

# Applying ML to assess fatigue and prevent injury in high performance swimming athletes

Hugo Veríssimo

Foundations of Machine Learning 24/25  
University of Aveiro  
Aveiro, Portugal  
hugoverissimo@ua.pt

João Cardoso

Foundations of Machine Learning 24/25  
University of Aveiro  
Aveiro, Portugal  
joaopcardoso@ua.pt

**Abstract**—Improving performance of elite athletes is the ultimate goal of any sports coach. For proper planning and continued improvement, a good balance between workload and rest is crucial. There are several metrics used to assess the impact of the workload that athletes are subject of, with the rate of perceived exertion being amongst the most relevant throughout decades of research in the field. For this project the authors used data from the 2019/2020 season from the local swimming club CAPGE, provided by Head Coach Daniel Tavares, in order to develop machine learning models for fatigue prediction. The data consisted of athlete's feedback and training load, of which a set of features was selected and average over different periods. The models used were logistic regression, support vector machine, and decision tree. Overall the models performed decently, with support vector machine standing out as the best performing. The model developed was documented and shared in Excel format for practical application, and will be followed up with further data collection.

**Keywords:** swimming, fatigue prediction, LogReg, SVM, DTree

## I. INTRODUCTION

The evolution in performance of high level athletes is highly dependent on their skill, motivation, and discipline. With the support of a knowledgeable coach, the evolution can be substantially improved, through careful tailoring of the training regimen. One of the most relevant metrics since the dawn of structure training in sports is the feedback from the athlete, commonly described as the rate of perceived exertion (RPE). This single metric comprises the athletes analysis and intuition of the effort that was carried and how ready they feel for another bout of training. In recent years, more and more sports coaches have relied heavily on collected data to better assess, plan and adjust the training plans of their athletes in a systematic way. This allows for a fine balance between intense workouts, that generate stronger stimuli for muscle development and sport specific skills, taking the balance between effort and fatigue as the crucial ratio to respect. Too high effort, may lead to injury, too low and some gains may be left on the table.

In the scope of the first project for Foundations of Machine Learning, we decided to partner with the local sports club CAPGE (Clube Associação de Pais da Gafanha da Encarnação) to treat the data (kindly shared by Head Coach Daniel Tavares) for estimating the fatigue of several athletes.

The data was curated and prepared to implement and fit several machine learning algorithms to estimate fatigue after workout.

With this approach, we aim to generalize the models for different athletes/sports, and make it available to the local club for implementation and further testing.

## II. STATE OF THE ART

Over the past decade there have been significant improvements in the field of ML applied algorithms for sports' related applications. Here we present the most relevant work in fatigue and injury prediction, which is not specifically on the sport we're assessing in our dataset, which lead to additional interpretations from the works analyzed to our own case of study. In general, the problem of class imbalance is seen throughout the literature, and different solutions are proposed, such as data gathering and preprocessing, over sampling and under sampling, with SMOTE (synthetic minority oversampling technique) being the most commonly used approach for over sampling. As early as 2010, Gabbet and colleagues modeled the risk of injury with a monodimensional approach using logistic regression, based on athletes rate of perceived exertion, showing that even with a monovariate approach to injury prediction useful results could be attained. In recent years, several authors have focused in alternative techniques such as Logistic Regression, Random Forest, Support Vector Machine, or Convolutional Neural Network on Multivariate Timeseries (the references for these papers can be found in the paper "A Narrative Review").

Besides model selection, feature engineering and selection is among the most debated topics. Several authors opt to include GPS data, metabolic consumption, mechanical load, RPE, detailed quantification of workloads, ratio between acute:chronic loads. Despite the multivariate imputation, data analysis often shows strong correlations between them, leading to overfitting problems (usually model independent).

In the work by Carey et al. (2018) different algorithms have been implemented to predict the risk of injury in an Australian football club. The data collection lasted for three seasons, consisting of absolute and relative training load metrics, derived from GPS, accelerometer, and RPE data. The prediction models used were regularized logistic regression, generalized

estimating equations, random forests, and support vector machines, with periods of 3, 6, and 21 days (these periods have been studied and verified as adequate for the case of Australian football). The periods served to calculate moving averages and exponentially weighted moving averages (EWMA). The latter allowed to account for the decay in significance of the training load the further it happened from a given day, in accordance with the work from Williams et al. (2016). From the results it was possible to verify that overfitting was very likely due to the multicollinearity between variables, which was confirmed by principal component analysis (PCA). The use of PCA with regularized logistic regression slightly improved the results.

More recent studies have employed ensemble algorithms, in order to take most of the different learning models selected, taking into account the need to balance the classes as is common practice for this type of problems.

In summary, the integration of machine learning techniques in sports fatigue and injury prediction has evolved from simple monovariate models to complex multivariate and ensemble approaches. Addressing challenges such as class imbalance, feature selection, and multicollinearity remains crucial for developing robust predictive models applicable across different sports contexts.

### III. METHODOLOGY

The methodology defined for this project consisted of three major steps: assessment and curation of the dataset, feature engineering and selection of features, followed by data normalization; training of the selected machine learning models; finalizing with model evaluation and subsequent training until optimum results were achieved. Figure 1 illustrates the methodology used in this project.

The programming language used was Python, and the packages available therein, with notable mention for scikit-learn. After data normalization for all features StandardScaler, the data was separated between training and testing data (80/20, respectively), using the train\_test\_split function from sklearn. The seeds for randomization were kept consistent across models, to ensure reproducibility, and avoid biases towards any model. The ML models selected for the project were LogisticRegression, SupportVectorMachine, and DecisionTreeClassifier, where the modeling approach and hyperparameters are detailed below. For the given hyperparameters available in each model the function RandomizedSearchCV was used with 8-fold cross-validation to minimize the risk of overfitting (the selected 8-fold CV was consistent throughout all the relevant stages, class weight estimation, training).

The metrics used to assess the different models are presented in Table I, and are consistent with those used in the literature.

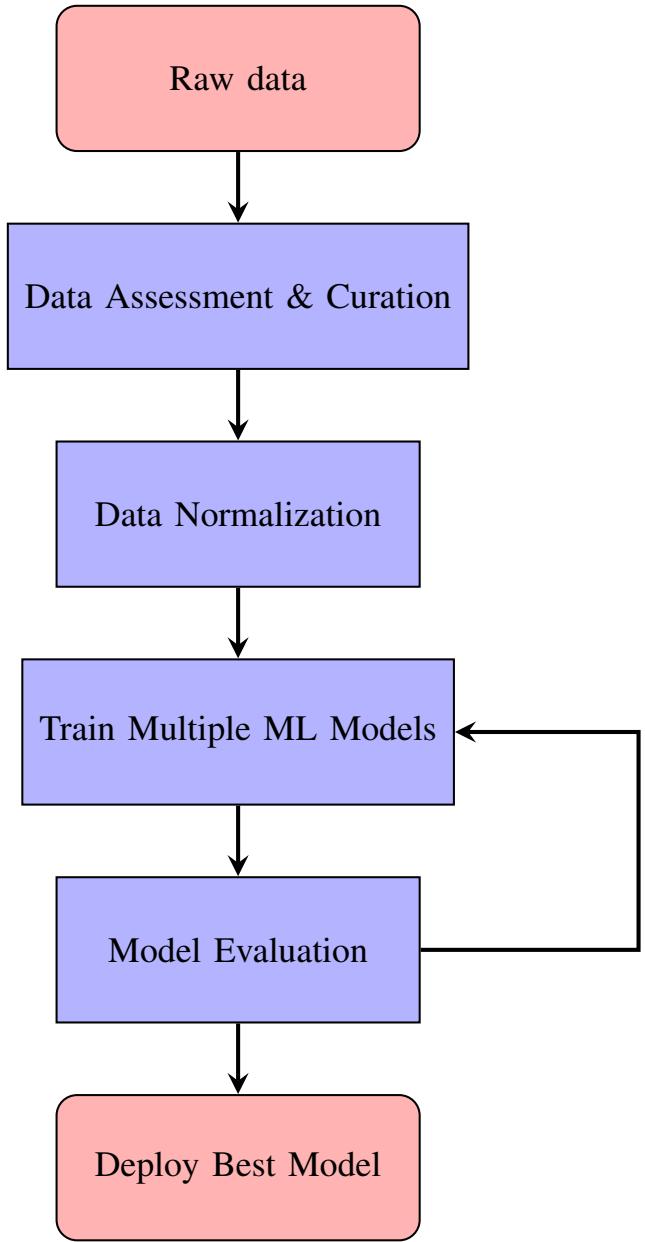


Fig. 1: Flowchart of the methodology for data processing and modeling. The evaluation results are used to refine the model training process.

TABLE I: Metrics for multiclass classification model evaluation.

Measure	Formula
Precision (per class $i$ )	$\frac{TP_i}{TP_i + FP_i}$
Recall (per class $i$ )	$\frac{TP_i}{TP_i + FN_i}$
F1-score (per class $i$ )	$2 \cdot \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$
Accuracy	$\frac{\sum_{i=1}^C TP_i}{\sum_{i=1}^C (TP_i + TN_i + FP_i + FN_i)}$

Given the class imbalance associated with the problem at hand, the importance of precision, recall and F1-score per class are especially relevant, along with the confusion matrix, to understand how the model is failing to correctly classify the various observations. The precision is the ratio of true positives for all the positives attributed (the higher, the better). Recall (or sensitivity) gives the ratio of true positives among true positives and false negatives, where lower values indicate a higher number of misidentified true positives. The F1-score provides a balanced assessment of the model, taking a harmonic mean of precision and recall, providing a particularly good way to assess datasets with imbalanced classes, while accuracy provides an overall assessment with the ratio of true positives and true negatives over the total predictions.

The learning curve was used for all models, to highlight the relation between the size of the training dataset and the performance of the adjusted model, providing a good insight into the quality of the fitting. In the case of increasing accuracy of the training curve while the validation curve remains constant or decreases below the training curve, is a clear sign of overfitting. In the case that both curves converge, but to low values (below 0.6) is a sign of under fitting and poor learning from the model.

The confusion matrix was used to assess the performance in estimating each class during training and testing, giving a clear idea how the model is trying to predict the classes, helping to better interpret the class weights attributed.

The methods and ML models used are consistent with those in the literature, considering the type of features and target in this project. Moreover, during the model refinement stage a reassessment of past steps was carried in order to ensure no gaps in the process.

#### IV. DATASET ANALYSIS

##### A. Data Description

The data used in this project was collected from the swimming club CAPGE during the season of 2019/2020, where each athlete has several observations corresponding to training days, where each of the features was collected. Not all athletes logged the same number of training days, nor present an equal distribution between low, average and high levels of intense training. The names of the athletes were removed to ensure privacy and confidentiality, keeping only the gender as a variable. The team is comprised by seven athletes, three male and four female. Most of the features are related to feedback from the athletes on different aspects of their lives (i.e. sleep quality, appetite, and rate of perceived exertion after training), while others are measurable (i.e. workload, variation in heart rate before and after training, weight variation). A notable feature to mention is the RPE, that is still deemed as one of the most relevant metrics for workload planning and fatigue assessment. All these attributes are classified between 1 — 10, each value corresponding to increasingly 'worse' categories (e.g., 1 great appetite / normal, 10 no appetite at all).

The fatigue index is calculated from these features, using weights attributed by the coach based on his empirical experi-

ence. The resulting fatigue index is between 0 — 100, which was categorized in four classes as seen in Table II.

TABLE II: Classification of fatigue index into categories based on numerical ranges.

Range	Initial Classes	Final Classes
$\geq 90$	Risk	Risk/Caution
$\geq 80$	Caution	
$\geq 40$	Optimal	Optimal
$< 40$	Low/Minimal	Low/Minimal

There is a big gap between fatigue classes due to the nature of training and performing high effort workouts in specific times of the training cycle. The dataset was provided in Excel format (per athlete), from which we imported and combined the data as a pandas DataFrame to apply the different models.

##### B. Dataset curation

The initial assessment evidenced the need for balancing our data. To start, we've reduced the number of classes, by combining the two higher risk classes ('Caution' and 'Risk'). With this, the number of observations was closer between 'Low/Minimal' and 'Risk/Caution', leaving us with an excess of observations for 'Optimal', as seen in Figure 2.

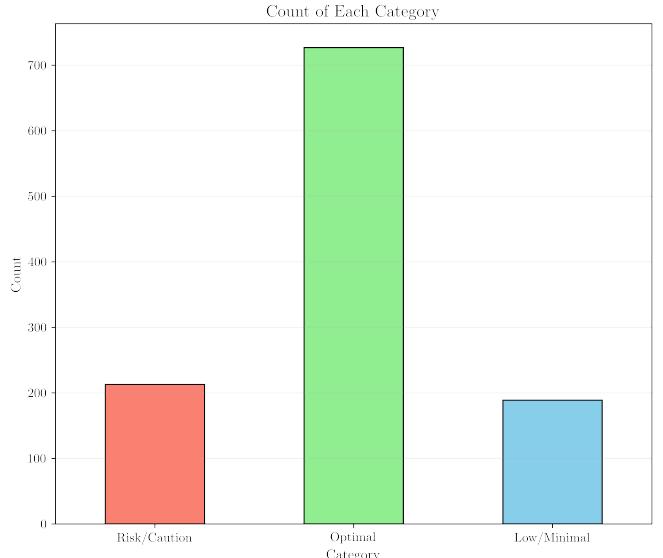


Fig. 2: Bar plot of the new classes, once 'Caution' and 'Risk' are combined into one.

At this stage, we opted to under sample our dataset to the number of observations of 'Risk/Caution', and over sample the observations in 'Low/Minimal', by imputing random samples from the pool of observations of 'Low/Minimal', ending up with 213 observations per class (regardless of gender). The use of SMOTE in this scenario would give continuous classes for our features, which wouldn't yield any physical meaning.

To assess how the different features vary among them and in relation to the target, we computed the correlation matrix as seen in Figure 3.

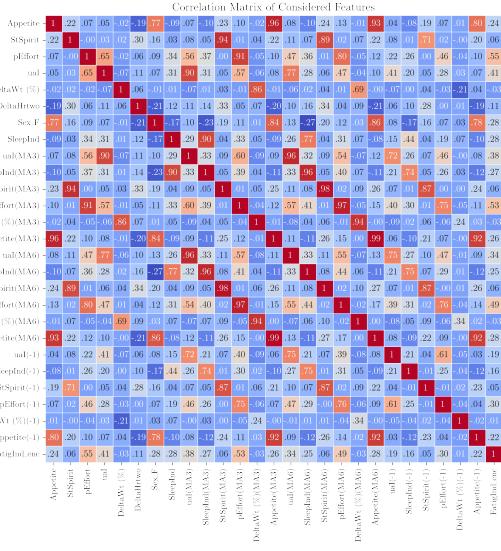


Fig. 3: Correlation matrix for all possible features considered for the models.

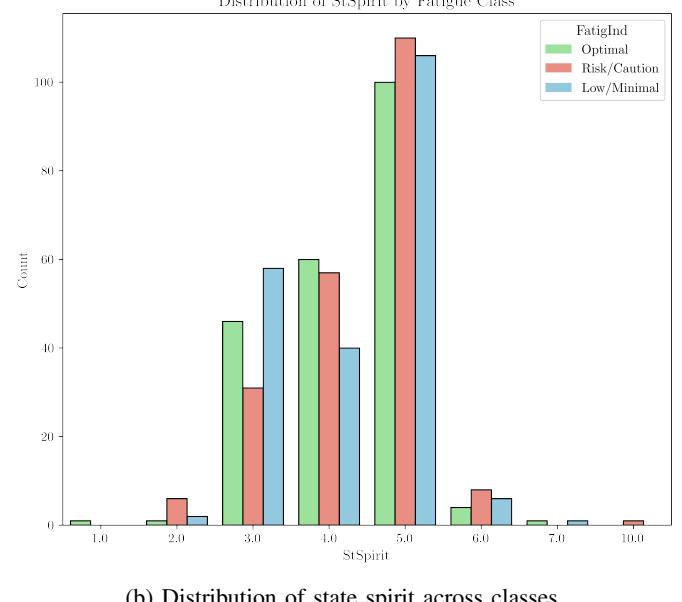
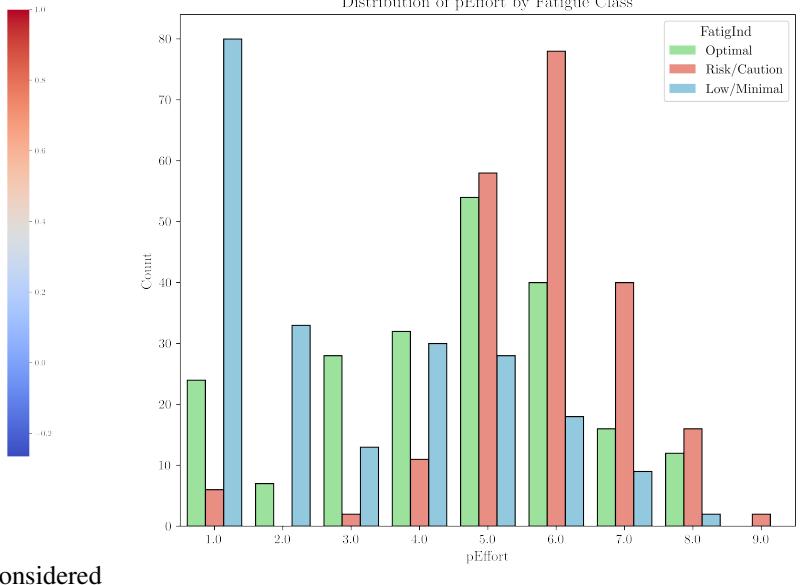


Fig. 4: Class distributions across different metrics: (a) perceived effort, (b) state spirit.

Considering that the weights used in the coach's original estimation of fatigue were identical regardless of sex, we performed some simple models in order to decide if it would be necessary to split it. We could verify that gender didn't have a significant impact in model performance, so we opted to use it as a feature.

Figure 5 illustrates the periodicity of higher training loads and subsequent lower intensity periods. It is important to refer that once the two higher intensity classes were combined the loss of granularity of how fatigue changes throughout the season is evident.

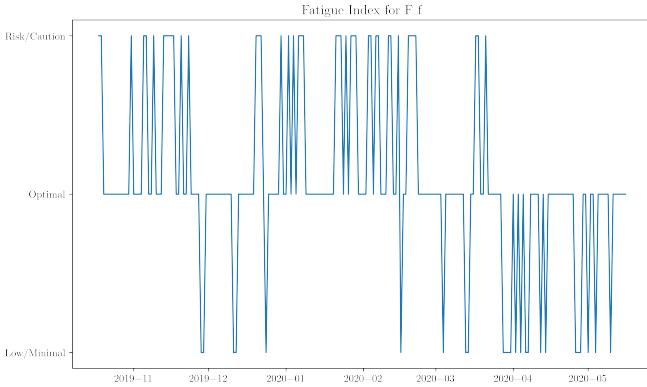


Fig. 5: Daily fatigue trends for athlete F\_f.

In order to keep the time series nature of training, increase and decrease of training intensity, and varying fatigue with training we included exponentially weighted moving averages (EWMA) [1].

This way memory is introduced in our model, and allow to consider the decay in the weights of events further away from any given day. The mathematical expression used to calculate EWMA for each selected feature is,

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

where  $\lambda_a$  is a value between 0 and 1 that represents the degree of decay, with higher values discounting older observations at a faster rate. The  $\lambda_a$  is given by:

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

At this point it was possible to select the most relevant features for modeling: *pEffort*, *uaI*, *SleepInd*, *Sex\_F*, *pEffort(MA6)*, *SleepInd(MA6)*, *ual(MA6)*, *Appetite(MA6)*.

## V. CLASSIFICATION MODELS

In this section the results for the three machine learning models are presented and detailed in equal fashion, to ease the discussion and interpretation in the next section. As was mentioned in the methodology, the modeling approach was the same for all, and here the training dataset results are presented first, followed by the test dataset results.

## VI. LOGISTIC REGRESSION

The logistic regression model was developed by setting different ranges for the hyperparameters, with  $C$ , the inverse of the regularization parameter, varying between 0.01 and 300, and allowing the selection of different cost functions ( $L1$ ,  $L2$ , or none). Due to the suboptimal performance of the 'Optimal' class in terms of precision, we also decided to adjust the class weights for 'Risk/Caution' and 'Low/Minimal' within a range of 0.1 to 2. The resulting model with the highest accuracy is illustrated in Figure 6.

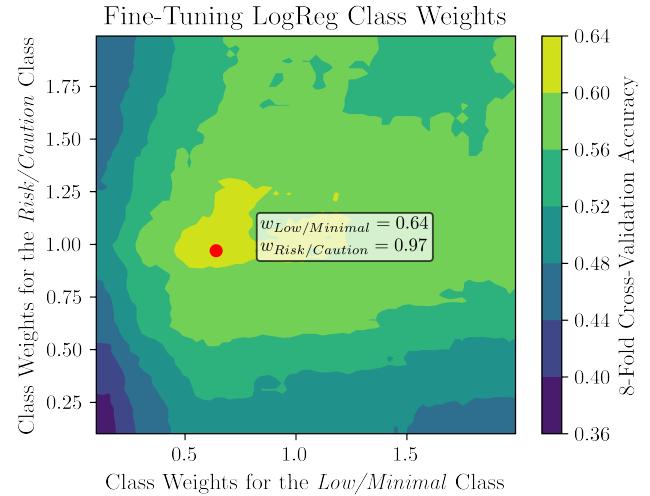


Fig. 6: Effect of the 'Risk/Caution' and 'Low/Minimal' classes weight on LogReg model accuracy.

Since the highest accuracy was achieved with class weights of 0.64 for 'Low/Minimal' and 0.97 for 'Risk/Caution', these values were selected, while the original weight for the 'Optimal' class was retained. These weight values were applied consistently across both the training and test datasets. For model optimization, the remaining hyperparameters were set to  $C \approx 2.13$ , the cost function was  $L1$  (Lasso regularization) and the solver used was SAGA (Stochastic Average Gradient Augmented). Despite the fact that it is most commonly used for large datasets, it was the best performing kernel of those available in the initial assessment, with the advantage that it allowed for regularization. The equation for the solver is given by,

$$\min_w \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y_i X_i^\top w)) + \lambda R(w) \quad (1)$$

The learning curve shown in Figure 7 illustrates the relationship between the training set size and the performance of the adjusted model, providing valuable insight to assess the model's behavior in terms of overfitting or underfitting. In this case, the learning curves from the training and the cross-validation converge after the largest training set size of 450, with no signs of overfitting.

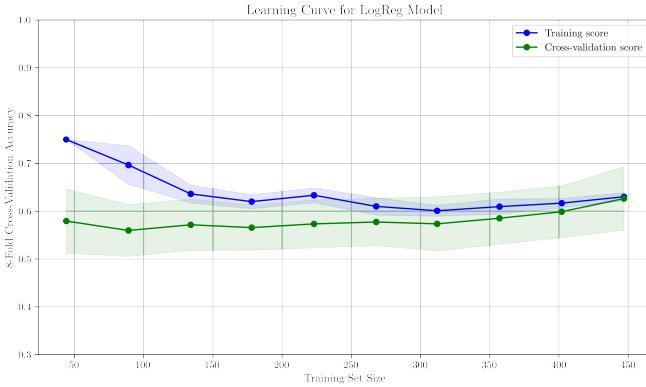


Fig. 7: LogReg model performance using learning curve representation across varying training data sizes.

Considering the confusion matrices of the training (Fig. 8) and test datasets (Fig. 9), it is possible to learn how well the model attributes true and false positives, and how they are distributed. For this particular model it is visible how similar they are, and how much better the prediction performance is for the classes in the extremities ('Low/Minimal' and 'Risk/Caution'), regardless of the weight attributed to the 'Optimal' class (even though a marginal improvement was observed compared to the initial weightless model).

The classification report provides a more straight forward comparison between the training (Table III) and the test datasets (Table IV), while showing class specific the performance metrics, further confirming the interpretation of the confusion matrix. Both classification reports present similar values, which is a good indicator that the model is properly fitted. As previously mentioned, the performance metrics are worse for 'Optimal' when compared to the remaining classes.

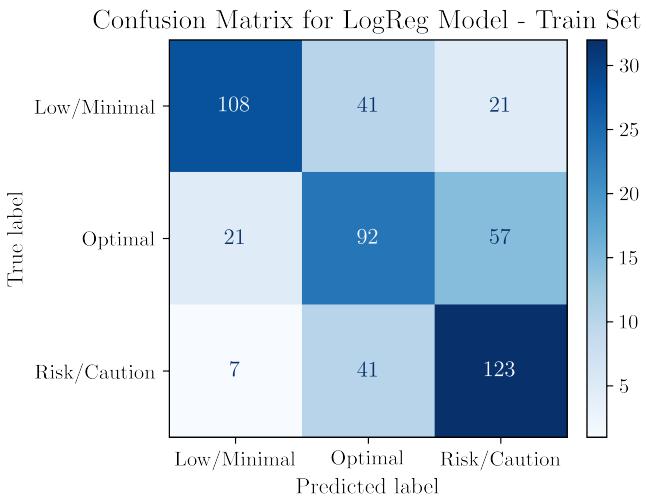


Fig. 8: Confusion matrix for the training data of LogReg model.

TABLE III: Classification report for LogReg model performance evaluation on training data.

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
<b>Accuracy</b>			0.63	511
<b>Macro avg</b>	0.64	0.63	0.63	511
<b>Weighted avg</b>	0.64	0.63	0.63	511

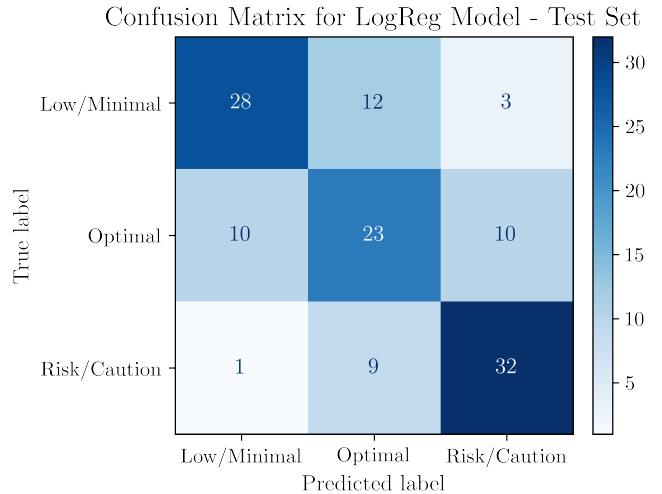


Fig. 9: Confusion matrix for the test data of LogReg model.

TABLE IV: Classification report for LogReg model performance evaluation on test data.

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
<b>Accuracy</b>			0.65	128
<b>Macro avg</b>	0.65	0.65	0.65	128
<b>Weighted avg</b>	0.65	0.65	0.65	128

## VII. SUPPORT VECTOR MACHINE

Given the increased complexity compared to LogReg, the support vector machine (SVM) model allows for manipulation of a larger number of hyperparameters, which consequently lead to longer computation times to achieve the best model. The hyperparameters used were Regularization Parameter (parameter that controls the penalty for misclassified training examples, i.e., cost function) ( $C$ ), the Kernel Coefficient ( $\gamma$ ), the kernel to be used, the highest degree possible (for a *poly* kernel), and the value of the independent term ( $\text{Coef}_0$ ), which controls the flexibility of the decision boundary [2]. The ranges of possible values can be assessed in Table V.

TABLE V: SVM model hyperparameters search space.

Hyperparameter	Possible Values
$C$	[0, 100]
$\gamma$	{scale, auto, 0.1, 0.01, 0.001}
Kernel	{linear, rbf, poly, sigmoid}
Degree	{1, 2, 3}
Coef <sub>0</sub>	[-5, 5]

As in the case for LogReg, an ideal class weight was estimated, however here only for 'Optimal' class, as it was the worst predicted class even at a training stage, with an equal split between 'Low/Minimal' and 'Risk/Caution'. The weight being smaller than 1 is penalizing the class, making it more likely to decide for one of the other classes when close to the decision boundary. The resulting class is then defined by,

$$C_{class} \leftarrow C \times w_{class} [2].$$

The weight for the 'Optimal' class that retrieved the highest accuracy was estimated as shown in Figure 10 ( $w_{Optimal} = 0.8$ ).

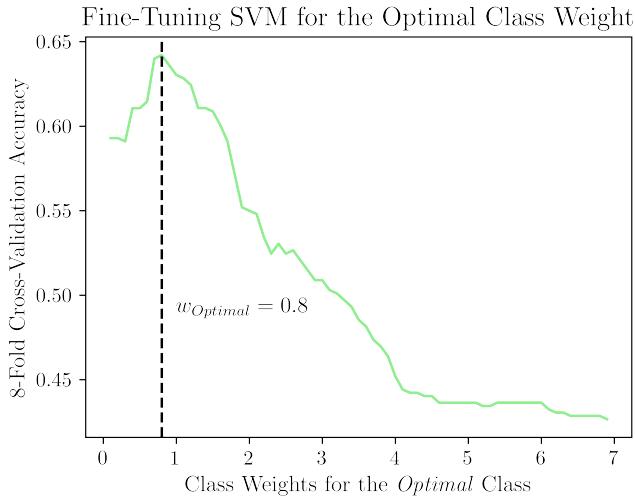


Fig. 10: Effect of the 'Optimal' class weight on SVM model accuracy.

The remaining classes kept the initially attributed unitary weight. The best parameters found, for this weights, were  $C$  6.93,  $\gamma$  auto ( $1/n_{features}$ ), and kernel RBF (Radial Basis Function).

$$\text{RBF kernel: } \exp(-\gamma \|x - x'\|^2)$$

In this case, the learning curves for both the training and cross-validation sets (Fig. 11) converge, with no signs of overfitting. However, it could be beneficial to expand the training dataset further, as the cross-validation score continues to improve. This suggests that the model might still benefit from more data, which could enhance its ability to generalize better.

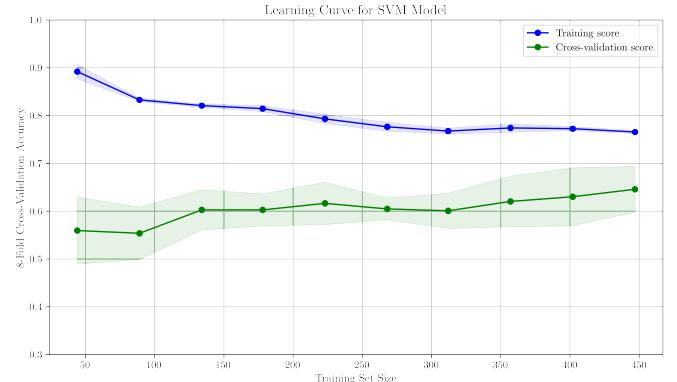


Fig. 11: SVM model performance using learning curve representation across varying training data sizes.

The confusion matrices for both training (Table 12) and testing datasets (Table 13) show a good performance in the classification of all classes, especially for those in the extremities (as already verified in the previous model).

The classification reports of the training (Table VI) and testing datasets (Table VII) show consistent results among them. However, it is noteworthy the low value of Recall for the 'Optimal' class in both cases, indicating several misclassification occurrences (observations that should be 'Optimal' but were classified as one of the other classes).

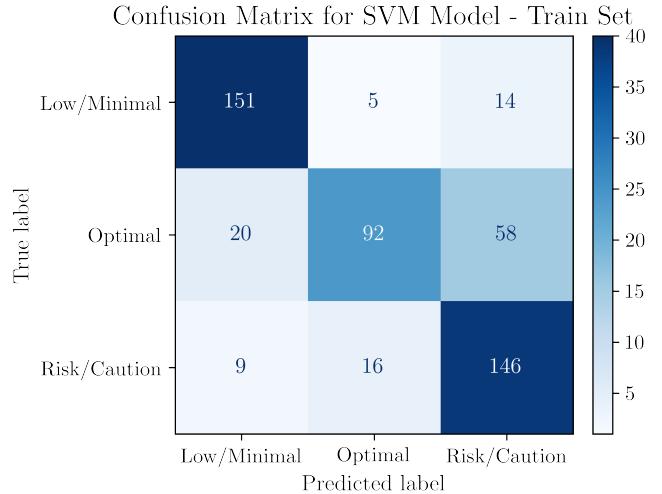


Fig. 12: Confusion matrix for the training data of SVM model

TABLE VI: Classification report for SVM model performance evaluation on training data.

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
<b>Accuracy</b>			0.76	511
<b>Macro avg</b>		0.77	0.76	511
<b>Weighted avg</b>		0.77	0.76	511

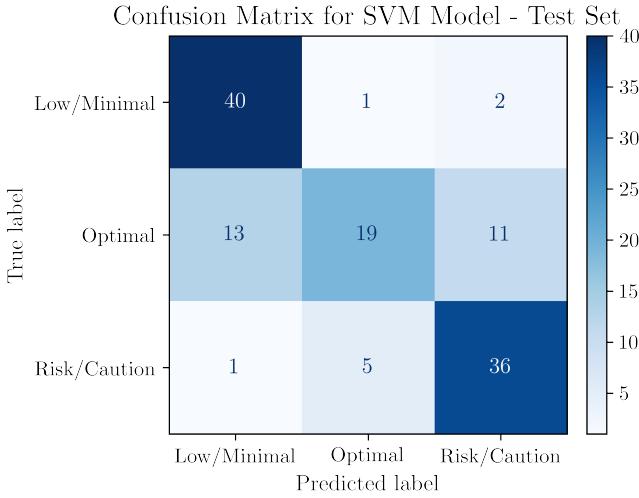


Fig. 13: Confusion matrix for the test data of SVM model

TABLE VII: Classification report for SVM model performance evaluation on test data.

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
<b>Accuracy</b>			0.74	128
<b>Macro avg</b>	0.75	0.74	0.72	128
<b>Weighted avg</b>	0.75	0.74	0.72	128

### VIII. DECISION TREE MODEL

The decision tree model was developed considering the hyperparameters available and setting ranges for the possible values to be estimated [3]. The ranges were selected taking in consideration the need to minimize the risk of overfitting. The max depth illustrates how deep the tree goes, and can be a sign of overfitting if it 'grows' too long. The min sample split constrains the tree in the number of splits allowed, as it requires more samples at each child node. A similar hyperparameter is min sample per leaf, meaning the minimum number of samples at any end node (making them more relevant). The split criterion evaluates the quality of a split when building the decision tree. The hyperparameters and ranges selected are presented in Table VIII.

TABLE VIII: Decision tree model hyperparameters search space.

Hyperparameter	Possible Values
Split Criterion	{gini, entropy}
Max Depth	[2, 3, ..., 8]
Min Samples to Split	[5, 6, ..., 20]
Min Samples per Leaf	[3, 4, ..., 10]

The max tree depth selected was 4, minimum samples to split 11, minimum samples per leaf 7, and the split criterion entropy ( $H$ ) is given by,

$$H(\text{node}) = - \sum_{\text{Class } j} p(\text{Class}_j | \text{node}) \log p(\text{Class}_j | \text{node})$$

The optimized DTree resulted in the structure as seen in Figure 14.

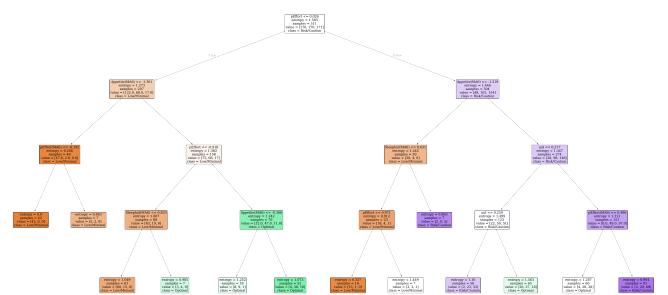


Fig. 14: Visualization of the decision tree model after optimization using randomized search.

The decision paths are characterized by the certainty in each node, highlighted by the color saturation for each of the classes ('Low/Minimal', 'Optimal', 'Risk/Caution'). The diminished proportion of end nodes with higher color saturation for 'Optimal' when comparing with the other classes is noteworthy, and consistent with the observations for the remaining models.

The learning curves were assessed from Fig. 15, which shows a significant improvement up until a training set size of 250, but with there's no significant gain with increasing set sizes, with the learning curves even diverging at 450 samples.

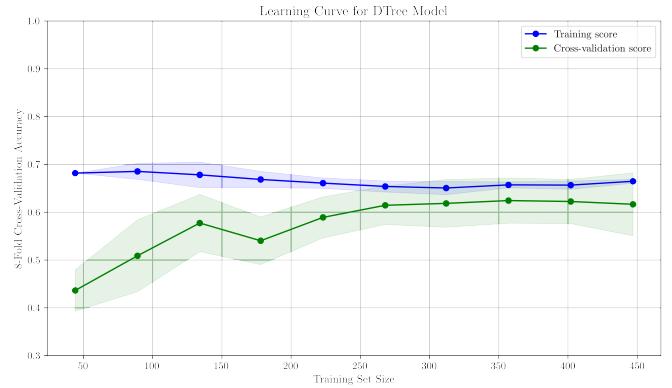


Fig. 15: Analysis of decision tree model performance using learning curve representation across varying training data sizes.

From the confusion matrices for training (Figure 16) and testing datasets (Figura 17) it is visible that the model had some difficulty between the 'Optimal' and 'Risk/Caution' classes, but with a satisfying prediction performance for 'Low/Minimal'.

The classification reports for both training (Table IX) and testing datasets (Table X) are well aligned with the interpretation of the confusion matrices. Furthermore, the performance metrics are consistent between datasets, with no indication of potential overfitting.

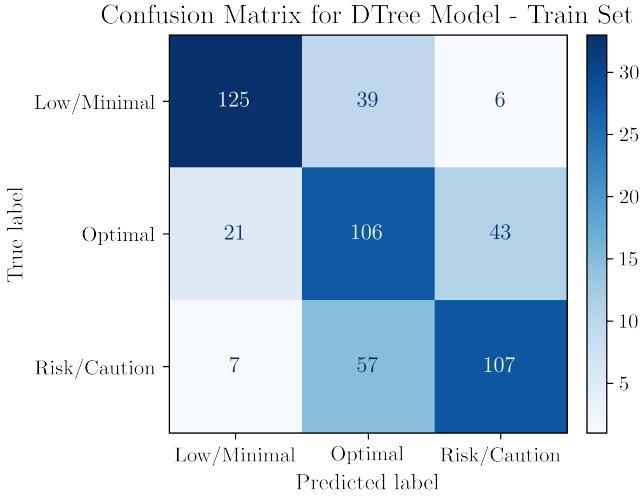


Fig. 16: Evaluation of decision tree model performance on training data performance via confusion matrix.

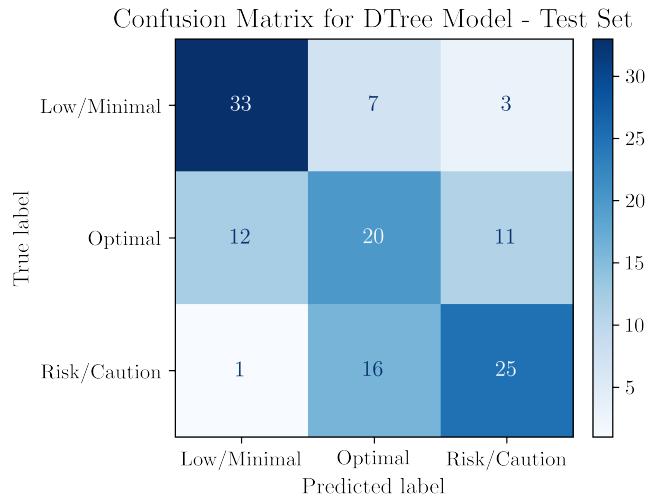


Fig. 17: Evaluation of decision tree model performance on test data performance via confusion matrix.

TABLE X: Classification report for decision tree model performance evaluation on test data.

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
<b>Accuracy</b>			0.61	128
<b>Macro avg</b>	0.61	0.61	0.61	128
<b>Weighted avg</b>	0.61	0.61	0.61	128

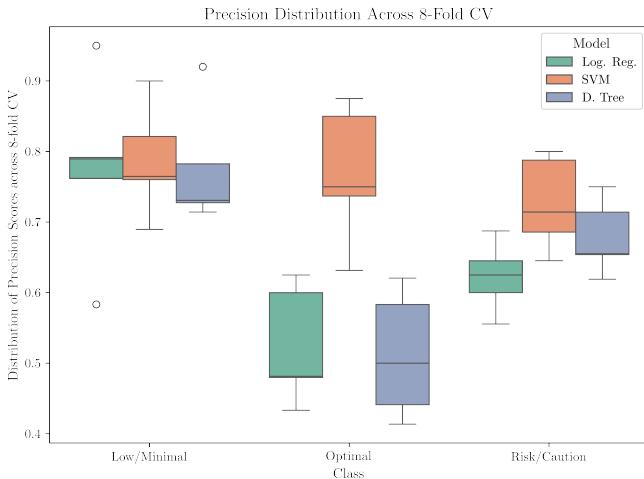
## IX. ANALISE DE RESULTADOS E COMP DE MODELOS

### A. *metricas ou resultados idk*

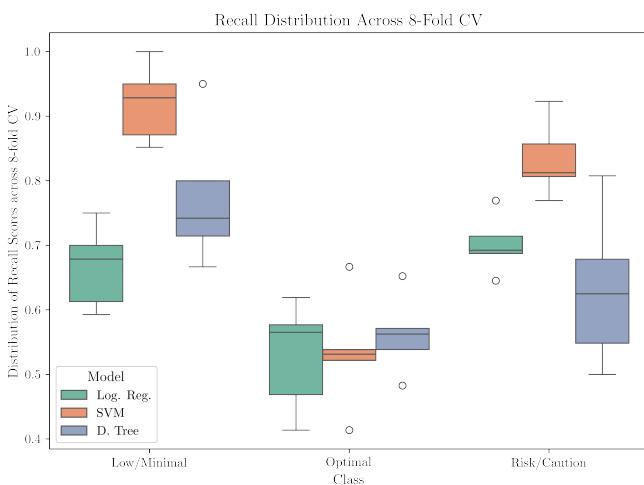
TABLE IX: Classification report for decision tree model performance evaluation on training data.

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
<b>Accuracy</b>			0.66	511
<b>Macro avg</b>	0.68	0.66	0.67	511
<b>Weighted avg</b>	0.68	0.66	0.67	511

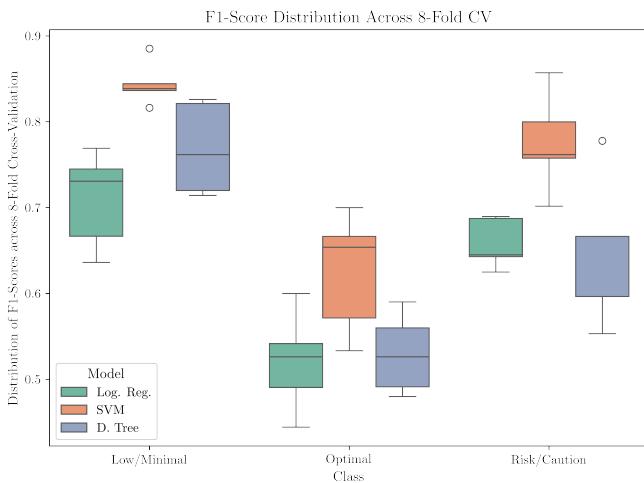
In order to clearly visualize and compare the different models, the data collected was represented using box plots, as depicted in Figure 18a, Figure 18b, and Figure 18c. The results were obtained from 8-fold cross-validation for the whole dataset in each of the models. In each iteration the models were training in 7 folds and tested with the eighth. The performance metrics were determined for each iteration and used afterwards to evaluate the global performance of each model. Each model used the optimized hyperparameters shared above.



(a) Precision scores across 8-fold cross-validation.



(b) Recall scores across 8-fold cross-validation.



(c) F1-Scores across 8-fold cross-validation.

Fig. 18: Boxplots showing the distribution of (a) precision, (b) recall and (c) f1-scores across 8-fold cross-validation for the different classes and models.

Boxplots, comparação entre modelos tendo em vista a disparidade de performance entre classes (particularmente para LogReg e DTee)

através da analise dos resultados do precision, recall e f1-score, é possivel observar alguns padrões. nomeadamente:

- o modelo SVM é o que tem sempre melhores resultados, destancando-se pela positiva em todas as métricas;

- a classe Optimal é destacada pelos seus maus resultados, revelando dificuldades na sua identificação e justificando a atribuição dos pesos na otimizacao dos modelos.

- o modelo decision tree parece apresentar uma variância significativa nos seus resultados across all metrics, o que pode indicar que é um modelo sensível aos dados utilizados tanto no treino como no teste

é importante referir que estas variações na variacao se deve ao facto de fazer 8-folds, e nao termos muitos dados, acaba por haver um desvio padrao maior nos resultados, por serem menos dados a ser testados

para alem disso , Facto de o nosso dataset ter dados repetidos para a class low/minmal, apesar de uma quantidade pouco significativa (12.5%), pode ocorrer ele ser treinado em dados q também irá testar, podendo haver um enviesamento positivo nas suas metricas.

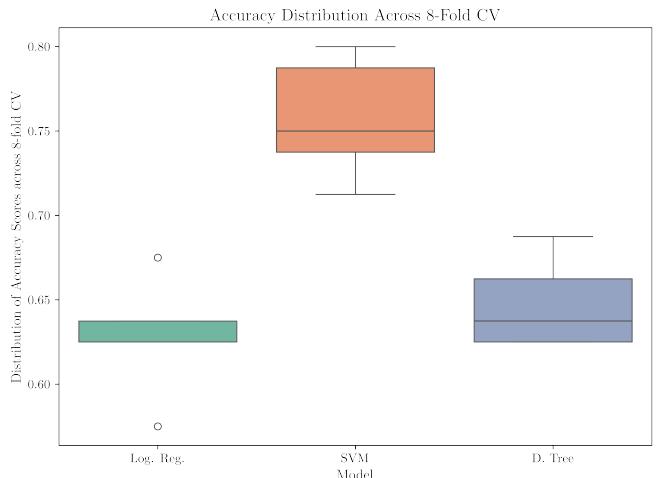


Fig. 19: Boxplots showing the distribution of accuracy scores across 8-fold cross-validation for different models.

de forma mais direta, através do accuracy, verifica-se mais uma vez o destaque do desempenho do modelo SVM quando comparado aos restantes, pelo que será o melhor modelo obtido.

#### B. representao no espaço?

em vista a auferir as razoes por detras de algumas conclusões a partir dos resultados expressos acima, decidiu-se representar the decision boundaries for each model on principal component axes, tal como os dados de teste e respetivas classificações reais. é importante notar que the decision boundaries são uma aproximação, uma vez que foi aplicado o PCA retendo apenas as duas primeiras direções,

pelo que alguma da variação foi perdida na criação dos decision boundaries, sendo perdido algum desempenho do modelo na visualização gráfica.

atraves da analise das seguintes figuras, pode-se verificar entao no fundo uma aproximacao do decision bondarie no espaço das componentes principais, revelando que categoriza as observacacoes mais a esquerda como low/minimal, as do meio como optimal, e as mais a direita como risk/caution (NS SE SE META ISSO, QUERIA FAZER UMA INTERPRETACAO MAIS GERAL DE TODOS IDK, OQ ACHAS?)

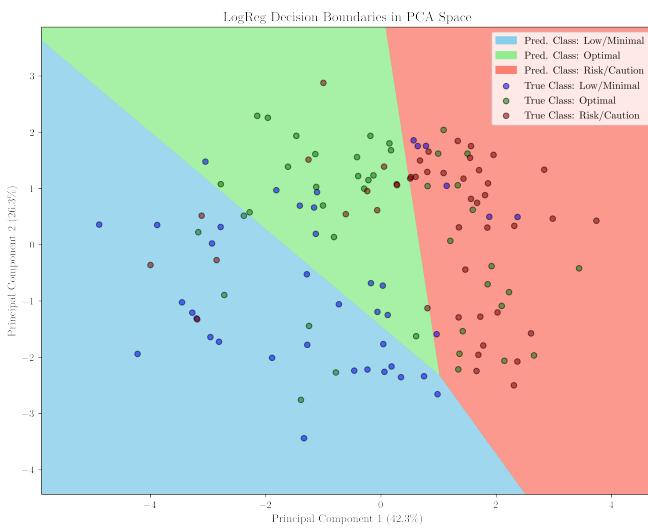


Fig. 20: Approximation of decision boundaries for logistic regression on principal component axes, showing true and predicted classes.

No caso da análise da Regressão Logística, observa-se uma maior precisão na classificação da classe Low/Minimal, o que também é confirmado pelo boxplot correspondente, que mostra uma mediana mais alta para essa classe em comparação com as outras. Isso indica que o modelo apresenta um desempenho relativamente consistente e confiável ao classificar instâncias dessa classe.

Quanto às restantes classes, Optimal e Risk/Caution, estas não parecem ter resultados tão precisos, não havendo uma separação mais significativa entre as classes, indicando que o modelo tem dificuldade em distinguir claramente entre essas classes, especialmente quando se observa a distribuição das previsões e as misclassifications no gráfico.

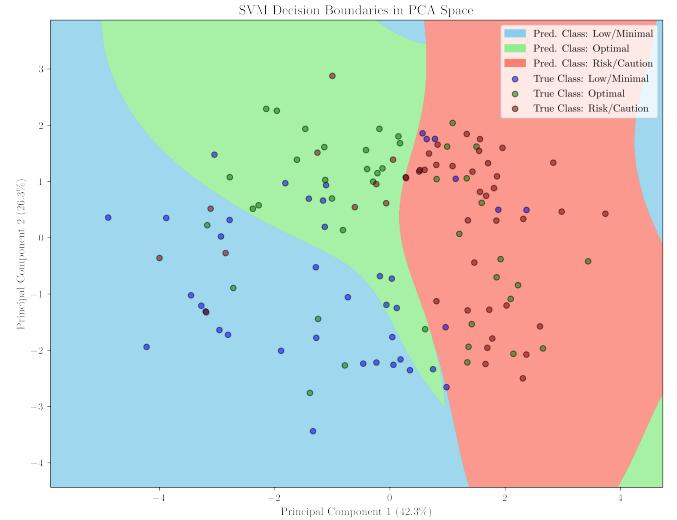


Fig. 21: Approximation of SVM decision boundaries on principal component axes, showing true and predicted classes.

Através da análise da figura 21, é possível observar claramente o comportamento do kernel RBF no SVM, evidenciado pela delimitação mais precisa das fronteiras de decisão entre as classes. Esta delimitacao, e atendendo a classificacao real de cada observacao, leve a querer que haja uma melhor precisão na classificação das diferentes classes, capturando melhor as complexidades dos dados, o que também foi confirmado pelos resultados da figura 18a.

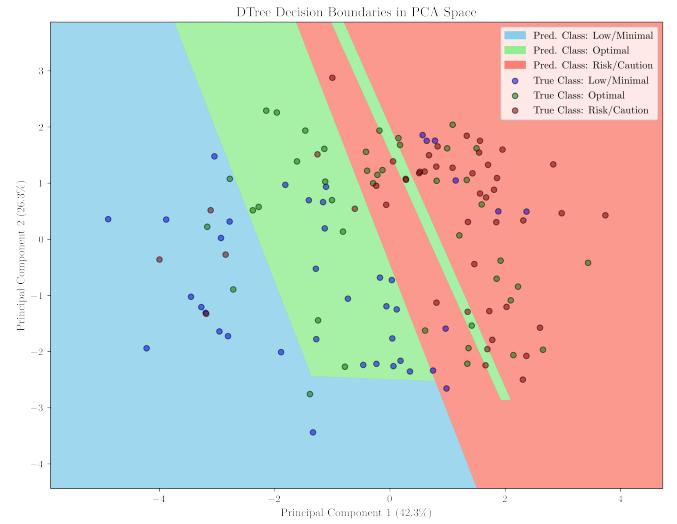


Fig. 22: Approximation of decision boundaries for decision tree on principal component axes, showing true and predicted classes.

por ultimo, a decision tree apresenta decision boundaries mais complexos, o que se pode dever as suas decisões discretas, poeque ao contrário dos outros modelos, que tentam ajustar uma superfície contínua para separar as classes, o decision tree divide o espaço de características de forma

hierárquica e discreta, porque em cada nó da árvore acontece uma divisão baseada em condições binárias

embora a árvore seja um modelo poderoso para capturar relações não-lineares, a sua representação gráfica neste contexto mostra que enfrenta dificuldades para separar claramente as observações entre as classes, com exceção da classe "Risk/Caution", que apresenta um comportamento mais favorável.

#### AUMENTO DIMENSIONALIDADE E OVERFITTING ?

sobre todos:

Estas análises gráficas ajudam a perceber a complexidade do problema e a dificuldade dos modelos em conseguir uma separação eficiente entre as classes. A sobreposição, especialmente da classe Optimal, destaca a necessidade de considerar ajustes nos modelos, transformações adicionais nos dados ou até modelos mais robustos que possam capturar melhor as relações subjacentes entre as classes e melhorar a precisão da classificação. (CONCLUSÃO OU AQUI?)

#### X. CONCLUSION

SVM apresenta um desempenho superior, com maior consistência e precisão em relação às outras abordagens analisadas.

The conclusion goes here.

#### ACKNOWLEDGMENT

The authors would like to thank... professora petia e treinador e organização do clube

#### XI. NOTAS

Discussão dos resultados: - Importante considerar os kernels selecionados, e o que isso pode representar dado o tamanho reduzido do nosso dataset

- Há coerência entre métricas (F1-score Test vs Train está à mesma distância quase sempre)

- Temos algumas zonas em branco ao longo do documento, temos de ver como resolver (não sei se será mais texto em alguma zona, mas gosto pouco de adicionar palha)

- Novelty and contributions (3) Compare your solution with the works of other authors (published references), try to propose a better solution, e.g. improve the performance of the ML model in solving the problem you work with.

#### SVM

Não introduzi os parâmetros como tabela, porque varia de modelo para modelo e achei que não fosse ficar tão bem. Queres indicar as expressões para todos os kernel ou só discutir brevemente? É que se não começamos a ter pano para mangas, particularmente só com o SAGA....

Nesta parte não introduziria mais texto, deixava de maneira mais simples e ilustrativa como no caso do LogReg

Considerar usar dataset com informação combinada de atletas para reter caráter temporal (progressão da época, aumento de cargas e combinação com features que tenham referência temporal [carga de treino do dia anterior])

Modelos sugeridos: Random Forest / SVM

temos com objetivo generalizar para que qq treinador possa ter noção do comportamento e do estado dos seus atletas,

podendo ajustar os seus treinos e cargas físicas consoante as medidas de fadiga.

para além disso tbm se quer ver quais as métricas mais importantes relacionadas com a fadiga

De acordo com a conversa com a professora é importante perceber como é que se deve definir a memória da nossa t-SNE, e a importância que isso tem no período da fadiga.

ns q dados fornecidos por um treinador de natação, tivemos de organizar os dados em folhas de excel, visto estarem por linhas e com gráficos e fórmulas de acordo com o treinador, bla bla, teve-se fazer oq? sabes melhor q eu pq foste tu q fizeste

os valores da fadiga foram convertidos para categóricos, pq e mais interessante classificar a fadiga, tendo em conta intervalos dados pelo treinador, do q numero que tornam mais difícil a interpretação dos mesmos

de seguida agruparam-se os dados todos num novo ficheiro excel, para posterior análise através do python, de forma mais facilitada

tem-se então dados diários, para X atletas, durante ns q tempo, ao longo da época tal, totalizando x observações (linhas)

[1] better way chronic load

[4] Predictive Modelling of Training Loads and Injury in Australian Football Carey 2018

[5] another review (Prediction models for musculoskeletal injuries in professional sporting activities: A systematic review)

[6] Murray et al. (2016

#### REFERENCES

- [1] S. Williams, S. West, M. J. Cross, and K. A. Stokes, "Better way to determine the acute:chronic workload ratio?" *British Journal of Sports Medicine*, vol. 51, no. 3, pp. 209–210, 2017. [Online]. Available: <https://bjsm.bmjjournals.com/content/51/3/209>
- [2] "Support vector machines," <https://scikit-learn.org/stable/modules/svm.html#svm-kernels>, 2024, accessed: 2024-11-25.
- [3] B. F. Mohtedi, "Indepth: Parameter tuning for decision tree," 2017, accessed: 2024-11-25. [Online]. Available: <https://medium.com/@mohtedibf/indepth-parameter-tuning-for-decision-tree-6753118a03c3>
- [4] D. L. Carey, K. Ong, R. Whiteley, K. M. Crossley, J. Crow, and M. E. Morris, "Predictive modelling of training loads and injury in Australian football," *International Journal of Computer Science in Sport*, vol. 17, no. 1, p. 49–66, Jul. 2018. [Online]. Available: <http://dx.doi.org/10.2478/ijcss-2018-0002>
- [5] D. Seow, I. Graham, and A. Massey, "Prediction models for musculoskeletal injuries in professional sporting activities: A systematic review," *TRANSLATIONAL SPORTS MEDICINE*, vol. 3, no. 6, pp. 505–517, 2020. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/tsm2.181>
- [6] N. Murray, T. Gabbett, A. Townshend, and P. Blanch, "Calculating acute:Chronic workload ratios using exponentially weighted moving averages provides a more sensitive indicator of injury likelihood than rolling averages," *British journal of sports medicine*, vol. 51, 12 2016.