CRANFIELD UNIVERSITY

Martí Rodrigo Corominas

Development of an Intelligent Approach for Delivering High Performing Training Solutions

School of Aerospace, Transport and Manufacturing
Management and Information Systems

MSc
Academic Year: 2021 - 2022

Supervisor: Prof. John Ahmet Erkoyuncu
Associate Supervisor: Dr. Bernadin Namoano
August 2022

CRANFIELD UNIVERSITY

School of Aerospace, Transport and Manufacturing
Management and Information Systems

MSc

Academic Year 2021 - 2022

Martí Rodrigo Corominas

Development of an Intelligent Approach for Delivering High Performing Training Solutions

Supervisor: Prof. John Ahmet Erkoyuncu
Associate Supervisor: Dr. Bernadin Namoano
August 2022

ABSTRACT

Predictive modelling is a state-of-the-art technique for which an objective variable is predicted given a set of environmental parameters. Many types of models exist and most of them can be built with different architectures.

This project comes from the necessity to improve the measurement of biometric indicators and understand how team performance can affect their level in a defence training context. The biometric indicators of interest are the player's heart rate and stress index, which are going to be related to information extracted from the GPS coordinates they have. The objective has been to study the evolution of the stress index during an exercise period, trying to determine how the other biometrical and positional factors influence its level.

After reviewing existing work in the medical field for stress monitoring and team performance in an educational context, it has been observed that no literature involved both topics relating them to each other, let alone in a defence environment. A list of quality indicators has been defined to assess the quality of the provided raw datasets, the information from which has been managed to build a single dataset that could be used to train a model.

The results of the quality assessment have shown that the recording frequency for the different indicators should be modified to a common value since the existing time difference between recordings has proven to be a complex issue to solve when building the model.

Regression, neural network, and random forest models have been tested, with the latter being the one offering the best precision. The heart rate, the duration of the exercise, and the distance from the player to the opposite team were the variables that played a major role in the prediction.

Overall, a valid prediction has been reached despite the missing gaps in the provided datasets. The key features to predict the stress index have been identified and recommendations in terms of data quality have been made so the predictions can be improved.

Keywords:

Predictive Modelling, Data Quality, Stress Index, Heart Rate, Neural Network, Regression, Random Forest

Word count: 8394

ACKNOWLEDGEMENTS

I would like to thank Dr. Bernadin Namoano and Pr. John Erkoyuncu for supervising this project and helping me with their guidance and knowledge. I would also like to thank Mr. Jim Sibson, Mr. Michael Bodie, and Ms. Tori Hull from Babcock International, for sponsoring this project, always being helpful and kind, and for their continuous support during the project.

And to my family and friends, the ones back at home but especially the ones I have found in this incredible journey that Cranfield University has been.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

SI	Stress Index
HR	Heart Rate
NN	Neural Network
ANN	Artificial Neural Network

1 INTRODUCTION

Babcock International is a private engineering company specialized in providing design and development of training exercises to defence, aeronautics, and nuclear sector companies and managing complex assets. For decades they have been one of the world's leading companies in engineering consultancy services and they are present all over the world (What We Do - Babcock International, 2022).

This project has been proposed in a defence-related context. It will address the necessity that Babcock is facing in trying to better predict some of the key indicators involved in a defence training exercise, while also improving the methodology to collect the necessary data for the analysis.

1.1 Background and motivation

Prediction models are an ever-in-development tool used in many contexts to predict how an asset will be affected by a change in its environment without having to change the conditions surrounding it. However, to be able to reach a model able to predict accurately an objective variable given a set of input variables, the model has to be trained with a complete set containing the input and objective variables (Lawton et al., 2022).

Usually, the process to train a model consists in splitting a dataset into two parts: a training set and a test set (Predictive Modelling: Splitting Data into Training and Test Set - FinanceTrain, 2022). The first one is used to adjust the model weights to predict as well as possible the objective variable. Then, the resulting model is tested against the test dataset to analyse how well the prediction works.

A common problem to create an accurate prediction model is the quality and type of the input data. There is a high number of models that can be used to treat each problem, and some of them struggle to work with datasets that are not complete. Also, each model processes data differently which makes some of them more suitable to manage different types of data (numerical, categorical…) while others work better with a single type common for all variables (Ali, 2020) (10 Predictive

Modelling Types (With Benefits and Uses), 2021). Therefore, tools like a data quality assessment are useful to know beforehand what the limitations of the datasets are and what kind of model suits best their analysis.

The motivation of this project comes from the necessity to ease the measurement of biometric indicators and understand the team performance from the participants to Babcock's defence training, to relate the player's HR with their Stress Index (SI). With a model that could predict the SI, the task of analysing the player's performance becomes easier, reducing the amount of recording devices to use and the process of transforming their captured information into conditioned data. Also, the recorded data so far shows a considerable amount of missing and repeated values, which makes it necessary to define a way to quantify its quality and fitness to be used for a prediction model.

1.2 Aim and objectives of the project

The project's aim then is to analyse the datasets containing information about the team exercises performed by the training participants, see their quality level, and use them to train and validate a prediction model for the SI. The information in the datasets includes geographical and biometrical information for each participant. The key data features are going to be listed and unified into a single file, while also assessing the quality of the source data in the process.

The objectives of the project can be listed below, in line with the aim defined above:

- Conduct a literature review to study existing literature in the project-related studies.
- Conduct a data quality assessment for the provided datasets, review its gaps and propose solutions to improve them.
- Identify the key data groupings to develop and validate an adequate prediction model, trying multiple architectures and types of models with the unified dataset as input information.

1.3 Thesis structure

The thesis structure is going to be as follows:

- Chapter 2 consists of the Literature Review. Previous studies will be analysed to observe the existing work in the project subject's related fields
- Chapter 3 consists of a methodology explanation, detailing how the project has been deployed throughout the months it has lasted.
- Chapter 4 consists of the data quality assessment, evaluating the quality of the raw datasets provided by Babcock.
- Chapter 5 consists of the explanation of the analysis that has been conducted.
- Chapter 6 includes the results and its discussion.
- Chapter 7 concludes the project with the conclusions and future work.

2 LITERATURE REVIEW

To begin the literature review, research will be conducted on studies that approach the evaluation of training and learning exercises in a team environment from a modelling perspective. The indicators considered to evaluate performance will be discussed and how different modelling solutions are proposed.

After gaining this general view, a more technical approach will be taken, analysing different types of models and their characteristics. Since the goal of the project is to create a model to predict biometric indicators, this part of the literature review will give us a list of candidates to be a possible modelling solution.

2.1 Learning measurement and team evaluation

In this subsection, a general approach to learning evaluation is taken, to study how knowledge retainment is studied from a modelling perspective.

Regarding the research of this project, literature on team evaluation can help to understand which factors are important when training is designed and delivered. According to (Pappas, 2015), simple tasks like observing learners' patterns before and after the taught lessons and setting performance goals or using assessments to evaluate knowledge retention, are one of the first steps in evaluating student performance. While these options can be useful for this project, it would be necessary to have a quantitative indicator to evaluate the learners' progress. To address this need, (Pappas, 2015) proposes the return of investment (ROI) as an indicator, although for this project it is not suitable due to the lack of knowledge we have about the training costs. (Raspopovic & Jankulovic, 2017) based their research around three dimensions defined by (DeLone & McLean, 2016) which can be evaluated quantitatively to measure the training effectiveness:

- Use dimension: Degree in which learning material such as presentations, papers, or other teaching materials are used.
- User satisfaction: Direct quantitative feedback from students via surveys.

- Net Benefits: Quantifiable savings in time or costs resulting from the training.

Although the study states a correlation between the three dimensions, it does not specify how the correlation is calculated. Especially for the last dimension, net benefits, it would be useful to know how they are directly related to both use and user satisfaction dimensions in a quantitative manner. However, these while given the nature of the training exercises the use dimension will not apply to this project, for the feedback and evaluation dataset studies the user satisfaction and net benefits indicators will be suitable for the study.

Further work using the net benefits dimension obtained through an online course, proposes to measure them through three learning domains (Dadd & Hinton, 2021), being:

- Cognitive: Related to the degree of retained knowledge from the training exercise.
- Affective: Evaluating quantitatively the changes in attitudes and interests of the learner.
- Psychomotor: Quantifying the improvement in the learner's skills in a range from novice to mastery.

(Keržič et al., 2021) additionally, provide a deeper insight into the topics discussed so far. Starting as well from the proposed dimensions by (DeLone & McLean, 2016), explore sub-indicators to construct the following model:

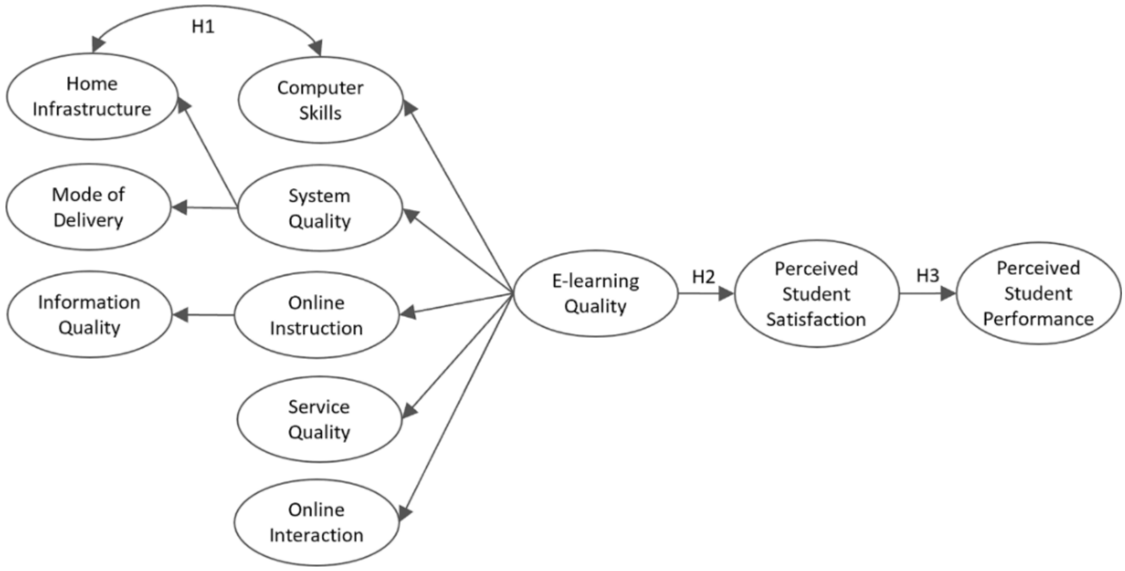

Figure 1 - Model defined by Keržič et al. Source: (Keržič et al., 2021)

Their assumptions are as follows:

- H1: Student computer skills are related to their home infrastructure.
- H2: E-learning quality has a direct positive impact on perceived student satisfaction.
- H3: Perceived student satisfaction has a direct positive impact on Perceived Student Performance.
- H4: E-learning quality has an indirect positive impact on Perceived Student Performance (through H2 and H3).

After gathering data from student surveys, it is found that the model explains 55% of the Perceived Student Performance. System and Service Quality have a strong impact on E-learning Quality, which in its turn influences greatly Student Satisfaction, thus proving H2. This last indicator influences greatly Student Performance, proving H3. On the other hand, H1 is not proven to be so strongly certain, thus not ensuring good computer skills in well-equipped home infrastructures.

Similar domains are proposed by (Aguinis & Kraiger, 2009) and (Salas et al., 2003) , but adapted to the military environment. The defined Learning, Behavioural and Cognitive outcomes evaluate affective changes in the participant's behaviour and the amount of knowledge they acquire through the training programmes, similar to the domains defined in the papers discussed so far. However, the reaction and organisational domains are recommended to be added to the analysis. The first is related to the feedback provided by players, integrating it helps ensure an impact on a later improvement process of the training exercise. The second, which is relevant for this project, provides an example of how some organisational indicators are used to evaluate the efficiency of the training exercise. It is key that team performance indicators are included in the analysis since, ideally, training exercises are supposed to enhance team performance (Alvarez et al., 2016).

Overall, these will be the most useful for this study, since they can be used to measure the training effectiveness and knowledge retention, adapted to an environment similar to one of the training exercises we will be studying. However, the measurement of behavioural and organisational domains will be a difficult one to execute, since the outcomes in them depend on the exercise that is being performed. None of the authors cited above provides a clear example of how they have been measured and it will depend on us to try to measure them according to their definition as well as we can with the information provided in the training datasets.

2.2 Data Quality Assessment

An assessment of the quality of the data will need to be taken to have an idea of the possible limitations the model could have given the quality of the provided data. There are a lot of studies evaluating data quality and its characteristics. However, for this project, there is no necessity for any special indicator. The information at hand from the datasets can be evaluated using the common indicators proposed by (Gupta, 2022).

(Pipino et al., 2002) have provided a theoretical approach to data quality, providing definitions for some indicators of interest:

- **Believability:** Defined as the extent to which data is true and reliable.
- **Completeness:** Defined as the extent to which data does not have missing values.

- **Objectivity:** The extent to which data is unbiased.
- **Timeliness:** The extent to which data is updated sufficiently (Antonopoulos, 2022).

These are furtherly developed by (Woodall et al., 2015), applying this knowledge to asset management, adding to the previous indicators the following ones.

- **Accuracy:** Defined as the extent to which the captured value corresponds to the actual one (Kim, 2020) (Kaushik, 2020).
- **Consistency:** The extent to which the data is recorded in the same format (Ghosh, 2020).

(Chan et al., 2021) extend this to asset maintenance, including evaluation and conditioning for automatic data cleansing and machine learning methods. To evaluate this properly and detect repeated values on the records, the "**uniqueness**" dimension is incorporated into the analysis, defined as the percentage of records that are duplicated and, therefore, need to be treated (Mihăiloaie, 2015).

Overall, the data from the datasets will need to be evaluated similarly, to bear in mind the limitations it may present when going into the model development. Also, an assessment like this can be an attractive opportunity for Babcock International to know where the flaws in their data recording processes are and which aspects need to be improved to be able to provide the best possible input to the prediction model.

2.3 Heart rate and Stress Index Analysis

Various domains have studied stress index and heart rate relationships. Most of the studies are in the medical field and are not applied to training defence exercises. However, they can provide an interesting approach on how to relate HR and SI, as well as perspectives on how to build and train models that predict one given the other.

Research by (Coutts et al., 2020), (Fujiwara et al., 2021), and (Faouzi, 2022), collect HR information using wristband sensors that participants wear 24 hours during their study periods. The data obtained are transformed into heart rate variability features (HRV), which are mainly based on the RR intervals and the high and low-frequency density. They are defined below:

- RR interval: Interval between R peaks for heartbeats. Once the noise in the signal is filtered out, they are referred to as NN intervals. A simple visual example can be seen below.

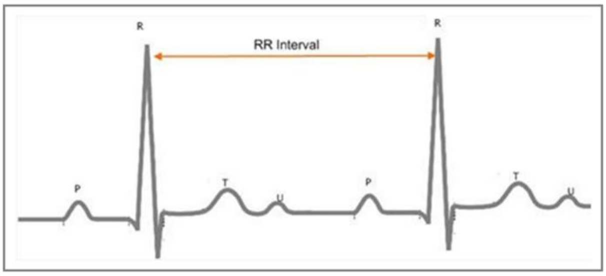

Figure 2 – Diagram of HR RR peaks and IBI. Source: https://www.ni.com (royalty free).

- AVRR, SDRR: Average and standard deviation of RR intervals
- pRR50: Percentage of successive intervals that differ more than 50 ms
- Low/High Frequency: Mean power of spectral density for HR measurements

These indicators are fed into their time series neural network models to predict the level of stress of participants. The assessment of the resulting models was performed by comparing the stress prediction to stress evaluation questionnaires obtained directly from participants' feedback.

Here (Liu & Ulrich, 2013) provide further details on which are the HRV features used in the two studies mentioned above. It includes features in the time domain as well as the frequency domain. For the first domain, the features are mostly based on studying the mean and standard deviation values for inter-beat intervals (IBI) and the percentage of them which are considered to be normal (quantity of reliable intervals once the noise is filtered out of the signal). Regarding the variability features of the frequency domain, the focus is on the power of the IBI for different frequency windows.

Their study is focused on predicting the level of stress from participants of the study only having their electrocardiogram (EGC) information. For validation, they use the galvanic skin response (GSR) as an indicator which traditionally has been used to measure stress. The level of accuracy they reach is quite high, which is interesting given that the indicators we'll have to study are quite similar. However, a detailed process on how to extract. For our study, the information could be complemented with variables similar to the ones defined by (Sun et al., 2008), and other biometrical and chemical indicators that evaluate quantitatively the permeability of the skin.

Here (Shaffer & Ginsberg, 2017) describe in detail the main HRV variables manipulated in this paper as well as the ones mentioned above. These features are fed into an autoregressive model which results in different prediction accuracies when adjusting certain parameters. The authors tested multiple widths of prediction windows to forecast the evolution of stress and they noticed that modifying this parameter influences the final accuracy of the model. This fact will have to be considered when building the model to relate HR and SI for this project since it will play an important role when trying to improve the model accuracy. The model with a higher accuracy score results in a prediction of 98% of accuracy in cases of extremely low or high stress and 85% for mild cases.

Overall, the discussed work has been useful to see how a medical approach is taken when trying to relate HR and SI. Since the research mainly focuses on the medical field, this project's goal will be to adapt it to the defence field, although similar methods will be used to transform the beats-per-minute (BPM) into HRV features. This will later be combined with the information extracted from the global positioning system (GPS) data that has also been provided to predict the evolution of stress.

2.4 Time series analysis

The articles above mention modelling as the methodology to study and predict the indicators they analyse. However, it is necessary to expand the literature time series analysis as the characteristic of the problem addressed is a regression problem. Their characteristics are explained below, and some cases of their application will be discussed to see if they are fit to be candidates to be a model solution for this project (Han et al., 2011).

2.4.1 Artificial Neural Networks

Artificial neural networks attempt to simulate the functioning of the human brain in a non-linear approach. They are based on a series of simple processing units (neurons) connected through weight functions, as seen in Figure 3.

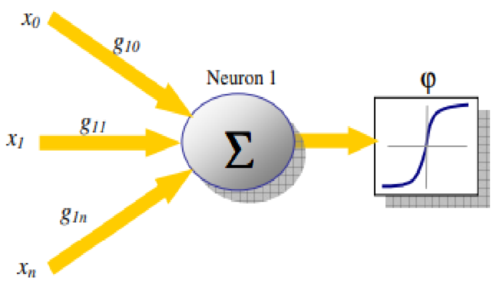

Figure 3 - Neuron diagram (Bechini G, 2007)

The input variables of the neuron are multiplied by the weights and added together. This is then multiplied by an activation function which can have multiple shapes like threshold, linear, etc. Depending on the result of this product, the neuron will activate or not, sending the output signal to the next neuron in the network. (Bechini G, 2007)

One of the most common methods to train these networks is to provide them with a training dataset, which has pre-defined inputs and outputs for the model. The weights are then adjusted and optimised, so the actual outputs of the network are as similar to the pre-defined outputs as possible (Ismail Fawaz et al., 2019).

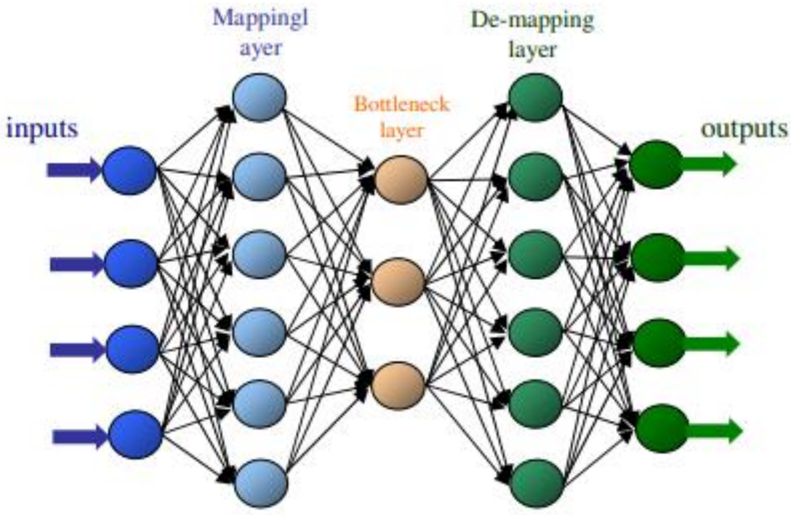

Figure 4 - Auto-associative neural network

A common type of ANN is auto-associative neural networks (AANN), also known as autoencoders (Sun et al., 2003) (Patel & Patel, 2016). They are a type of ANN composed of a mapping layer, a bottleneck layer, and a de-mapping layer. This particular architecture has the advantage that in case of a change in the input signal (noise or bias), the network can still replicate the original signal, filtering out the noise.

However, the major drawback of ANN is that, besides being difficult and slow to train, they are quite unexplored still. It is not known which type of architecture suits best each problem, and therefore the time needed to optimise the solution can be quite extensive.

In addition, (Agatonovic-Kustrin & Beresford, 2000) provides an overview of one of the most remarkable uses ANN has had in the medical field is the analysis of images to diagnose cancer diseases. Many related articles study models to predict patient conditions based on image processing via an ANN model (Li et al., 2017) (Islam, 2019). While it is useful to see an application in a practical field of an ANN model, the fact that this case specifically analyses images is not too useful for this project.

ANNs are also commonly used to work with time series models, such as the one it is aimed to develop for this project. (Mohamed, 2019) discusses how categorical and numerical data is used to predict solar radiation in a set of cities. This approach is similar to the one for this project, firstly organising the available data into a single input dataset. Some statistical indicators are also studied for the prediction output, which will be useful in the evaluation process of this project.

2.4.2 Decision trees

Decision trees are a simple and understandable approach to modelling. Their algorithm consists of a set of decision rules defined by the values of the variables from the training dataset, which split the data into groups (*What Is a Decision Tree?*, 2022) (Dhar et al., 2021).

The design process for a decision tree is as follows. The base node (or root node) contains the whole of the data, and it is split into two or more branches according to the different values of a single one of the variables. The data points are distributed according to this rule and stored in decision nodes. Each of the decision nodes is going to be split into more branches depending on other rules defined by the value of the other variables. The splitting process is going to continue until reaching a leaf node, in which the predicted result has a predefined probability of coinciding with the output training data. (Chauhan, 2022)

The disadvantage of this type of model is the risk of overfitting the model, creating more branches than needed. This is particularly delicate in the case of large datasets, such as the case of this project. Pruning processes are applied to reduce this risk, according to which a node that, given a certain probability threshold of reaching a successful outcome, shows a result lower than the threshold is not furtherly explored.

A common process to also reduce the overfitting problem is using a particular case of decision trees algorithm known as random forest. It consists in producing a set of decision tree models and given an input, selecting the most common response as the prediction output. This makes the prediction to be more accurate but less understandable when interpreting the path followed to reach a certain output.

(Podgorelec et al., 2002) provide a general overview of decision trees used in medical applications, providing a clear explanation of their functioning. However, a very interesting case study is presented by (Šprogar et al., 2001) in which they develop a model for a real case study using vector decision trees that can predict with a 75% accuracy the patient's diabetes diagnosis. This real case study will be useful since it uses datasets with continuous and categorical data like the ones that have been provided for this project.

2.4.3 Regression

Linear regression is one of the simplest algorithms to use in machine learning. It is used to find the best linear equation that fits a given data set. Although it is simple and quick to compute, it has a major drawback that its name implies: the resulting model is linear, which can struggle to predict the desired outcome if the input data is not linear.

A good solution for this problem is a Gaussian Process Regression. This method takes a non-parametric approach to this problem, allowing the prediction model to have as many parameters as needed (Vassallo, 2019) (Wang, 2022).

Let's consider two groups of data points: one group is to train the model and the other is to test it. The Gaussian Process considers a group of functions to which the training data points belong and approximates the probability of the response values of the test data points belonging to each one of them. With these probabilities, a mean and a deviation function are calculated, therefore having a zone of probability to which the predicted response can belong (Bailey, 2019).

This type of model has proven to be successful in previous studies, like by (Shashikant et al., 2021), in which they predict with high accuracy the risk of having diabetes with HRV data from patients

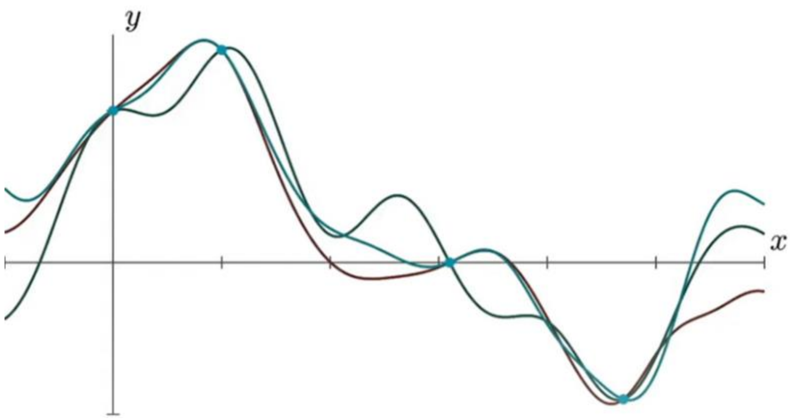

Figure 5 - Gaussian process prediction functions

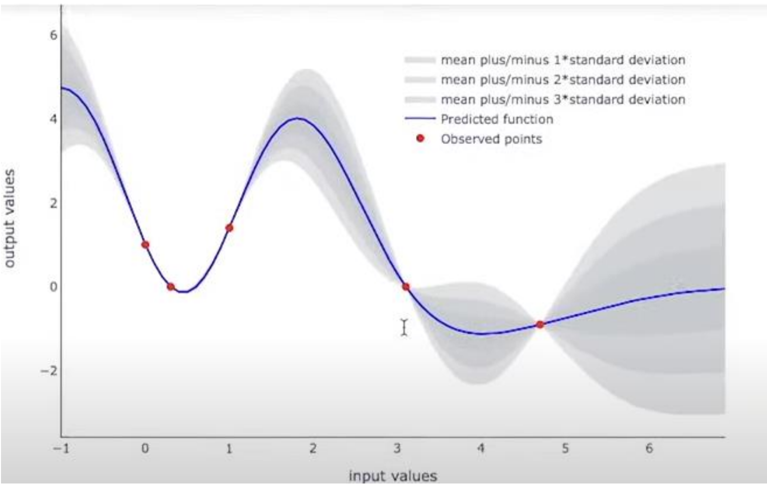

Figure 6 - Gaussian process output

2.5 Research Gap

In this literature review, it has been observed how modelling is approached in learning and team evaluation, giving an idea of which indicators are commonly considered in terms of knowledge retention. At the same time, it has been reviewed how machine learning techniques have been used in a medical environment, and how modelling is approached in terms of predicting a diagnosis given a dataset of patient information.

This project will try to combine both perspectives into a single prediction model that can predict the training participants' stress index considering both medical and team environment factors, something that has not been explored before. Also, to do so in a defence training context will be something innovative since most of the case studies analysed happen in a medical environment.

3 METHODOLOGY

As observed in the previous sections, after establishing the objectives and aim of the project, a literature review has been conducted to study how the development of a prediction model was approached from a team perspective, as well as what indicators are usually used when managing biometric indicators. After the literature review, the practical study will be conducted with the data provided from training participants. The complete project flowchart can be seen in Figure 7.

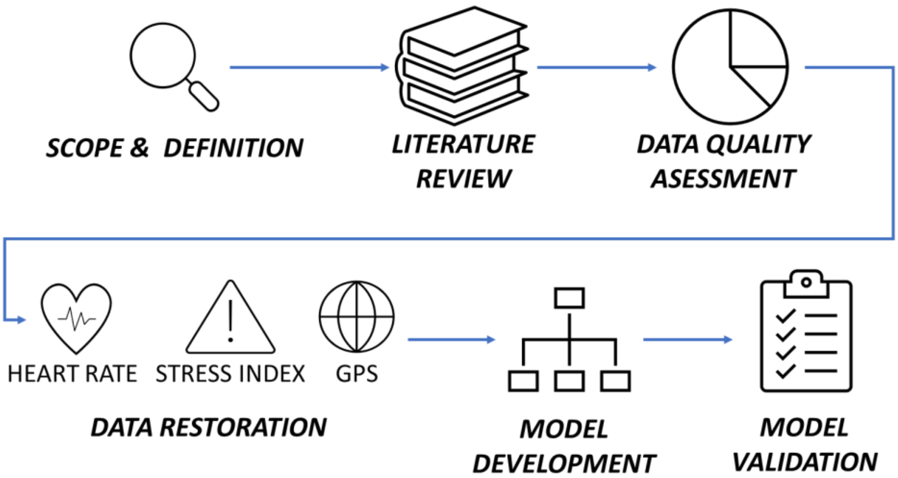

Figure 7 - Project flowchart

Three datasets have been provided which contain information about the exercises performed by the training participants. Two of the datasets have information on biometric indicators for participants, evaluating the heart rate and the stress index, while the third one provides the GPS coordinates for each player. An additional dataset has been provided with the personal information for participants, such as age, height, weight, and background expertise.

Before proceeding to feed the information to the model, the data needed to be evaluated in terms of quality and organised in a single dataset. A data quality assessment has been taken, to evaluate the provided datasets according to several quality indicators. The three datasets have been analysed separately in terms of completeness, accuracy, and uniqueness. Also, since the base for the general analysis has been the common timestamps in the HR and SI data as it is explained in Section 4, the HR and SI datasets have been compared with the GPS dataset in terms of timeliness.

Once the data has been assessed in terms of quality, a preliminary analysis using MS Excel has been conducted. This analysis has combined the features of the multiple datasets into files that compare HR to SI, HR to GPS, and SI to GPS. Its goal has been to notice any trends that the correlations between these indicators may show and, therefore, have a general idea of how the model could predict the objective variable (SI).

In the model building stage, a single input dataset has been developed containing the data by second, to properly predict the SI given the other indicators. It has been based on the timestamps that contain both HR and SI registries, and the speed and distance values have been added when registries coincided between datasets. The calculation of the speed and distance indicators will be a bit more elaborate, as it is explained in Section 295.3. Individual parameters such as player's age, height, and others are also included in the complete input dataset. The complete list can be also found in Section 295.3.

To approach the model development stage, the software Orange was first used to try to build both NN and Random Forest prediction models. However, this software lacked computing power and the results had a very low percentage of precision. Weka was then used as modelling software, given its higher potential when computing especially NN prediction models.

First, the Neural Network models were developed, taking an empirical approach, building and testing multiple models starting from networks with only one layer of hidden neurons. Multiple numbers of neurons were tried for the layer and the one that showed the best prediction score was kept as optimal. Then, another layer was added to the model, and it was proceeded to test the numbers of neurons. The process was continued until a four-layer model was tested and the results showed the last layer did not improve the predicted results.

After testing NN models, regression models have been tested. Like in the previous case, an empirical approach has been taken and multiple values for the batch size and ridge margin have been tested.

Finally, Random Forest models have been used as another candidate to create a prediction model. In this case, Python has been used as modelling tool since Weka did not offer the possibility. Several tests have been run, starting with an empirical approach until reaching a point where the prediction score did not improve considerably.

4 DATA QUALITY ASSESSMENT

Before proceeding to organize the provided data into a file apt to be fed into a prediction model, an assessment of the quality of the datasets must be taken to know how good the data is. This is to have an idea of what the difficulties are going to be when trying to build a model and what gaps of knowledge may exist in the data that has to be used to predict the objective variable.

The data from the original datasets have been assessed in terms of quality, measuring the following indicators:

- **Completeness:** Defined as the percentage of registered data points for HR, SI, and GPS data compared to the total number of expected data points for said indicators. The expected number of data points is calculated with the average recording frequency for each indicator in the time window between the earliest and the latest data points recorded.

$$n^{\underline{o}}\ expected\ datapoints = \frac{latest\ registry - earliest\ registry}{average\ recording\ frequency}\ (1)$$

Equation 1 - Number of expected datapoints

$$\%\ Completeness = \frac{n^{\underline{o}}\ of\ registered\ datapoints}{n^{\underline{o}}\ of\ expected\ datapoints}\ (2)$$

Equation 2 - Completeness definition

- **Timeliness:** Defined as the percentage that the recording windows for the SI and the HR represent compared to the total GPS recording windows. Due to this comparison, the GPS dataset is not evaluated in terms of this indicator.

$$\%\ Timeliness = \frac{latest\ HR\ registry - earliest\ HR\ registry}{latest\ GPS\ registry - earliest\ GPS\ registry}\ (same\ for\ SI)\ (3)$$

Equation 3 - Timeliness definition

- **Uniqueness:** Defined as the percentage of unique registries compared to the total number of registered data points.

$$\% \, Uniqueness = \frac{n^{\underline{o}} \, of \, unique \, registries}{total \, n^{\underline{o}} \, of \, registries} \quad (4)$$

Equation 4 - Uniqueness definition

- **Accuracy:** Defined as the percentage of data points within the acceptable margins of acceptance compared to the total number of registries.

$$\% \, Accuracy = \frac{n^{\underline{o}} \, of \, acceptable \, registries}{total \, n^{\underline{o}} \, of \, registries} \quad (5)$$

Equation 5 - Accuracy definition

In Figure 8 can be observed the charts comparing the four indicators for the HR and the SI datasets.

Figure 8 - Data Quality indicators for the HR and SI datasets.

As the indicators are defined, correlations exist between them that can explain their values. In terms of timeliness, the resulting value is shockingly low, by average or by player indifferently. While the GPS records are taken during the complete five-day data period (although not at a constant frequency), only two specific time windows record the SI and HR, highlighted in a red square below (Figure 9)

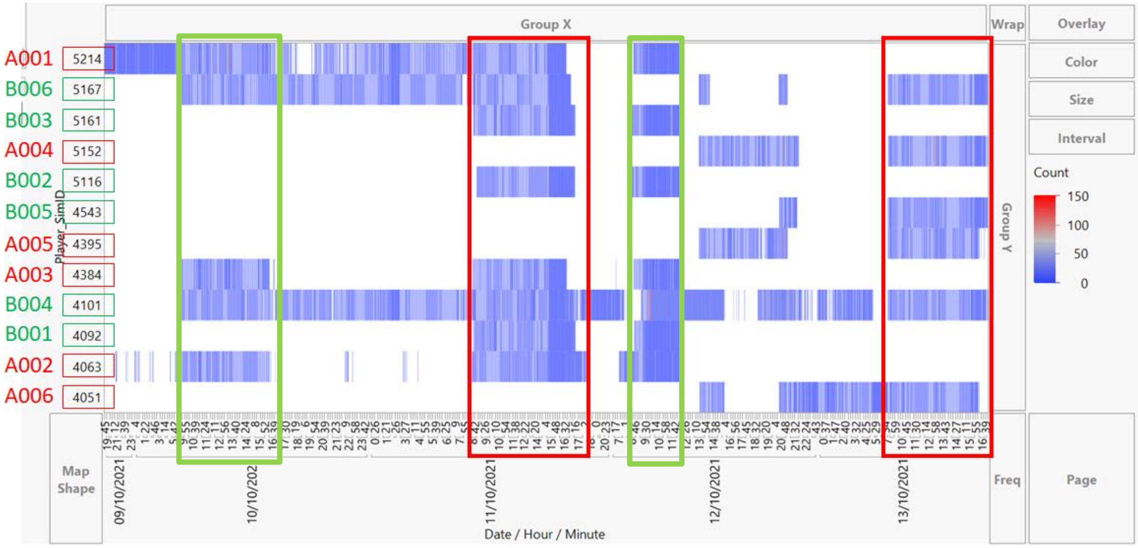

Figure 9 - GPS timestamp recordings along the 5-day training period. The areas inside the rectangular shapes are the time windows for which the HR and SI are recorded.

This is an important issue to solve since there are other periods (green squares) in which the recordings are defined in specific time windows for different players, making us believe they could be other training exercises with no biometric data recorded that make the timeliness indicator lower.

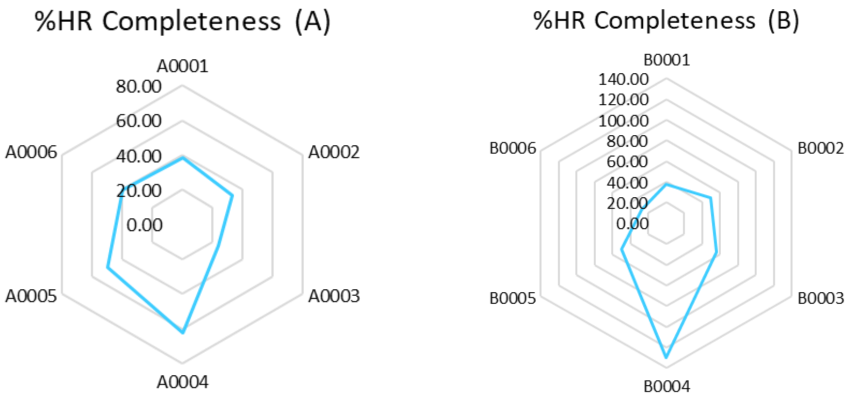

Figure 10 - Completeness indicator for the HR dataset.

The percentage of completeness has been observed to be low for most players (Figure 10). Being calculated with the average recording frequency, this means that most registries have a high recording frequency at a specific recording window but then some scattered datapoints exist in the complete recording window. This makes the expected number of datapoints higher, and it is why the percentage of completeness decreases. Therefore, the takeaway from this indicator is either that these gaps between registries need to be filled or that the sporadic registries need to be deleted for the analysis since the data needs to be continuous to approach the study from a time series perspective.

In terms of uniqueness, although the indicator is quite high in most players and datasets, there have been some specific cases in which a low indicator has been observed. Such is the case for player B0004 in the GPS dataset, which can be related to an unexpected case of high completeness.

Primary_Id	Expectedpoint	receivedData	percentageMissing	Completeness	repeated	Uniqueness
A0001	7728	2959	61.71	38.29	219	92.60
A0002	7539	2524	66.52	33.48	180	92.87
A0003	6062	1457	75.97	24.03	88	93.96
A0004	3294	2054	37.64	62.36	175	91.48
A0005	3251	1619	50.20	49.80	133	91.79
A0006	3205	1283	59.97	40.03	53	95.87
B0001	3312	1242	62.50	37.50	72	94.20
B0002	3306	1632	50.64	49.36	140	91.42
B0003	3311	1865	43.67	56.33	158	91.53
B0004	9512	12373	-30.08	130.08	2915	76.44
B0005	2469	1223	50.47	49.53	81	93.38
B0006	9512	2548	73.21	26.79	82	96.78

Table 1 - Completeness and Uniqueness indicators for GPS dataset.

As can be observed in Table 1 the low percentage of uniqueness (red) is corresponding to a high percentage of the completeness indicator (green). The received number of datapoints is much larger than expected, thus indicating that this excess of datapoints are repeated records. This is an important issue since like it has just been justified, the completeness value is quite low. In cases like B0004, it could be easily misleading the result since the real completeness value is probably low like in the other players, but it is the repeated registries that are incorrectly raising its value.

In terms of accuracy, most registries are within expected range of values.

Overall, it has been observed that the provided datasets have difficulties in terms of having a common recording frequency, which generates gaps between the timestamps for the different indicators. This may be a problem when building and training a model, since some architectures struggle to predict accurately an objective value when the datasets are not complete.

It is advised that the recording process should be improved so these frequencies are unified, making the registries complete and, therefore, having a prospect of reaching a better accuracy for the prediction model. Also, it has been detected those repeated values have been registered multiple times, which should also be avoided since it can confuse the model. It is true though, that this only happens on certain and very specific occasions.

So, in summary, we could say that in terms of accuracy and uniqueness, the data recording process is quite good, but in terms of completeness and timeliness, it has to improve and try to have a common recording frequency to have complete records of the player's data.

5 ANALYSIS

5.1 Data pre-processing

The first step of the analysis has been to manipulate the original datasets to be able to combine them in pairs for the preliminary analysis. The first step has been to delete the duplicated rows, followed by calculating the average of each numerical indicator by minute and by player. This, as it will be explained later, is to ease the analysis process for this preliminary stage.

Then, the data by pairs of indicators has been compared to see the shared timeframe of recordings, and a single sheet where both indicators can be seen has been created to later create visual charts that can be used to compare.

5.2 Preliminary analysis

Following the pre-processing stage and before joining the datasets into a single file to input into a prediction model, the relationships between different indicators have been studied. With the data combined by pairs of indicators explained in the previous section, charts have been created to visualize and find out about the indicator correlations. A priori, such correlations were the following:

- **Stress Index raise leads to Heart Rate raise:** When the player is put under stress, it is expected that the heart rate will rise. This correlation is unidirectional.
- **Player Speed raise leads to Heart Rate raise:** When the player is put under physical effort, it is expected that the heart rate will rise. The correlation is unidirectional.
- **Distance Player-Own team raise leads to Stress Index raise:** When the player is far from his own team, it is expected that the stress index will rise. This correlation is unidirectional.
- **Distance Player-Opposite team decrease leads to Stress Index raise:** When the player is close to the opposite team it is expected that the stress index will rise. This correlation is unidirectional.

- **Duration of exercise leads to Stress Index raise:** When duration increases, it is expected that players will become fatigated and, therefore, easy to stress. This correlation is unidirectional.

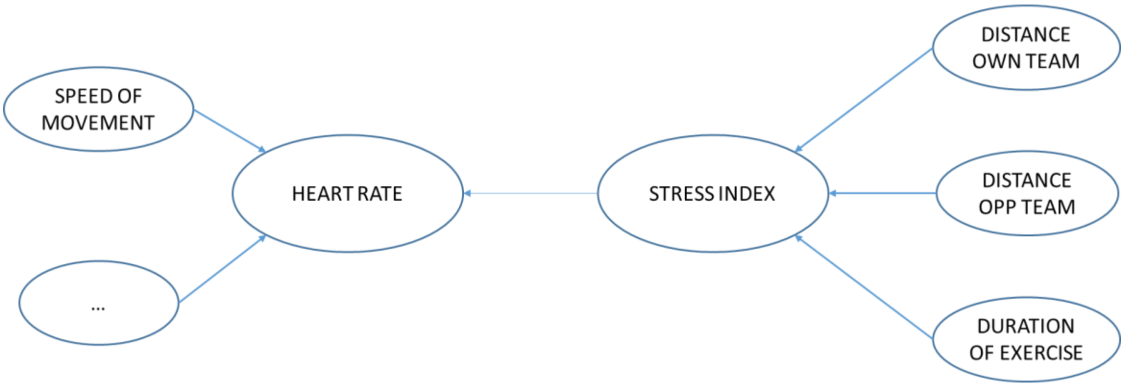

Figure 11 - Correlation between variables.

These correlations have been first explored using Excel sheets. Charts have been drawn to visually detect if these hypotheses show signs of being true, and in such cases, the intention is to prove them statistically.

The main issue to be solved for the Excel analysis has been the fact of the timestamps for the three main datasets do not coincide between them, as has been mentioned in Section 4, Figure 9. The recording frequency for the GPS coordinates is much lower than the one for the stress index and the heart rate (Figure 12). This adds to the problem that the recording windows for the GPS are much larger than the ones for the other two indicators. Therefore, only two specific time windows of GPS data can be analysed and compared to biometric data.

	nº of records	mean recording frequency
Stress Index	48154	1 seconds
Heart Rate	25275	2 seconds
GPS	32779	26 seconds

Figure 12 - number of records and recording frequency for each indicator.

Since this Excel analysis is a preliminary stage, it has been carried out by calculating the mean for each indicator by minute, to ease the process and solve the frequency difference between indicators.

5.3 Model development

For the model input data, a separate excel file has been created to analyse each correlation, since the model needed to be trained with the exact observations by second, and not the mean values by minute. Based on the timestamps that had both HR and SI values recorded, the GPS coordinates have been added with the information that can be extracted from them when existing (speed and distances).

Given that the speed is calculated from the GPS coordinates, the problem when it comes to relating it to the HR is the difference in recording frequencies. To solve this, each speed datapoint has been considered to be constant for thirty seconds. Therefore, a speed registry taken at, for example, the second half of the 9:45 am minute will be associated with all the HR values registered in the same 30-second period.

The same approach has been taken when relating the SI with the distance indicators. However, since the distance is considered to be less variable than the speed, the constant period for each distance registry (to own or opposite team) has been estimated in one minute long.

The complete dataset has been standardized to minimise the alteration that different data magnitudes could have in the training of the model. Also, it has been detected that the HR recordings in beats per minute units are redundant with the proposed frequency variables explained in 2.3, given that the RR peak intervals are related to bpm units as follows:

$$HR\ [bpm] = \frac{60000}{RR\ interval\ [msec]}\ (6)$$

Equation 6 - HR and RR interval correlation

This linear correlation can alter the model training process when the weights are assigned to each variable to predict the output. Therefore, the frequency domain variables have been removed from the complete dataset. The complete list of parameters included in the dataset is:

- **Player primary ID**: number identifying the player.
- **Heart Rate**: HR value.
- **Intensity**: Categorical value depending on the value for the HR and the effort it supposes for each player.
- **Team**: team to which the player belongs.
- **Age**: Player's age.
- **Weight**: Player's weight.
- **Height**: Player's height.
- **Speed**: Player's speed (if recorded).
- **Own distance**: Player's distance to its own team (if recorded).
- **Opposite distance**: Player's distance to the opposite team (if recorded).
- **Duration**: Duration of the exercise at the moment of the timestamp.

Overall, every registry (or line) of the dataset is going to present a full association for these variables, except for the GPS-related indicators (speed, own distance, opposite distance) that have not been registered continuously in time as has been explained above.

6 RESULTS AND DISCUSSION

6.1 Preliminary analysis

The results of the preliminary analysis have been useful to notice that some trends exist when relating the multiple indicators that are going to be analysed to try to build a model.

In the case of relating SI and HR, seems that the correlation is quite clear, showing a rise in HR when a SI peak appears (Figure 13). However it is true that not all SI peaks translate into an HR raise, and therefore there could be another factor that affects the last indicator.

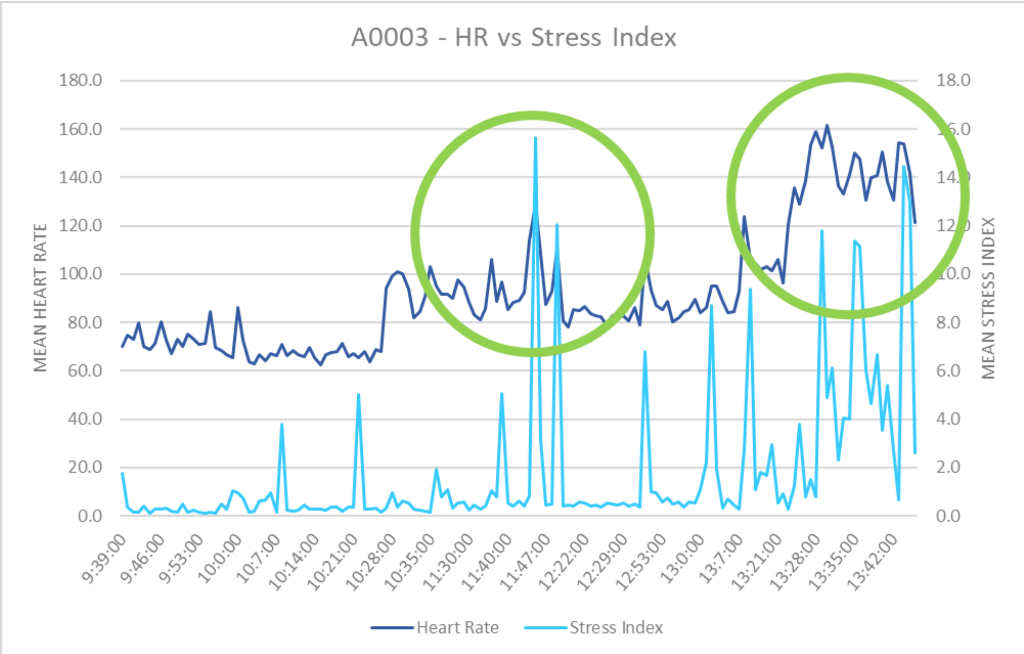

Figure 13 – HR vs SI correlation for player A0003.

This could be the case for the player speed, as can be seen for player A0003 (Figure 14). Some mild speed peaks do not show an increase in the heart rate rhythm towards the beginning of the exercise (circled in red), while a strong speed peak later in the exercise comes with a visible peak in the HR value.

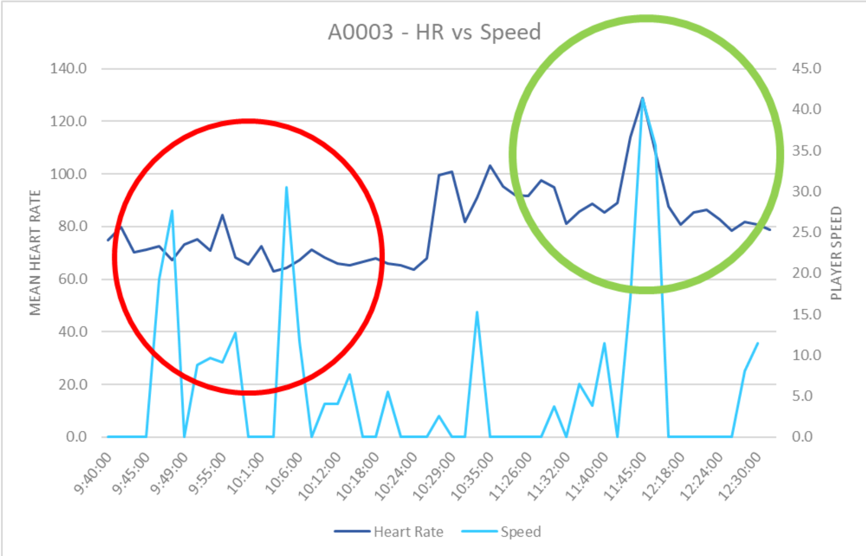

Figure 14 - HR vs Speed correlation for player A0003.

Similar is the situation when trying to relate the SI to the two distances defined in Section 5.2 (player-opposite team and player-own team), which show coincidence at some specific points, but discrepancies at others (Figure 15).

Figure 15 - SI vs Own Distance and Opposite Distance correlations.

Overall, in some cases, a correlation between variables exists, although not always the variation of one indicator influences the others in the same way. After this, the model building stage is approached with a better idea of how the output variable (SI) might be affected when changing the conditions for the other parameters.

6.2 Prediction models

Three types of prediction models have been tested to predict the Stress Index given a set of parameters. As it has been explained in section 5.3, the complete list included biometric and GPS indicators as well as individual parameters for each player to predict the SI.

6.2.1 Neural Network model

As has been reviewed in the Literature Review (Section 2.4.1), there is no specific approach when deciding on an architecture to start with a NN analysis. By empirical approach, in this project, we started with a single layer of hidden neurons. Multiple amounts of neurons were investigated and the most accurate result was selected to add another layer.

After trying models with three layers of hidden neurons, it has been found that adding a fourth layer makes the model suffer from overfitting, worsening the predicted results.

Model	Mean Absolute Error	Root Mean Squared Error	R2
Multilayer 50	1.0644	3.166	0.1986
Multilayer 40	0.5777	2.8	0.0745
Multilayer 30	1.1651	2.5612	0.3039
Multilayer 30-15	0.33	1.0271	0.6033
Multilayer 30-25	0.3416	1.0395	0.5558
Multilayer 30-15-25	0.2392	0.8295	0.5591
Multilayer 30-15-30	0.3004	0.9071	0.6241
Multilayer 30-15-35	0.3181	0.864	0.6243
Multilayer 30-15-45	0.2642	0.7788	0.575
Multilayer 30-15-45-20	1.0968	1.7726	0.5474
Multilayer 30-15-45-10	1.1332	2.1074	0.5481

Table 2 - Results for the NN models.

To measure the accuracy of each model, the predicted results are plotted against the expected values. Ideally, the prediction should equal the expected values and, therefore, the trendline for the plot should be a line with equation $prediction\ value = test\ value\ (equivalent\ to\ y = x)$. With this plot and the trendline, the R^2 value is observed, to measure how well the plot points fit the trendline, as well as to measure the slope of the resulting equation since $R^2 \cong m$ where m is defined according to $y = mx + n$. The best model has been the Multilayer Perceptron with three layers (30,15,35 neurons) with an R^2 value of 0.6243 (Figure 16).

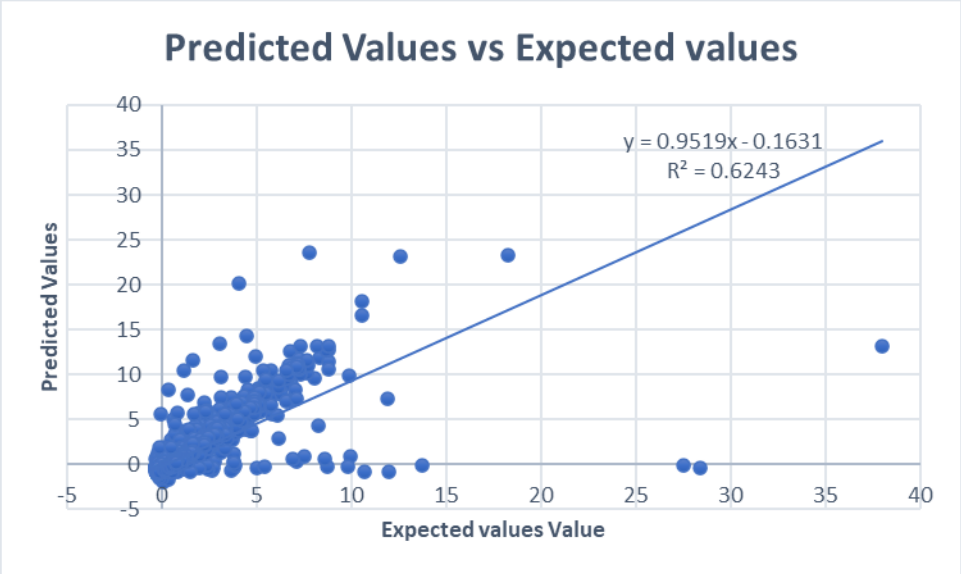

Figure 16 - Results for the Multilayer Perceptron model (30, 15, 35)

The predicted results do not show an excessive deviation from the expected values, and although the *m* value is quite high (0.91), the R^2 value is only 0.6243. The model does not show an overfitting problem as the duration of the exercise increases, which had been observed in other tests, although it is true that towards the end of the exercise the predictions become less accurate. Also, this model predicts better the SI around points of sudden variation of the expected values, although some excessively negative values (not valid) are observed for the prediction around those areas.

This type of model (NN) has been slow to simulate, so this value is considered the best this architecture can provide, also considering the overfitting problem observed with four hidden layers.

6.2.2 Regression model

The regression models have been tested with different values of ridge and testing batch size. Three combinations of these parameters have been tested and the results have been the same for all of them, thus indicating that the result reached was the best that this type of model could reach.

The resulting prediction model has an *m* value of 0.8301 and an R^2 of 0.6084 (Figure 17), thus making this prediction less precise than the one from the NN model. However, the results of these two types of models for the accuracy indicators are not very distant, and while it is true that the NN model predicts the SI more accurately, the Regression has been much quicker to simulate.

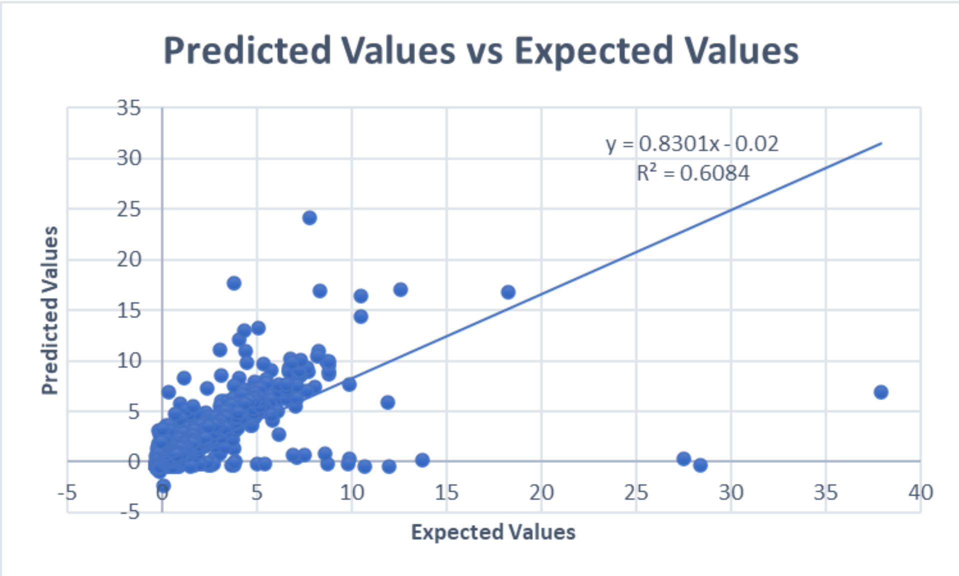

Figure 17 - Results for the tested regression models.

6.2.3 Random forest model

The last type of model to analyse has been the random forest model. The main difference with the other two has been the software used to build the model. While the NN and regression were developed using Weka, for the random forest Python has been used.

The same indicators have been used to measure the model's accuracy. The value for *m* is 0.938 and R^2 is 0.9132 (Figure **18**). These results are much higher than the ones obtained, which makes the model a better fit to predict the SI. It is also observed that the model does not increase the difference between the expected results and the prediction as the duration increases.Figure 19 how the first 20 records are practically identical for the prediction and expected values

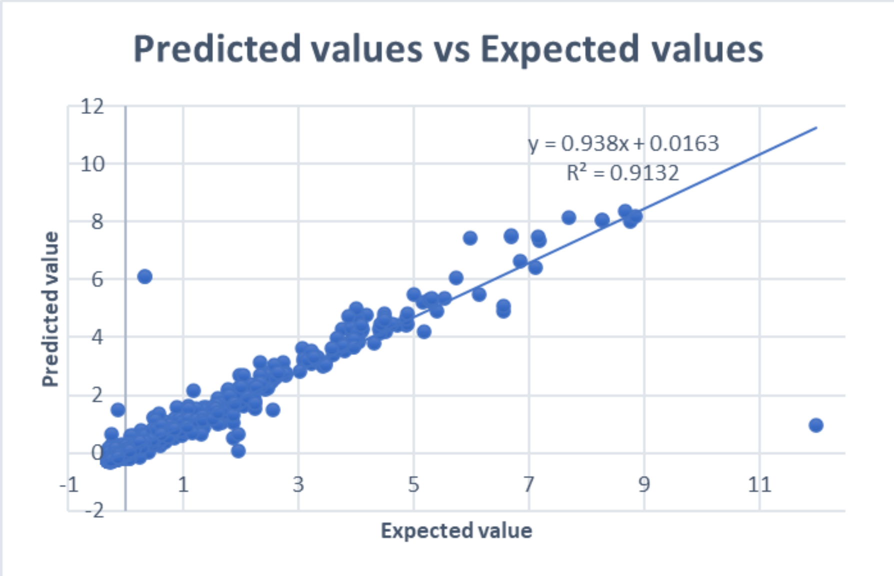

Figure 18 - Results for random forest model.

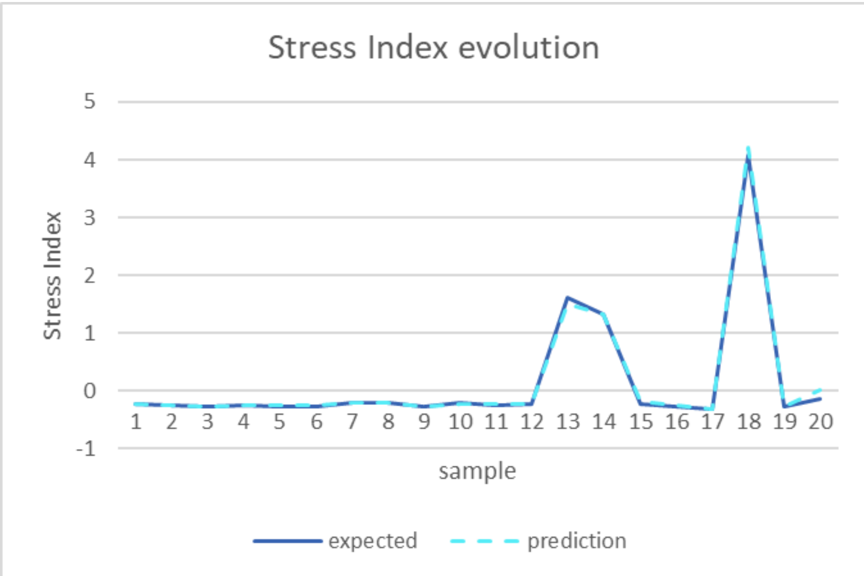

Figure 19 – Evolution for the stress index expected and prediction results.

The random forest model offers the possibility to visualise the importance of each variable, assigning the weight they have when predicting the output variable. They are:

variable	% importance
heartRate	0.4612
duration	0.1663
opp distance	0.1391
intensity	0.0732
own distance	0.0583
speed	0.0502
Height	0.0315
Weight	0.0119
Primary_Id	0.0049
Age	0.0024
Team	0.0012

Table 3 - Relative weight for each variable for the prediction in the random forest model.

6.3 Discussion

Once observed the results, it is necessary to discuss the takeaway from each section.

The preliminary analysis has been useful to detect some tendencies about the correlations that some indicators have between themselves. The correlations defined in Section 5.2 have been observed in some cases, confirming the biometric indicators can be affected like expected when certain exercise parameters vary. However, the tendencies were not clear enough to prove strong correlations since, in many cases, the indicators did not react in the way defined by the proposed correlations (Figure 14).

Therefore, it is deduced that many are the parameters that can affect the level of an indicator, specifically the SI which is the objective to predict and that its evolution is not only related to a single environment parameter. The three models that have been developed and built had the objective of predicting the SI based on learning how all the indicators can affect its evolution.

The regression model has been the one with the weakest results, given that it is the simplest model to build. Its methodology is to simply approximate a regression line that fits as well as it can the training dataset, which in this case has resulted in an R^2 of 0.6083. This result is probably affected by the fact that some of the variables fed into the model are not numerical but categorical, which this type of model struggles to manage.

Similar is the case of the multilayer perceptron (NN) model, for which the result has been an R^2 of 0.6243. Although this model architecture has much more potential than the regression model, it struggles to learn from datasets that present missing values in their records. As has been discussed in Section 4 and Section 5.3, the input data contains gaps in the fields corresponding to the GPS-related indicators (speed, own distance, opposite distance), which makes this type of model struggle when predicting the output variable. Also, as happened with the regression model, the categorical values are difficult for this model to analyse, so they probably worsen the prediction results.

On the other hand, the random forest model works by splitting the data according to threshold values, thus making it easier to manage categorical values like, in the case of our dataset, the Player Id or the Intensity. Also, this model works better with empty registries, so the results do not get too affected by it and the final R^2 value has been 0.9132.

Getting the importance (Table 3) **Error! Reference source not found.**for each variable has been a remarkable takeaway from the random forest model since it confirms that HR is the variable that mostly affects (46%) the level of the SI. It has been observed that the duration of the exercise also plays an important role in the prediction (16%), similarly to the distance from the player to the opposite team (13%). On the other hand, other variables like the ID for the player (Primary_Id), its team, and age are variables not quite relevant for the study, with an importance below 1%.

Remarkably, the distance to the opposite team indicator is on the top 3 of indicator importance. It is an indicator that directly comes from the GPS coordinates and, given the recording frequency difference between the SI and the GPS datapoints, it shows a lot of gaps in its registries. Nevertheless, it makes its way into the top 3, thus indicating that probably, if it had more complete registries, its importance would increase.

Overall, after conducting a literature review on how stress and heart rate are monitored and how team performance is evaluated, the unified dataset has allowed us to feed into the model the key variables that influence the evolution of the Stress Index. The data quality assessment of the raw datasets made us realize that the data had a lot of gaps due to the low values of timeliness and completeness of the GPS, SI, and HR indicators. The complexity in terms of empty registries made us adapt the input dataset, so the model could be trained with a file as complete as possible, which has been key for the performance of the models.

7 CONCLUSION AND FUTURE WORK

This project had the aim of building a model that could predict the evolution of the SI while also assessing the quality of the data provided in its measurements. Starting with a literature review, it has been detected that most stress monitoring studies were focused on a medical approach, while none of them was based on team performance evaluation. Having reviewed the most common approaches for these two domains, a single dataset to contain all the information from the raw files has been created.

The main challenge has been to develop a single file as complete as possible since the provided raw datasets had a lot of missing registries. The recording frequency for each indicator was different, leading to low percentages of timeliness and completeness. The accuracy and uniqueness showed no major issues, with most records being unique and within acceptable limits.

With the preliminary analysis, some correlations between indicators have been observed. However, they were not strong enough to be proved statistically. Therefore, it has been left to the modelling stage to observe how well the models could predict the SI given the other variables.

The NN and Regression models had a low precision score for their predictions, mainly because the data had categorical values. However, the random forest managed much better this type of variable and reached a higher percentage of precision, determining the HR, the duration of the exercise, and the distance from the player to the opposite team as the main variables to influence the SI.

With these results, it remains for future studies to keep exploring the random forest models to predict the SI since many of their parameters can be modified to see if the precision increases. Also, this project has only focused on three types of models, while trying other types could help to improve the prediction. It also remains for future projects to try to relate the studied datasets with biometric and GPS information to the feedback and external evaluation datasets that were also recorded from the training exercises.

REFERENCES

10 Predictive Modeling Types (With Benefits and Uses) | *Indeed.com*. (n.d.). Retrieved August 17, 2022, from https://www.indeed.com/career-advice/career-development/predictive-modeling-types

Agatonovic-Kustrin, S., & Beresford, R. (2000a). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, *22*(5), 717–727. https://doi.org/10.1016/S0731-7085(99)00272-1

Agatonovic-Kustrin, S., & Beresford, R. (2000b). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, *22*(5), 717–727. https://doi.org/10.1016/S0731-7085(99)00272-1

Aguinis, H., & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology*, *60*, 451–474. https://doi.org/10.1146/ANNUREV.PSYCH.60.110707.163505

Ali, R. (2020). *Predictive Modeling: Types, Benefits, and Algorithms* | *NetSuite*. https://www.netsuite.com/portal/resource/articles/financial-management/predictive-modeling.shtml

Alvarez, K., Salas, E., & Garofano, C. M. (2016). An Integrated Model of Training Evaluation and Effectiveness: *Http://Dx.Doi.Org/10.1177/1534484304270820*, *3*(4), 385–416. https://doi.org/10.1177/1534484304270820

Antonopoulos, C. (2022). *The Importance of Data Quality Timeliness*. https://www.measuredresultsmarketing.com/importance-data-quality-timeliness/

Bailey, K. (2019). *Gaussian Processes for Dummies ·*. https://katbailey.github.io/post/gaussian-processes-for-dummies/

Bechini, G. (2007). Performance Diagnostics and Measurement Selection for On-Line Monitoring of Gas Turbine Engines. *Ph.D. Thesis, Cranfield University*.

Chan, P., Lebel, F., Owusu-Banahene, E., Wilmes, K. (2021). Data Cleansing for Asset Maintenance Applications. *Ph.D. Thesis, Cranfield University*.

Chauhan, N. S. (2022). *Decision Tree Algorithm, Explained - KDnuggets*. https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html

Coutts, L. v., Plans, D., Brown, A. W., & Collomosse, J. (2020). Deep learning with wearable based heart rate variability for prediction of mental and general health. *Journal of Biomedical Informatics*, *112*, 103610. https://doi.org/10.1016/J.JBI.2020.103610

Dadd, D., & Hinton, M. (2021). *Performance Measurement: Evaluating the Effectiveness of E-Learning - Open Research Online*. PMA 2021: Performance Management Association Biannual Conference. http://oro.open.ac.uk/77931/

DeLone, W. H., & McLean, E. R. (2016). Information Systems Success Measurement. *Foundations and Trends® in Information Systems*, *2*(1), 1–116. https://doi.org/10.1561/2900000005

Dhar, A., Mukherjee, H., Dash, N. S., & Roy, K. (2021). Text categorization: past and present. *Artificial Intelligence Review*, *54*(4), 3007–3054. https://doi.org/10.1007/S10462-020-09919-1/TABLES/11

Faouzi, J. (2022). *Time Series Classification: A review of Algorithms and Implementations*.

Fujiwara, K., Miyatani, S., Goda, A., Miyajima, M., Sasano, T., & Kano, M. (2021). *Autoencoder-Based Extrasystole Detection and Modification of RRI Data for Precise Heart Rate Variability Analysis*. https://doi.org/10.3390/s21093235

Ghosh, A. (2020). *What is Data Consistency?* https://thecustomizewindows.com/2020/01/what-is-data-consistency/

Gupta, A. (2022). *The 6 Dimensions of Data Quality | Collibra*. https://www.collibra.com/us/en/blog/the-6-dimensions-of-data-quality

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.

Islam, R. (2019). *Brain Tumor Detection from PET (Positron Emission Tomography) using Artificial Neural Network (ANN).* Khulna University of Engineering and Technology. https://www.academia.edu/41145223/Brain_Tumor_Detection_from_using_Artificial_Neural_Network_ANN_Brain_Tumor_Detection_from_PET_Positron_Emission_Tomography_using_Artificial_Neural_Network_ANN_Positron_Emission_Tomography_Image

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: a review. *Undefined*, *33*(4), 917–963. https://doi.org/10.1007/S10618-019-00619-1

Kaushik, S. (2020). *(46) Understanding the difference between Data Accuracy and Validity | LinkedIn*. https://www.linkedin.com/pulse/understanding-difference-between-data-accuracy-validity-kaushik/

Keržič, D., Alex, J. K., Alvarado, R. P. B., da Silva Bezerra, D., Cheraghi, M., Dobrowolska, B., Fagbamigbe, A. F., Faris, M. A. I. E., França, T., González-Fernández, B., Gonzalez-Robledo, L. M., Inasius, F., Kar, S. K., Lazányi, K., Lazăr, F., Machin-Mastromatteo, J. D., Marôco, J., Marques, B. P., Mejía-Rodríguez, O., … Aristovnik, A. (2021). Academic student satisfaction and perceived performance in the e-learning environment during the COVID-19 pandemic: Evidence across ten countries. *PLOS ONE*, *16*(10), e0258807. https://doi.org/10.1371/JOURNAL.PONE.0258807

Kim, F. (2020). *What is Data Accuracy, Why it Matters, and How Companies Can Ensure They Have Accurate Data? - Data Ladder*. https://dataladder.com/what-is-data-accuracy/

Lawton, G., Carew, J. M., & Burns, E. (2022). *What is Predictive Modeling?* https://www.techtarget.com/searchenterpriseai/definition/predictive-modeling

Li, Z., Wang, C., Han, M., Xue, Y., Wei, W., Li, L. J., & Fei-Fei, L. (2017). Thoracic Disease Identification and Localization with Limited Supervision. *Advances in Computer Vision and Pattern Recognition*, 139–161. https://doi.org/10.48550/arxiv.1711.06373

Liu, D., & Ulrich, M. (2013). Listen to Your Heart: Stress Prediction Using Consumer Heart Rate Sensors. *Final Project, Stanford CS*, *229*.

Mihăiloaie, C. (2015). *Performance Magazine Data quality dimensions – from Accuracy to Uniqueness - Performance Magazine*. https://www.performancemagazine.org/data-quality-dimensions-from-accuracy-to-uniqueness/

Mohamed, Z. E. (2019). Using the artificial neural networks for prediction and validating solar radiation. *Journal of the Egyptian Mathematical Society 2019 27:1*, *27*(1), 1–13. https://doi.org/10.1186/S42787-019-0043-8

Pappas, C. (2015). *8 Tips To Measure Your Online Training Effectiveness - eLearning Industry*. E-Learning Industry - Corporate e-Learning. https://elearningindustry.com/8-tips-to-measure-your-online-training-effectiveness

Patel, J., & Patel, A. (2016). Artificial Neural Networking in Controlled Drug Delivery. *Undefined*, 195–218. https://doi.org/10.1016/B978-0-12-801559-9.00010-7

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data Quality Assessment. *Communications of the ACM*, *45*(4), 211–218. https://doi.org/10.1145/505248.506010/FORMAT/PDF

Podgorelec, V., Kokol, P., Stiglic, B., & Rozman, I. (2002). Decision trees: an overview and their use in medicine. *Journal of Medical Systems*, *26*(5), 445–463. https://doi.org/10.1023/A:1016409317640

Predictive Modelling: Splitting Data into Training and Test Set - Finance Train. (n.d.). Retrieved August 17, 2022, from https://financetrain.com/predictive-modelling-splitting-data-training-test-set

Raspopovic, M., & Jankulovic, A. (2017). Performance measurement of e-learning using student satisfaction analysis. *Information Systems Frontiers*, *19*(4), 869–880. https://doi.org/10.1007/S10796-016-9636-Z/TABLES/8

Salas, E., Milham, L. M., & Bowers, C. A. (2003). Training evaluation in the military: Misconceptions, opportunities, and challenges. *Military Psychology*, *15*(1), 3–16. https://doi.org/10.1207/S15327876MP1501_01

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*, 258. https://doi.org/10.3389/FPUBH.2017.00258

Shashikant, R., Chaskar, U., Phadke, L., & Patil, C. (2021). Gaussian process-based kernel as a diagnostic model for prediction of type 2 diabetes mellitus risk using non-linear heart rate variability features. *Biomedical Engineering Letters*, *11*(3), 273–286. https://doi.org/10.1007/S13534-021-00196-7/TABLES/9

Šprogar, M., Kokol, P., Zorman, M., Podgorelec, V., Yamamoto, R., Masuda, G., & Sakamoto, N. (2001). Supporting medical decisions with vector decision trees. *Studies in Health Technology and Informatics*, *84*, 552–556. https://doi.org/10.3233/978-1-60750-928-8-552

Sun, Y., Moss, G. P., Prapopoulou, M., Adams, R., Brown, M. B., & Davey, N. (2008). Prediction of Skin Penetration Using Machine Learning Methods. *Undefined*, 1049–1054. https://doi.org/10.1109/ICDM.2008.97

Sun, Y., Peng, Y., Chen, Y., & Shukla, A. J. (2003). Application of artificial neural networks in the design of controlled release drug delivery systems. *Advanced Drug Delivery Reviews*, *55*(9), 1201–1215. https://doi.org/10.1016/S0169-409X(03)00119-4

Vassallo, D. (2019). *Exploring Gaussian Process vs Linear Regression – David Vassallo's Blog*. https://blog.davidvassallo.me/2019/03/15/exploring-gaussian-process-vs-linear-regression/

Wang, J. (2022). *An Intuitive Tutorial to Gaussian Processes Regression*. https://github.com/jwangjie/

What Is a Decision Tree? (2022). https://www.mastersindatascience.org/learning/introduction-to-machine-learning-algorithms/decision-tree/

What We Do - Babcock International. (n.d.). Retrieved August 17, 2022, from https://www.babcockinternational.com/what-we-do/

Woodall, P., Gao, J., Parlikad, A., & Koronios, A. (2015). Classifying data quality problems in asset management. *Lecture Notes in Mechanical Engineering*, *19*, 321–334. https://doi.org/10.1007/978-3-319-09507-3_29

APPENDICES

Appendix A - Cures Approval Letter

15 June 2022

Dear Mr Rodrigo Corominas ,

Reference: CURES/16346/2022
Title: Development of an Intelligent Approach for Delivering High Performing Training Solutions

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your CURES application, reference CURES/16346/2022 **has been reviewed. You may now proceed with the research activities you have sought approval for.**

If you have any queries, please contact CURES Support.

We wish you every success with your project.

Regards,

CURES Team