-up on the implementation and improve data collection for further refinement.