Opleiding I&E

Modeling of the rating of perceived exertion for training sessions in wheelchair tennis.

Lars Schrijver

Supervisors:
Arie-Willem de Leeuw & Arno Knobbe

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

15/12/2020

Abstract

In the past, the training load for athletes was determined by solely relying on the expertise of the coach. Now, with the current technological developments, it is possible to perform objective measurements and investigate what the optimal training load is. In this thesis, we study the relationships between training characteristics and the Rating of Perceived Exertion (RPE) for training sessions in wheelchair tennis. Here, the training attributes are measured by sensors on the wheelchair of the athlete and the Rating of Perceived Exertion is asked from the athlete after each training session. While investigating the data, we found out that there was not enough data for male wheelchair tennis players. This is why we focused on the female athletes for the rest of the thesis. After all the preprocessing was finished, 24 training sessions were suitable for our thesis.

We have used several machine learning techniques to model the Rating of Perceived Exertion. First, we applied LASSO regression and regression trees. Moreover, we used subgroup Discovery to investigate cases where big or small values give sub-optimal results. The results showed us the accuracy of the machine learning techniques. The accuracy is defined with the R^2-score. The R^2-score score was 0.108 for the LASSO regression and 0.180 for the regression trees. So, the model built with regression trees predicted the training load the most accurate. The parameter α of the LASSO regression model was more stable than the parameter ccp_α of the regression trees model. The regression trees model showed three outliers for the ccp_α. However, because these parameters have different purposes, they were both stable enough. LASSO regression showed us that the Average Velocity and the percentage of the time the athlete spends in the rotation speedzone of higher than 100 are their most important variables to quantify the relationship between the selected Rating of Perceived Exertion and the training characteristics. The coaches could try to focus more on interval training, because of the fact that average speed was an important factor for predicting the RPE. Also, it would be helpful to include specific upper-body repeated power ability drills in the physical preparation, because heavy rotating of the wheels gave the athletes a heavier training session on a physical level. With Subgroup Discovery, we were able to find out that the subgroup where there have been made less than or equal to 351 turns to the left and the number of times the speed is below zero is higher or equal to 145 is a characteristic for a relatively easy training. One thing a coach or an athlete has to keep in mind is that attributes that impact the athlete the most are the attributes where the training volume and intensity both are high. These are training attributes where the feature itself defines the intensity (a high velocity) and the value defines the volume (200 times this training session).

Contents

1 Introduction

1.1 Research context

Many people are unaware that tennis is one of the most popular sports in the world. Tennis is a sport where you defend your half of the court and try to hit the ball with a racket over the net and between the lines on the other side. According to WorldAtlas, tennis has 1 billion fans and is the fourth most popular sport in the world. In this thesis, we will focus on wheelchair tennis. Wheelchair Tennis is one of the forms of tennis adapted for wheelchair users. In wheelchair tennis, you can win as much as 200,000 pounds in a single tournament. Wheelchair tennis was introduced by a man called Brad Parks in 1976. Parks became disabled after he had gotten in a skiing accident. When Brad was working on his rehabilitation he met a fellow patient called Je Minnebraker. The two then decided to play some tennis from within their wheelchairs across the street from the hospital. When they were playing some people from the next court shouted: \Why do you guys even bother?". After that, the guys wanted to make it a commonly accepted and professional sport. 40 years later they build a sport that is recognized as the most professionalized disabled sport in the world [15]. The rules are similar in comparison to those applied by tennis. The size of courts, balls, and rackets are the same, but there are two major di erences. The athletes use specially designed wheelchairs and the ball is allowed to bounce twice on your side of the court before you hit it.

To perform in wheelchair tennis, both tactical and physical aspects are important. For the latter, it is important to construct the training schemes such that the athletes are optimally prepared for the competition. Research shows us that wheelchair tennis matches are medium to high in aerobic tness. It also shows that the athletes have a work:rest ratio between 1:1 and 1:4 [4]. This gives us an indication of the physicality in wheelchair tennis. This physicality is de ned by the training load (TL) an athlete has endured in training. A daily change in training load (TL) will potentially optimize the athletes' physical performance. It can also have a positive e ect on the amount of overuse injuries and illness [2]. Training load (TL) is dependent on training volume and training intensity. If you understand the TL completed during exercise, you achieve the outcome that was intended without compromising your health [18]. Former research found that the Rating of Perceived Exertion is a valid method for monitoring training load, which is something to remember when working on this thesis [17].

The reason for investigating this topic is that not many scientists have conducted research to obtain a clear answer to our research question with data from actual professional wheelchair tennis players. Another big reason for this thesis is that when we nd out what makes a certain training tougher than other training sessions, we can help tennis players prevent injury caused by excessive physical e ort. Data-driven insights, indicating the impact of training sessions and their speci c attributes, can help coaches who coordinate training sessions and training schedules, but also help athletes to acquire more insight in how to optimize their performances. One of the problems of investigating this topic is that it is hard to gather the data needed for this thesis. That is why we will investigate this in consultation with the Koninklijke Nederlandse Lawn Tennis Bond (KNLTB). The Koninklijke Nederlandse Lawn Tennis Bond organizes wheelchair tennis training sessions and has people in operation who can connect sensors to the wheelchairs. These two things are essential for gathering the right and enough data.

In this thesis, we are searching for a relation between training attributes and the amount of physical load that is perceived by an athlete. The goal is to help players and coaches give insight in what e ect a certain training attribute has on the athlete and what they should avoid when the athlete is in danger of physical overload. It can also help athletes deliver better performances, which will result in more prize money. In the past, coaches and players guessed the optimal training load, but now it can be professionalized with the current technical developments. This brings us to the following research question:
\How can we quantify the relationship between the selected Rating of Perceived Exertion and the training characteristics obtained from sensor data in wheelchair tennis?"
We research this with the help of linear modelling and Subgroup Discovery in order to select key features and provide predictive models.

1.2 Thesis outline

We will start by giving an overview of the de nitions and background that is needed for understanding the remainder of this thesis in Chapter 2. In Chapter 3, we will we will discuss the data that is used for this thesis. Hereafter, We will discuss the decisions we made regarding the machine learning techniques used for this thesis in chapter 4. In chapter 5, we will tell more about the results these machine learning techniques provided. At last, we critically look at our research in chapter 6 and elaborate on our ndings in chapter 7.

2 Theoretical basis

In this section, we will list all the theoretical terms, ideas and models that are used throughout this research. We will discuss the Sport Science and Data Science concepts separately.

2.1 Sport science

In this subsection, we will elaborate on the important Sport Science de nitions that are relevant in our research. Listed below are some terms that will help understand the sport side of this thesis.

2.1.1 Work:Rest ratio

This ratio shows how the amount of physical work compares to the amount of rest an athlete endures for a certain workout or during a match. If the ratio is for example 2:1, the athlete works two times as much as the amount of rest [19].

2.1.2 Training volume

This term refers to the duration of a workout or training. This can for example be reported in terms of time and distance. In terms of time, you can think of minutes per day and hours per week. In terms of distance covered, you can think of 80 kilometers a week for a runner or 350 kilometers a week for a cyclist [18].

2.1.3 Training intensity

This term shows us how hard an athlete trains. This can be measured with many methods. Some of the methods are , including heart rate, oxygen consumption, weight lifted, power output, blood lactate concentration, or the rating of perceived exertion (RPE) [18].

2.1.4 Training Load (TL)

There are two variations of the Training Load (TL), external and internal TL. External TL represents the physical work performed during the training session or match. Internal TL stands for the associated biochemical (physical and physiological) and biomechanical stress responses. This is the TL that is experienced by the athlete [20]. External Training load is the function of the training volume and the training intensity, which are mentioned in section 2.1.2 and 2.1.3 of the theoretical basis. The external training load can be calculated using the following equation:

$$\text{External } TL = \text{training intensity} \qquad \text{training volume}$$

The data that needs to be gathered in order to calculate the external training load, can be found in a few ways. Some of the options are with a heart rate monitor, the amount of distance covert or the total weight lifted. An option to measure the internal training load is with the commonly known session-RPE method [18].

2.1.5 Rating of Perceived Exertion (RPE)

The Rating of Perceived Exertion (RPE) describes the level of exhaustion at the end of an exercise or training session on a scale from 0 to 10 [1]. The RPE is conducted by asking the athletes how tough the training session was perceived right after the training is completed. The highest value 10 is described as the highest physical load that can possibly be perceived in a training or match. The lowest value of zero can occur when the athlete experiences the lowest physical load possible in a training or match [2]. The RPE method for quantifying training load is a great option because of its low cost and because it is easy to understand and relatively simple to implement [18]. When the Rating of Perceived Exertion is multiplied with the duration of a training session, we call it the sRPE.

2.1.6 Training attributes/characteristics

These training attributes tell you something about the actions performed by athletes while training or playing a match. It shows you how much or how long a certain action has occurred. This can for example be how many sprints from 2.5 meter/seconds an athlete has performed this match.

2.2 Data science

For this part of the theoretical basis, we will discuss the terms related to the data science side of this thesis. It will help us understand the data science methods and techniques.

2.2.1 Cross validation

We use cross validation in order to prevent the predictive model from over tting and to tune our parameters. There are multiple di erent types of cross validation, but we only discuss the ones that we used in this thesis.

2.2.2 K-fold Cross validation

K-fold cross validation is a technique in which the dataset is randomly divided into k pieces of the same size. The k-1 samples are utilized for training the predictive model. The sample that is not used for training will be used for testing the model. This routine is than repeated k times. Every sample is used for testing once and K-1 times for training the predictive model. By gathering all test scores, we can determine the accuracy for a certain model and parameter settings. If you repeat this for di erent models or parameter settings, it is possible to nd out what model is the most accurate. A bene t of this technique is that it can give an accurate outcome without over tting even though you have a small dataset available [5].

2.2.3 Leave-one-out cross validation

Leave-one-out cross validation is a special kind of cross validation where k is equal to the number of data points [11]. It repeatedly uses one data point for testing and the others for training the predictive model. Every datapoint is used for testing once [12]. This technique is only relevant for smaller datasets, because of the computational cost it requires.

2.2.4 Nested cross validation

Nested cross validation makes use of inner fold cross validation and outer fold cross validation. There are k outer folds that we use for testing the accuracy of the model. The other k-1 folds are used to train the model. In the inner fold we try to nd the best values for our parameters. With these parameter settings we test the model in the outer fold cross validation for new data. In the outer fold we predict the target variable for new data in order to calculate the accuracy of the model. The accuracy is calculated by comparing the predicted values with the actual values [13].

2.2.5 LASSO (Least Absolute Shrinkage and Selection Operator) regression

Regression models can calculate the predicted risk or possible outcome. If you apply standard regression models, there is a big chance the model will over t by including many variables to predict. Furthermore, standard regression models tend to have the problem that they don't do well predicting corner cases. LASSO regression is a tool that can select the most useful variables which addresses both of these problems [5]. To elaborate on that, a LASSO regression model takes in a parameter called , which determines how nonzero weights should be handled. If is zero, you are basically using linear regression. This can be determined by making use of one of the di erent types of cross validation [6]. This is essential to avoid over tting when the number of datapoints is lower than the number of features [7].

2.2.6 Regression trees

Decision trees are a form of machine learning. When you provide data, the tree will slowly build itself by providing tests where the answer is the branch you have to follow. This process happens until every leave is a possible outcome to your problem. Decision trees are employed to solve regression (target is of numerical value) problems and classi cation (target is of nominal value) problems. Regression trees and model trees are special types of decision trees developed for regression problems. However, the main di erence between model trees and regression trees is that the leaves of the regression trees have a constant value, while model trees hold linear models in their leaves which can predict numeric values for a given data sample [14]. For this thesis, we use the decision tree regressor package to implement the regression trees.

To prevent over tting, it is important to apply pruning. Here, we have two options: Pre-pruning and post pruning. To elaborate some more on pre-pruning we have selected two techniques.
1. There can be another split at a certain node if the number of observations proving that node is higher than a number X.
2. There can be another split at a certain node if the residual sum of squares can be decreased.

In post pruning, we rst build the model and afterwards remove some branches. To decide on which branches to remove, we use complexity measures such as Minimal cost complexity [8]. This pruning method recursively nds the node with the \weakest link". The subtree, with the largest cost complexity that is smaller than a parameter called ccp_, will be chosen. So you set to a value which is the highest complexity that is acceptable for you [16]. If ccp_ =0, the biggest tree will be chosen because the complexity penalty term is essentially dropped. If ccp_ has a high value, only the root node will remain [21].

2.2.7 Subgroup discovery

We have so far discussed two forms of regression that we will use in order to build a predictive model. This will give us linear dependencies, but we also want to investigate possible non-linear relationships [9]. SD is a data mining framework that is capable of finding some subgroups in the dataset that gives us surprisingly high or low results. Here, the exceptionality of a subgroup is quantified by a quality measure. This measure combines the size of the subgroup and the difference between the target value of the subgroup and the entire data set. The user of SD can also set some constraints for the framework. For example, it is possible to set a search width and the number of attributes in the subgroup. There are certain key parameters/options when you apply subgroup discovery on a dataset. Obviously, you have to select a target attribute and a quality measure like Z-score. You also have to select which strategy you want to use to find subgroups in your dataset. The strategy we want to elaborate on is called beam search. This strategy works from top to bottom, starting with the consideration of single condition subgroups. The defined quality measure determines if a subgroup is added to the beam and at what position. If a subgroup is added, it can be expanded with a new condition depending on the parameter that defines the maximum number of conditions in a subgroup. This parameter is called search depth [9]. In addition, you set the search width (beam width). The beam width determines how much are stored during the search. If you increase the search width, you also increase the computational time [10]. Finally, SD will provide subgroups with a quality. The algorithm tests a lot of possible features (hypotheses) and will sort them on their quality. For validating these subgroups, we compare those with a high quality to random descriptions/subsets that are predicted to have an average quality. This will give us a threshold, If the quality is higher than this threshold, the subgroup is statistically significant. The subgroup with the highest quality may appear to be significant, even though this is a result of the many models tested.

2.2.8 R^2 (coefficient of determination) regression score function

There are multiple ways that you can check the accuracy of a model. For our research we made use of the R^2 score to test the accuracy of our regression models. R^2 (coefficient of determination) is a statistical measure that shows the proportion of variance of the regression model. The formula for this Coefficient is:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{SS_{res}}{SS_{tot}}$$

where:

SS_{res} = Sum of squares of the residual errors
SS_{tot} = Total sum of the errors

The R^2-score can be between 0 and 1. The closer the R^2-score is to 1, the more accurate the model is. If the R^2-score is 0, the model predicts just as good as taking the mean from the training target variable.

3 Data

In this chapter, we will discuss the data that we used in this research. To provide initial insights into the data, the datasets will be explored by using visualizations and summary statistics. Additionally, it is explained how the datasets are preprocessed. Finally, some limitations of the available datasets are discussed.

3.1 The data sets

For this research, we used two datasets that were gathered by the Koninklijke Nederlandse Lawn Tennis Bond (KNLTB). These are two spreadsheets gathered at training sessions from professional wheelchair tennis players. The first dataset contains the Rating of Perceived Exertion (RPE). The second dataset contains sensor data gathered from the sensors placed on the wheelchairs. For instance, the total duration and the number of times a sprint from longer than 3 meters has been performed. This data is accessible online and we have received permission to use this data collection for scientific purposes.

3.1.1 The RPE dataset

This dataset contains 1051 rows and 19 columns of data. The rows tell us how many training sessions are recorded and the colums are the training features. A few features from this dataset are training date, training id, duration and RPE. Also, 16 pro wheelchair tennis players have contributed to this dataset. 465 training sessions have a corresponding RPE value. The RPE value of other training sessions have not been documented or the player did not fill it in.

3.1.2 The sensor dataset

In table 6 (see appendix A), a list has been provided with the attributes that are relevant in this thesis and their explanation. The thing all the attributes have in common is that they are almost all about Velocity (m/s), Acceleration(m/s^2), Rotation Velocity (deg/s) or rotation acceleration (deg/s^2). There is a lot of different information available about these attributes like their average and maximum values. It is also possible to find the value for these attributes when the athlete makes a turn or moves in a curve. For the speed, distance and acceleration also different zones are considered. Here, a zone is defined by a lower and upper limit on one of the aforementioned characteristics. The features that are considered are the percentage of the time or distance covered in a specific zone. For example, there is a feature called Distance sz 3.5+ which describes the distance covered in a speedzone of higher than 3.5 m/s. So, how much distance an athlete has moved with a speed higher than 3.5 m/s. The amount of left and right turns can also be found in the dataset. In table 1, a summary is shown with information about the features relevant to this thesis. There are in total 71 features relevant for this thesis and they are categorised on the information they hold. Some features tell us more about the velocity of the athlete and others about the acceleration.

Total amount of features	71
Features related to speed	23
Features related to acceleration	15
Features related to rotation speed	12
Features related to rotation acceleration	11
Other features	10

Table 1: Information about the features that are relevant to this thesis. There are 71 features.

3.2 Preprocessing the data

First, we match both data sets on training session to obtain information about for how many sessions we both have sensor data and a Rating of Perceived Exertion (RPE). We could now see how much training sessions had complete sensor data and a corresponding RPE. After this, We have one dataset with on the left side the sensor data and on the right side the corresponding RPE value. In this dataset, it was easy to see how many data each individual player had. We found out that the male wheelchair tennis players had a small amount of data that is why we focused on the female tennis players for the rest of our thesis. This is why we removed the data collected for male athletes from the dataset and proceeded with the data from female athletes.

In this dataset, a lot of data was irrelevant for this thesis. So, the next step was to remove all the data that we had no use of regarding our analysis. This was data that was not related to the training load like the date of the training and all sort of id's. Now that we have removed the columns that we don't need, we have two remaining categories of variables. We have feature variables and one response variable. The RPE is our response variable, because it is a result from the sensor data which are the feature variables. The response variable will be stored in the variable y. The feature variables will be stored in x. A simplified table from the final table is shown in table 2. Before we can analyze our dataset, we split the data in a training and test set. The test set will be as small as possible, because the small size of our dataset. After these actions, data of 24 training sessions is ready to be analyzed.

Training session attributes		Target Variable
Avg. velocity	....	RPE
1.5	....	8
....	....	....

Table 2: A simplified representation of how the final dataset looks. This dataset has Training session attributes on the left side that determine what kind of training it was and how much training load an athlete has provided. On the right side, we have the Rating of Perceived Exertion per session, which indicates how tough the training was perceived by the athlete.

3.3 Limitations of the data

There are a few limitations with the data that we have to deal with. For one, we only have 24 data points, because we have focused on female athletes and we need a full set of sensor data and corresponding RPE values. Another limitation to our dataset is that we have more training session attributes than data points, which enlarges risks of over tting. The distribution of the RPE values is heavy on the middle values as shown in gure 1. This makes it hard to predict "extreme" high or low values, because the machine learning techniques have a lower amount of data for training with these values.

Figure 1: The RPE distribution in the nal dataset. This shows us how many datapoints we have in our dataset and how they are distributed.

4 Methods

In this section, we discuss the methodology. The di erent parts in our approach can be seen in Figure 2 and in this section we focus on the Data Mining part. We discuss the di erent approaches that we have taken. Moreover, as we are in a situation where the number of predictors is larger than the number of distinct data points, we speci cally explain how these risks of over tting are minimized.

Figure 2: An overview of the di erent steps of our data-driven approach to study the relationship between the selected Rating of Perceived Exertion and the training characteristics obtained from sensor data in wheelchair tennis. We started with receiving the datasets and information from the KNLTB. We merged the datasets and preprocessed them. After that, we used data mining tools in order to extract information and build a model. We ended with advising the coaches and athletes.

4.1 LASSO regression with cross validation

The rst technique we have applied is LASSO regression. As explained in Chapter 2, this technique contains a parameter called that determines the number of predictors in our model. To nd the value of this parameter and an estimate on how our model performs on unseen data, we combine LASSO regression with nested Cross Validation. First, we have to scale the dataset. Scaling the dataset was necessary, because otherwise the value would faster penalize features with an average higher than other features. To scale the data we set the lasso regression parameter called normalize to true.

We started with using leave-one-out cross validation (section 2.2.3) to nd the optimal value for and corresponding model. This model predicts the RPE for training sessions and compares them to the actual value. The di erence of these values will be added to the total error. The optimal value for is then determined by nding the model with the lowest error. LASSO regression also provides insight in which variables are the most important by giving each variable a coe cient. The further the coe cient is from 0 the higher the impact it has on the prediction. This can help us understand which variables impact a training session the most. The LASSO regression algorithm with Leave-one-out cross validation doesn't give us a realistic view of the accuracy of the model, because the model has already seen the test datapoints when calculating the R^2-score. This the reason that we moved on to nested cross validation.

Nested cross validation is explained in section 2.2.4 of this thesis. For this thesis, we use $k=1$ in order to feed the LASSO regression algorithm as much data as possible. Every outer fold has his own optimal value for the parameter and model. This model will use a datapoint, the model has not seen before, to make a prediction for the target variable. Eventually, we have 24 models with 24 values for , which are crucial for two reasons. These models have produced 24 predictions which will be compared to their actual values in order to measure the accuracy of these models. As a measuring tool for the accuracy we use the R^2-score. The R^2-score is de ned in section 2.1.5 of this thesis. Secondly, we want to nd out more about the stability within these models. The consistency of the values will help us with checking the stability of this algorithm.

4.2 Regression trees with nested cross validation

The second technique we have applied is regression trees. For this approach, there are many similarities with LASSO regression. First of all, the structure of the algorithm is the same. We use nested cross validation for both machine learning techniques. The goal is still to optimize a parameter to build the best possible model. We still use the R^2-score to measure the accuracy of the model. However, there are also some differences. First, we use the decision tree regressor package to implement the regression trees. Also, we now try to optimize a different parameter, the ccp_alpha. This parameter is used for pruning in order to prevent the regression tree from overfitting. More information on this parameter can be found in section 2.2.6. Also, we use a different technique for scaling the dataset, because we can't use this technique with regression trees. We now scale the dataset by using the following formula:

$$X = \frac{X_{voor} - X_{voor}:\mathrm{mean}()}{X_{voor}:\max() - X_{voor}:\min()} \tag{1}$$

where:

X = The normalized dataset
X_{voor} = The original dataset

4.3 Subgroup discovery

With the help of the generic subgroup discovery tool Cortana, we aim at finding informative and high quality subgroups from our dataset. The Subgroup Discovery algorithm that is utilized for our research uses beam search to look for subgroups in a large group of candidates. There are a few parameters we have to set in order for our algorithm to give the desired results.

For our quality measure we have chosen for Explained Variance, because it works in a similar way as the R^2 score used with the other machine learning techniques. We have also researched another quality measure. This quality measure is called absolute Z-score. The Z-score measures how may standard deviations the mean of the subgroup is away from the mean of the population. The higher the value of the Z-score, the bigger the difference between the population and the subgroup. The reason we also investigated this quality measure is because we wanted to investigate cases where the subgroups differ a lot from the mean and population. The target variable is set to the RPE. For this thesis, we don't have a lot of data, so the search depth (d) is set to two. This is sufficient, because of that after a certain point the quality of the subgroups does not increase when the number of conditions increases. For the beam width w, we have chosen the value 100, because when we increase it no extra high quality subgroups are found.

At the end, we validated our results by using a distribution of false discoveries. We have calculated the threshold for finding significant results by using swap-randomization on the target attribute.

5 Results

5.1 Introduction in the results

In this section, we will present the results that the machine learning techniques have provided. We will discuss the results from the LASSO regression model, the regression tree model and subgroup discovery, separately.

5.2 LASSO regression

In this part of the results section, we will discuss the ndings from the LASSO regression algorithm. We start with using leave-one-out cross validation to get our rst model and corresponding value for the parameter . As shown in gure 3, the model that makes the lowest error for predicting the Rating of Perceived Exertion per session has anvalue of 0.068.

Figure 3: A graph containing the error that the models make when predicting the target variable. The values of the models can be found on the x-axis and the error on the y-axis.

We also wanted to research what variables de ne the LASSO regression formula. There are three variables that have an e ect on the prediction of the target variable. These variables with their corresponding coe cient are shown in Table 3.

Variable	Coe cient
LiSpAvg	3.017358
% Rot sz 100+	4.208297
Freq Speed 3.5+	0.007255

Table 3: Variables that have an e ect on the prediction of the target variable with their corresponding coe cients. The average speed (LiSPAvg) has a smaller impact on the target variable than the percentage of the time that the rotation speed of the wheels of the wheelchair is higher than 100 degrees per second. The lowest attribute in the table stands for how frequent the speed goes over 3.5 metres per second.

We will now look at the average error this model makes when predicting a certain RPE value. The results from this machine learning algorithm can be found in gure 4. For example, this algorithm predicts 2.9 points high when the actual RPE value is 3.

Figure 4: The average error of the model as a function of the RPE values.

To get a full and accurate view of the LASSO regression algorithm and the results, we have to use nested cross validation. The reason for this is that the cross validation algorithm tests on a datapoint that it has already seen during training.

Nested cross validation will tell us more about the model's ability to predict the RPE for new data. That is the reason this part of the results will discuss the results from the LASSO regression algorithm when the optimal values for are determined by nested cross validation. Every time we predict the RPE for a test datapoint, we use the best possible model (with the best value for the parameter). This model is determined by the inner fold of the nested cross validation. The stability of the value of the parameter can be found in figure 5. On the y-axis we find the values for . The values are all between 0.0593 and 0.0907. This means the values are all in an interval of 0.0314 from each other.

Figure 5: The best values for the parameter per model are displayed in this figure.

When we use the predictions of our model for new data we can calculate the R^2 score:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{41.036934}{46.0} = 0.10789$$

So, The R^2-score of our model is approximately 0.108.

5.3 Regression trees

In this part of the results, we will discuss the results from the regression trees algorithm. This machine learning technique builds a tree in order to predict the Rating of Perceived Exertion (RPE). We will not place all the regression trees in this thesis, but we want to show you at least one. One of the trees can be found in figure 6. You walk through the tree answering to statements that are displayed at the top of each node. If the statement stated at the top of a box is true, the left arrow should be followed. You do this until you are at a leave and have found the predicted value.

Figure 6: This regression tree is used for predicting the Rating of Perceived Exertion in one of the outer folds. When you answer the questions in the tree, you will be able to make a prediction of the target variable. Whenever the statement on top of a box is true, you follow the left arrow. Otherwise you should follow the right arrow. The mse stands for the mean squared error. If the mean squared error is zero, the model does not make a mistake for this prediction for the training set. Overfitting will then be a probability. The samples show us how many datapoints there are left in that node. The values tell us the value for the RPE of the samples in that box.

It is now time to see how well these regression trees perform when predicting the target variable. In gure 7, you will nd the average error that the regression trees make for predicting the RPE values that are listed on the x-axis. For example this model predicts 1.5 points low when the actual RPE value is 9.

Figure 7: This gure shows us how accurate our model is on predicting a certain value for the target variable.

To calculate the accuracy of this model for new data and check the stability of the model, we use nested cross validation. In figure 8, the optimal values for the ccp are listed. These ccp values are used to predict the RPE for new data. On the y-axis we find the values for ccp. We can see here that the values of the ccp are all between 0 and 0.250. The values are consistent except for three outliers, which are lower.

Figure 8: The best values for the hyper-parameter per model are displayed in this figure.

When we use the predictions of our model top predict the RPE for new data we can calculate the R^2 score:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{37.72416667}{46.0} = 0.17991$$

So, The R^2-score of our model is approximately 0.180.

5.4 Subgroup discovery (SD)

in this section, the results will be presented that we gathered with the subgroup discovery program Cortana. We list the subgroups that were obtained with two di erent quality measures. These quality measures are Explained Variance and absolute Z-score. In table 4, the ve subgroups are presented with the highest quality according to the quality measures Explained Variance and absolute Z-score.

Quality measure: Explained Variance

Quality	Average	Discription of subgroup
0.528261	3.75	Left p_turns <= 351.0 AND Freq Speed< 0 >= 145.0
0.444851	4.2	Left p_turns <= 351.0 AND Freq Distance< 0 >= 158.0
0.444851	4.2	Left p_turns <= 351.0 AND Freq Speed 0.75<1.5 >=43.0
0.444851	7.8	Freq Acc 2.5+ >= 1.0 AND Freq Acc 0.5< 1 >= 61.0
0.417391	5.6	Freq Speed 3.5+<=22.0 AND distTotal <=8937.0

Quality measure: abs Z-score

Quality	Average	Discription of subgroup
3.250418	3.75	Left p_turns <= 351.0 AND Freq Speed< 0 >= 145.0
2.919202	3.666667	% Rot sz 50< 100 = '0,12' AND Freq Distance 10+ >= 73.0
2.919202	3.666667	Left p_turns <= 351.0 AND Freq Distance< 0 >= 187.0
2.919201	8.333333	Freq Speed 3.5+>= 15.0 AND Freq Distance 10+ >= 109.0
2.907263	4.2	Left p_turns <= 351.0 AND Freq Speed 0.75<1.5 >=43.0

Table 4: The ve subgroups with the highest quality are listed in these tables. The quality measure that is used for the rst table is Explained Variance and absolute Z-score is used for the second one. The quality, listed in the rst columns, tells us how the quality is according to the used quality measure. The average shows us what the average RPE value was for this subgroup. The subgroup itself is listed in the third column.

As you can see in the above tables, the subgroup with the highest quality is the one where there have been made less than or equal to 351 turns to the left and the amount of times the speed is below zero is higher or equal to 145. This subgroup has an average value for the target variable of 3.75.

The threshold and quality of the best subgroups, derived from two of the quality measures, can be found in Table 5. We nd that the quality of the best subgroup is larger than the threshold for the Absolute Z-score but not for the Explained Variance. This means the subgroup is statistically signi cant for the quality measure absolute Z-score, but not for Explained Variance.

Quality measure	Signi cance level	Threshold	Quality of best subgroup
Explained Variance	5%	0.58	0.53
Absolute Z-score	5%	3.2	3.25

Table 5: Subgroups have a quality value which is measured by quality measures. The quality measures give us subgroups of di erent qualities.To see if a certain subgroup is statistical signi cant, we compute the threshold. If the quality is higher than the threshold, the subgroup is signi cant. In this table, we nd that threshold and can compare it with the quality of the subgroup to nd out if the subgroup is signi cant.

6 Discussion

in this section, we try to interpretate the results we have gathered in our research.

This research is not generalizable for all the wheelchair tennis players around the world, because we have specifically researched the female athletes from the Netherlands. During our research we saw that the male wheelchair tennis athletes behaved in a different way, which makes it impossible to generalize this research for the entire wheelchair tennis community. It is hard to tell if it is even generalizable for the entire female wheelchair tennis community, because athletes from other countries can behave in a different way or train under other conditions.

We will now discuss the models and their respective accuracies and stabilities. The accuracy of the models was not as high as we had hoped. One possible reason is the low amount of datapoints that were fit for this thesis. If we have a closer look at the reason why the R^2-score is low, we find that the models have problems predicting the low (relatively easy session) and high (relatively though session) values of the RPE. Evidence for this can be found in figures 4 and 7. A possible explanation for this is that the model has the lowest training data for predicting these RPE values. If the machine learning technique can't train on data with these "extreme" values, it will also not be able to predict these values for new data.
To check the stability of the models in this thesis we look at the stability of the parameters of our models. Model instability causes additional variance, which we want to avoid. The stability of the parameters of both models are somewhat similar. The parameter values of the LASSO regression model lie close to each other and have no outliers. The parameter values of the regression trees algorithm are further from eachother and have 3 outliers. The stability of the LASSO regression model is a bit better, but it is hard to compare them because of the fact that the parameters are different. This is why find that both models are stable enough.

There are some differences between the machine learning techniques in this thesis. First of all, The accuracy from the Regression Trees algorithm is higher than the accuracy of the LASSO Regression algorithm. Another difference is that the features that have an important effect on predicting the RPE value are different for all the machine learning techniques.

To elaborate more on these features, we see in the research that a few training attributes were more important to calculate the RPE than others. For example the average speed was important, which suggests if the training consisted of a lot of fast moving it will be perceived as tough. This is in line with the research we found [18]. That research showed us that training load is dependent on the training volume and intensity which are both high when the average speed is high. Another thing that we found is that rotating the wheels with more than 100 degrees per second is also perceived as though. Research found that the movement velocity is an important workload factor in wheelchair tennis [23]. The movement velocity is generated by a high rotation velocity. These training features should be avoided if the athlete is tired or coming back from an injury. In our best subgroup, there have been made less than or equal to 351 turns to the left and the number of times the speed is below zero is higher or equal to 145. This subgroup has an average value for the target variable of 3.75. This subgroup stands for a training that is perceived as "easy". What we learn from this is that left turns can be though for the athlete. More people are right handed, which could be the reason for this. This subgroup was significant with the absolute Z-score measurement but not with the Explained Variance. This shows us that we can use this information but have to stay critical, because of the low amount of data.

With the results from this thesis we want to give the coaches advice for the training sessions of their athletes. The coaches could try to focus more on interval training, because the average speed was an important factor for predicting the RPE. This, so called repeated sprint ability is crucial for wheelchair tennis athletes [22]. Also, it would be helpful to train the physical strength of the arms more, because heavy rotating of the wheels gave the athletes a heavier training session on a physical level. Therefore, we recommend to include specific upper-body repeated power ability drills in the physical preparation [22]. Lastly, the turns to the left were important in combination with the amount of times the speed is below zero. Coaches can interpret this and choose to train the left arm more and add more reverse movement in though training sessions to switch things up.

For improvements of this research my first suggestion would be to do this research again when there is more data available. This will improve the accuracy significantly. Also, it would be a very interesting to have a more balanced dataset with more "extreme" values for the target variable. An interesting idea for further research would be to add the speed of the strokes of the players to the dataset in order to see in what way that effects the training load. It will also be interesting to conduct this research for male wheelchair tennis athletes. Further research can also involve personal models, because every person is different and perceives a training session in another way.

7 Conclusion

In this section, we will answer our research question:
How can we quantify the relationship between the selected Rating of Perceived Exertion and the training characteristics obtained from sensor data in wheelchair tennis?

We have demonstrated that the relationship between the Rating of Perceived Exertion and training characteristics in wheelchair tennis can be obtained by using LASSO regression, regression trees and Subgroup Discovery. We have measured the accuracy of the two machine learning techniques with the R^2-score. The two machine learning techniques we used were LASSO regression and regression trees. According to the R^2-score of the models, the regression tree model is more accurate than the LASSO regression model. For these two models we have also gathered their respective values for the parameters and ccp_ . With these parameters and their values it is possible to conclude that the models have a somewhat similar stability, but the stability of the LASSO regression model is better. This has to be taken lightly, because these parameters have different purposes. According to this thesis, it is possible to conclude that the regression tree model works best for this specific problem.

We will now look at some features that are important for this thesis. LASSO regression showed us that the Average Velocity and the percentage of the time the athlete spends in the rotation speed zone of higher than 100 degrees per second are the most important variables to quantify the relationship between the selected Rating of Perceived Exertion and the training characteristics. The average velocity shows us that athletes perceive a training as tough when the velocity is not dosed well. The rotation speed zone of higher than 100 degrees per second makes it clear that the training load depends for a large proportion on how hard the athlete makes the wheels rotate. There was one subgroup that was statistically significant for one of the quality measures. In this subgroup, there have been made less than or equal to 351 turns to the left and the number of times the speed is below zero is higher or equal to 145. This subgroup has an average value for the RPE of 3.75. If there have been made a low amount of left turns and a lot of movements backwards a training can be described as a training with a medium low amount of physical load.

Interesting possibilities for future research are to research the male athletes in order to compare the results and find out how the two groups differ. Also, it would be interesting to research foreign female wheelchair tennis players to compare them with wheelchair tennis athletes from the Netherlands.

References

[1] Naclerio, Fernando, Chapman, Mark, & Larumbe-Zabala, Eneko. (2015). Use of the Rate of Perceived Exertion Scales in Resistance Training: A Comment on Mayo, Iglesias-Soler, and Fernández-Del-Olmo (2014). Perceptual and Motor Skills, 121(2), 490-493.

[2] Moreira, Alexandre, Gomes, Rodrigo V, Capitani, Caroline D, Lopes, Charles R, Santos, Audrei R, & Aoki, Marcelo S. (2016). Training intensity distribution in young tennis players. International Journal of Sports Science & Coaching, 11(6), 880-886.

[3] Fahlman MM and Engels HJ. Mucosal IgA and URTI in American college football players: A year longitudinal study. Med Sci Sports Exerc 2005; 37: 374–380.

[4] Sánchez-Pay, A, Torres-Luque, G, & Sanz-Rivas, D. (2015). Match activity and physiological load in wheelchair tennis players: A pilot study. Spinal Cord, 54(3), 229-233.

[5] Ranstam, J, & Cook, J. A. (2018). LASSO regression. British Journal of Surgery, 105(10), 1348.

[6] Cady, F. (2017). The data science handbook.

[7] Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc. 1996;58:267–288.

[8] Fahrmeir, Ludwig, Kneib, Thomas, Lang, Stefan, & Marx, Brian. (2013). Regression (2013 ed.). Berlin, Heidelberg: Springer Berlin / Heidelberg.

[9] Knobbe, Arno, Orie, Jac, Hofman, Nico, Van der Burgh, Benjamin, & Cachucho, Ricardo. (2017). Sports analytics for professional speed skating. Data Mining and Knowledge Discovery, 31(6), 1872-1902.

[10] De Leeuw, Arie-Willem, Meerhoff, Laurentius A, & Knobbe, Arno. (2018). Effects of Pacing Properties on Performance in Long-Distance Running. Big Data, 6(4), 248-261.

[11] Meijer, Rosa J, & Goeman, Jelle J. (2013). Efficient approximate k -fold and leave-one-out cross-validation for ridge regression. Biometrical Journal, 55(2), 141-155.

[12] Kim, Deok Ryun, Ali, Mohammad, Sur, Dipika, Khatib, Ahmed, & Wierzba, Thomas F. (2012). Determining optimal neighborhood size for ecological studies using leave-one-out cross validation. International Journal of Health Geographics, 11(1), 10.

[13] Parvandeh, Saeid, Yeh, Hung-Wen, Paulus, Martin P, & McKinney, Brett A. (2020). Consensus features nested cross-validation. Bioinformatics, 36(10), 3093-3098.

[14] Samadi, Mehrshad, Jabbari, Ebrahim, & Azamathulla, H. Md. (2012). Assessment of M5 model tree and classification and regression trees for prediction of scour depth below free overfall spillways. Neural Computing & Applications, 24(2), 357-366.

[15] https://www.usta.com/en/home/stay-current/national/history-of-wheelchair-tennis.html

[16] https://scikit-learn.org/stable/auto_examples/tree/plot_cost_complexity_pruning.html

[17] Gomes, Rodrigo V, Moreira, Alexandre, Lodo, Leandro, Capitani, Caroline D, & Aoki, Marcelo S. (2015). Ecological Validity of Session RPE Method for Quantifying Internal Training Load in Tennis. International Journal of Sports Science & Coaching, 10(4), 729-737.

[18] Coutts, Aaron James, Gomes, Rodrigo Vitasovic, Viveiros, Luis, & Aoki, Marcelo Saldanha. (2010). Monitoring training loads in elite tennis. Revista Brasileira De Cineantropometria & Desempenho Humano, 12(3), 217-220.

[19] Lloyd Jones, Molly C, Morris, Martyn G, & Jakeman, John R. (2019). Effect of Work: Rest Ratio on Cycling Performance Following Sprint Interval Training: A Randomized Control Trial. Journal of Strength and Conditioning Research, 33(12), 3263-3268.

[20] McLaren, Shaun J, Macpherson, Tom W, Coutts, Aaron J, Hurst, Christopher, Spears, Iain R, & Weston, Matthew. (2017). The Relationships Between Internal and External Measures of Training Load and Intensity in Team Sports: A Meta-Analysis. Sports Medicine (Auckland), 48(3), 641-658.

[21] https://online.stat.psu.edu/stat508/lesson/11/11.8/11.8.2

[22] Sánchez-Pay, Alejandro, & Sanz-Rivas, David. (2020). Wheelchair tennis, from health to competitive analysis: A narrative review. Journal of Human Sport and Exercise, 16(2), Journal of human sport and exercise, 2020, Vol.16 (2).

[23] Filipcic, Ales Filipcic, Tjasa. (2009). Analysis of movement velocity and distance covered in wheelchair tennis. Kinesiologia Slovenica. 15. 25-32.