Contents

1 INTRODUCTION

The ability to be able to quickly accelerate the body from a resting position is highly regarded as an important performance metric in many sports. This explosive ability is related to the maximal power output of an individual. An assessment of maximal power output is common practice for trainers and coaches in general. Maximal power output can be evaluated commonly for the arms and legs. The lower limbs maximal output power specifically the legs can be measured using different exercises namely, sprint running, sprint pedalling on bike, and vertical jumps.

Jumping is a complex human movement motion that requires motor coordination of the upper and lower body segment. The propulsive action of the lower limbs during a jump motion is an indicator for maximal output power. As a result, many reliable and valid tools have been developed to measure the flight time, as well as the amount of force generated throughout vertical jumps.

As this approach has been popular amongst coaches and sport scientists different vertical jump forms for the estimation of maximal output power has been developed. The two most commonly used are the counter movement jump (CMJ) and squat jump (SJ). The CMJ involves an $< 90\,\text{deg}$ bend in the knees with motion of the hands upon take-off from ground. The squat jump limits the motion of the hand as they must be placed on the hips. The use of these two different jump types provide slightly different information on athletic performance. The SJ allows a deep bend in the knees and the motion strictly involves the legs capability of generating adequat power to create maximal take off height. The CMJ is an indicator of using both upper body and lower body synchronization capabilities to mutually create maximal power output. Generally, as the CMJ is a more flexible jump it results in higher flight times, compared to the squat jump.

In general, the public does not have access to athletic facilities and hence cannot evaluate their jump performance. Tools involved in evaluating jump performance could also be cost prohibitive, hence various studies [1,2] have attempted to predict jump height value based on different metrics such as height, weight, age, race. In this study we attempt to devise a set of metrics and questions that will be used to train models for predicting the flight time of a person. We train our models using a data set from students, and discuss the results in the following sections.

2 QUESTIONS SELECTED AND JUSTIFICATION

2.1 SELECTED QUESTIONS

Table 2.1 presents the list of the selected questions to be asked from the human subjects of this study. The questions are selected based on various biological and experimental studies which are referred to in the justification part (Section 2.1.1).

Table 2.1: The features and parameters that are suggested to be used in this study.

#	Question	Answer Format	Justification
1	What's your height?	Feet and inch (or, other units with unit name)	See Section 2.1.2
2	What's your weight?	Pounds (or, other units with unit name)	See Section 2.1.2
3	What's your gender?	Two options: "Male" or "Female"	See Section 2.1.3
4	How many minutes a day you exercise? (on average)	Minutes	See Section 2.1.4
5	Which categories best describes your race?	Multiple Choice: Asian, Black, White, etc.	See Section 2.1.5
6	What's your age?	Years	See Section 2.1.6
7	How fast do you run?	Scale 1-10 (10 being the fastest)	See Section 2.1.7
8	What is the maximum number of squats that you can do?	Number	See Section 2.1.9
9	Do you play any sports involving short sprints and jumps?	Yes or No	See Section 2.1.8
10	Do you care about your jump height?	Yes or No	See Section 2.1.7

This document should only be use for personal use, and not published or distributed.

2.1.1 Justifications

The vertical jump has been widely studied as a performance metric for assessing athletes. Researchers have examined various factors that potentially could contribute to the counter movement jump (CMJ) performance. Below outlines various metrics such as age, body composition, weight, height, etc which could help predict a recreational athletes performance on such tasks.

2.1.2 Height and Weight

Height and weight are general anthropometric parameters that have been widely used as a predictive variable. For instance, Aragon-Vergas et al. [1] explored the usage of various physiological parameters and selected the body weight and height as predictive variables. The study reported them as effective features to be used in combination of several other parameters for the vertical jump predictive analysis.

2.1.3 Gender

Individual's gender is highly correlated to how high they can jump vertically. According to Zeinal Abidin et al. [2], male subjects are reported to jump about 26% higher than female subjects, on average.

2.1.4 Exercise

In general, regular physical exercise and overall fitness plays a significant role in the athletic performance. Specifically, Genuario et al. [3] discovered that there is a moderate relationship with the correlation coefficient of about 0.5 between the muscular strength and vertical jump performance.

2.1.5 Race

Empirically, race is believed to be a significant factor in many types of athletic activities. For example, it is observed that Africans are, on average, more successful the in track and field sports. There are also many studies in the literature that confirms such a relationship. Guth et al. [4] conducted a study and analyzed the effect of certain genes on athletic performance in vertical jump, running. It is further concluded that a few genes can be associated with favorable athletic profile.

2.1.6 Age

As an individual ages, the muscles become weaker and weaker. In specific, after the age of 30, the density of muscles starts to reduce and at the same time the amount of body fat will increase . According to Baechle [5], as the age increases the overall muscle strength reduces which can directly affect the vertical jump ability.

This document should only be use for personal use, and not published or distributed.

2.1.7 Self-Report

Self-report is one of the most popular methods used to measure physical activity. While the reported values can be affected by various psychological factors, it is still highly correlated with the actual quantities of interest [6]. Here, we ask several questions regarding the expectation of individuals about their athletic performance. In addition, to reduce the influence of psychological factors, subjects are asked about a few questions that are indirectly related to the vertical jump performance such as their physical and mental conditions before the jump and so on.

2.1.8 Sprint Time

Vertical jump is a typical assessment metric for display of explosive strength and power. The vertical jump has a high correlation to the power generated by the leg extensors, thus similar assessment of sprint and fast leg movement exercises correlate to these metrics. In [7] a correlation has been found between the jump squat performance of an athlete, when their weight has been taken into account. Also, in [8] a strong correlation between 5 m sprint times and strength metrics for counter movement jumps has been found. [9]

2.1.9 Polymeric Training

As performance in vertical jump is highly correlated to performance in sport where leg movement is involved greatly[1], many different training types have been studied to enhance vertical jump performance. A common training employed is plyometric training (PT), widely suggested to improve countermovement jumps (CMJ)[10]. PT involves high-intensity eccentric and rapid powerful concentric contractions, using body weight. The performance on squats, involves power and strength in legs, and is correlated to vertical jump performance of an athlete[3].

2.1.10 Body Composition

Body composition on vertical jump performance has shown a negative correlation. In [11, 12, 6] a study has been done on high school athletes which showed students with higher body fat, had a dramatic lower vertical jump result as body fat increased above 10% in males and 19% in females. Though a clear relationship could not be shown for body fats below 10%. Body fat composition has shown to be a strong predictor for performance of recreational male athletes in the vertical jump test in a regression model [3]. To assess body fat composition one can ask about the definition of the individuals abdominals[4].

3 Proposed Method

In this section the proposed method for the prediction of the flight time based on the provided features will be explained. First, the background concepts that are used directly in the our approach are discussed in Section 3.1. Afterward, in Section 3.2 we will discuss the experimental setup under which the prediction take place during the train and test phases. Section 3.3 elaborates on the details of the final implementation of the method proposed in this paper.

3.1 Background

3.1.1 Ridge Regression

Linear regression is an approach employed to model the relationship between a scalar function y with input feature vector X. The simple linear regression model is formulated as below:

$$h(x) = W^T X + b. \tag{3.1}$$

Where, W is the weight vector and b is the bias parameter.

Usually, the model parameters (i.e., w and b) are adjusted to fit the data. The weights can be determined using various cost functions. A popular choice is the ordinary least squares cost function:

$$Cost(W,b) = \frac{1}{N}\sum_{n}^{N}(W^T X_n + b - y_n)^2. \tag{3.2}$$

The above is minimized subject to the choice of and results in a hypothesis function that best matches the training set. Finally, at the prediction phase, this model is then used to determine values for the new input vectors of interest.

In a special branch of linear regression models called ridge regression, the models variance is reduced by simply adding a regularization term to the cost function:

$$Cost(W,b) = \frac{1}{N}\sum_{n}^{N}(W^T X_n + b - y_n)^2 + \lambda \sum_{d} w_d^2. \tag{3.3}$$

Where λ is a hyper-parameter that controls the contribution of the regularization term in the total cost.

Linear regression method are one of the most popular regression methods in the literature. The main reason behind this fact is that a least squared linear regression can be fitted using few training samples. In addition, there are analytical solutions for the optimum model parameters that facilitates training these models. However, linear regression assumes independence in the input variables. Also, in problems that the target is not linearly related to the features it works poorly. Whereas non-linear regression models can easily handle many such cases. Though linear regression has been well studied and is relatively easy to deploy in many different systems, hence its popularity.

3.1.2 K Nearest Neighbors

It is a supervised classification and regression algorithm. This algorithm is intuitive and easy to understand, thus it is used in many applications. This method does not require training,

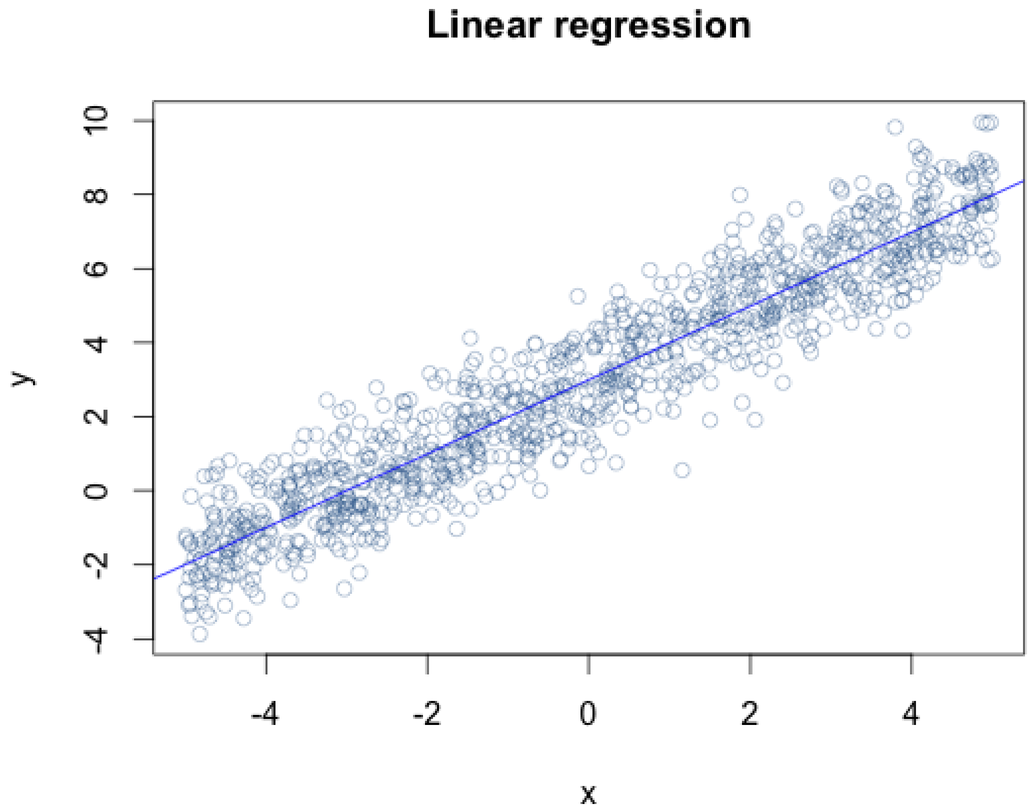

Figure 3.1: An example of a linear regression fitting. Image courtesy of r-bloggers.com.

and can be readily employed using labelled data. It is best to choose labelled data that result in boundaries within the data which adequately represents the potential classes of data the system will be subject to.

In K-nearest neighbor the objective is find the K labeled points y_i, $i = 1, 2, ...k$ which are closest to the input point x using a distance measure. The input is then assigned the mean of the target values corresponding to these points. In the case of K=1, the input is assigned the label that of the closest neighbor. For data inputs that are continuous, the most employed measure is the Euclidean distance:

$$D(x, y) = ||X - Y||_2^2 = \sum_k (x_k - y_k)^2 \tag{3.4}$$

Other popular distance functions employed in KNN algorithms are the Manhattan distance, and the Minkowski distance which are out of scope of this study.

Some notes to consider when using training sets, is with features that vary greatly in their units. The larger units will dominate the distance measures, resulting in a bias. This can be resolved by standardizing the feature space, to avoid larger features dominating the calculations although this could ensue its own biases. A common standardization scheme used is adjusting the mean value of each feature to zero while adjusting the variance value to one.

Another important step in KNN algorithms is to determine the value of K. A popular method employed for this step is k-fold cross validation in which the training data is split into k partitions. One partition is used for validation, and the model is trained on the rest k-1 partitions. This increases run-time by a factor of K. Figure 3.2 demonstrates the effect of different Ks on the final KNN decision boundary.

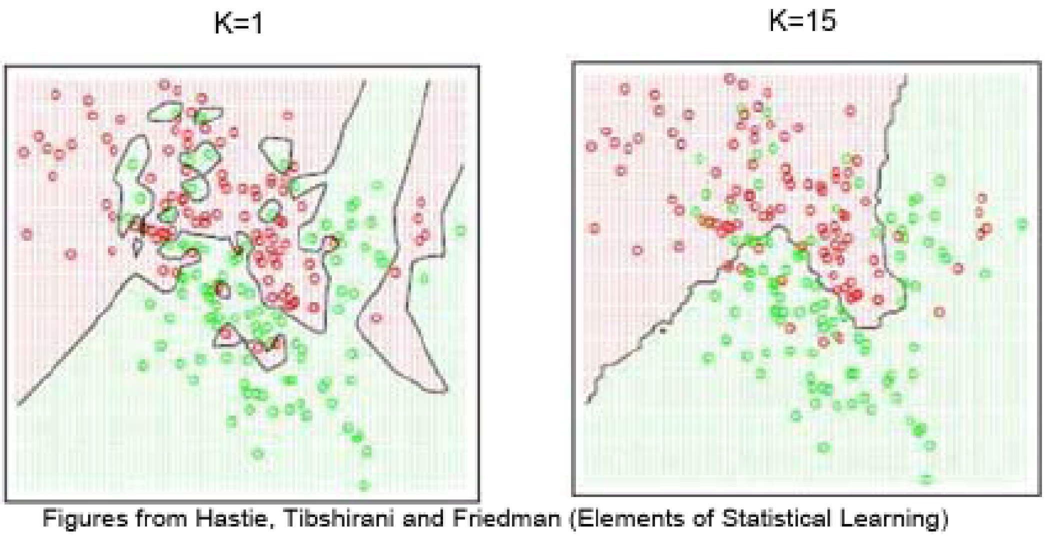

Figure 3.2: An example of a KNN model. The dicision boundaries for K=1 and K=1 are compared against each other. As it can be seen form this figure, the decision boundary for the choice of K=15 is much smoother than the K=1 case.

3.1.3 Random Forest

One of the most well-known learning algorithms within the ensemble learning framework is random forest. As it is evident from the name, in a random forest the prediction is made by the aggregation of the predictions from several decision trees. In other words, a forest of predictors is created from several decision trees. In the random forest setting, commonly, each tree is trained using a method called bootstrap learning. In bootstrap learning a learner is trained on a random subset of training dataset instead of the whole available dataset. Specifically, if the available training dataset is denoted by D, bootstrap aggregating suggests using n new datasets denoted by D_i that are n different random subsamples of D. Accordingly, using this method, each tree within the random forest ensemble is different from other trees as it is trained on a different batch of training samples. For regression problems, the average value of

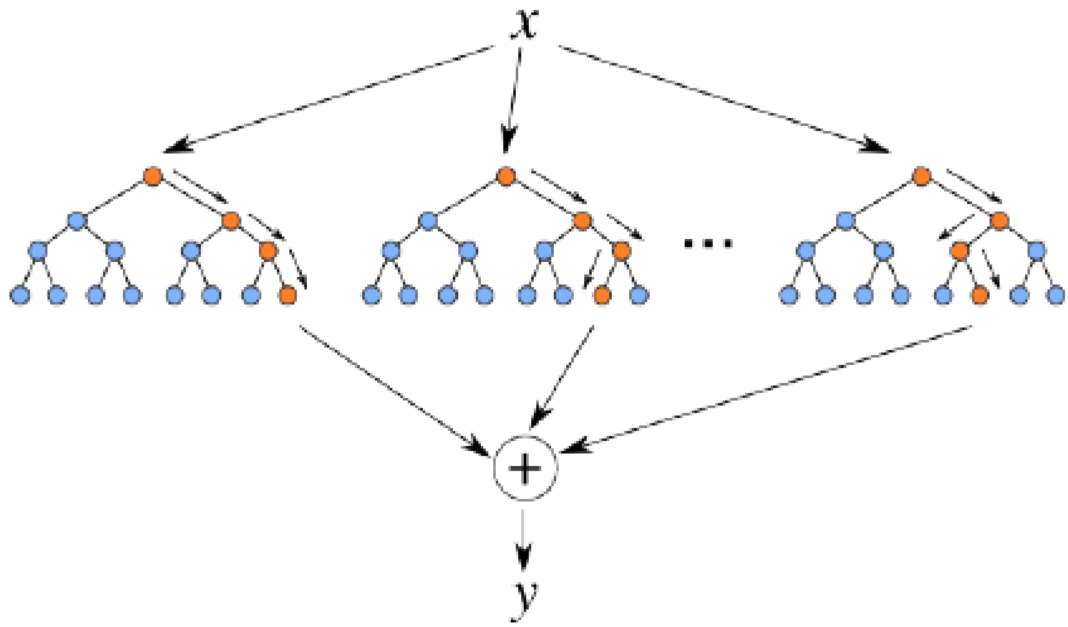

Figure 3.3: An illustration of random forest regression algorithm. (Source: kazoo04. hatenablog. com)

predictions of all trees inside a random forest is calculated and returned as the final random forest prediction and for classification problems the final prediction is defined by a voting between tree predictions:

$$Pred_{regression} = \frac{1}{B}\sum f_b(x') \tag{3.5}$$

$$Pred_{classification} = \frac{1}{B} Mod(f_1, \ldots, f_B) \tag{3.6}$$

In practice, compared to strong learners, the diversity within a random forests plays a significant role in the reduction of the model variance and bias [13, 14]. Another advantage of using random forest is that by the trained model can be analyzed to infer feature importance [15, 16].

Recently, random forests become one of the most attractive learning algorithms within the healthcare research community [17, 18, 19, 20, 21]. Another reason which makes them favorable in many healthcare scenarios is that while random forests can achieve performances comparable to SVMs, random forests models are still easy to interpret.

3.1.4 Recursive Feature Elimination

Recursive feature elimination is one of the simplest but yet effective methods for feature selection. In a nutshell, the algorithms starts with training a model using all of the available

Algorithm 1: Recursive feature elimination

```
Tune/train the model on the training set using all predictors
Calculate model performance
Calculate variable importance or rankings
for Each subset size S_i, i = 1...S do
    Keep the S_i most important variables
    [Optional] Pre–process the data
    Tune/train the model on the training set using S_i predictors
    Calculate model performance
    [Optional] Recalculate the rankings for each predictor
end
Calculate the performance profile over the S_i
Determine the appropriate number of predictors
Use the model corresponding to the optimal S_i
```

Figure 3.4: Steps of the recursive feature selection algorithm. (Source: topepo.github.io / caret/ recursive-feature-elimination)

features. Afterwards, in each step we select a single feature to be eliminated form the current set of features and measure the effect of doing so on the performance. Usually, this process is repeated until the accuracy of the trained predictor start decreasing. Figure 3.4 demonstrates the detailed steps of the recursive feature selection algorithm.

3.2 Experimental setup

In this study, an electronic jump mat is used to measure the time of flight. Each subject is asked to jump for two times. First, a training jump so the subject becomes familiar with the correct jump style. The second time; however, is recorded as the time flight corresponding to each subject.

3.2.1 Questions Provided

Table 3.1 presents the final features that are used in the rest of this paper. In summary, the final list of features asked from each subject asks for height, weight, age, gender, race, favorite color, and amount of physical activity. A complete discussion of these features and their predictive power is discussed in Section 4.

It is also worth mentioning that the questionnaire is sent to the subject a few days before the actual jump tests, for the train dataset. For the test data collection; however, the time interval between the questionnaire was about a few weeks.

Table 3.1: The final features and parameters that are provided to be used in this study.

#	Question	Comment
1	What is your age?	Any non-negative integer value (ex. 21, 37)
2	How often do you exercise?	1. Everyday; 2. More than once a week, 3. Once a week, 4. Once a moth, 5. None
3	I consider myself a competitive person	1. Agree; 2. Neither agree or disagree; 3. Disagree; 4. None of the above
4	What is your height?	In inches
5	What is your weight?	In pounds
6	What gender do you identify with?	1. Female; 2. Male; 3. None of the above
7	Are you suffering or had an injury?	1. Yes; 2. No; 3. None of the above
8	What is your favorite color?	1. Red; 2. Blue; 3. Green; 4. Black; 5. White, 6. Yellow, 7. Other
9	How many hours do you sleep at night?	Average number, in hours
10	Race? 1. White 2. Black, African Am. 3. American Indian or Alaska Native	1. White 2. Black, African Am. 3. American Indian or Alaska Native, etc.

This document should only be use for personal use, and not published or distributed.

3.2.2 Final Evaluation

During the evaluation phase, the final models which are trained on the train dataset are used without any modification. In specific, a total of 11 subjects were selected as test subject and predicted flight time is compared with the ground truth measured values. A complete discussion of the data gathered from the test phase is presented in Section 4.

3.3 Final Implementation

In the beginning of the analysis, we have converted the categorical feature to numerical features. The resulting numerical feature vector is then statistically normalized such that the mean of each feature is adjusted to zero and its variance is adjusted to one.

For the implementation of a flight time predictor based on the provided features we tried different learning models such as linear regression, ridge regression, random forest, and K nearest neighbor algorithms.

It is noteworthy that based on the fact that there are 10 features available, while the total number of training samples is about as twice. Therefore, as it is expected, overfitting many of these learning algorithms. In order to overcome this issue we have devised a two step approach. First, we have removed a subset of features that are perceived to be insignificant to the final predictor such as favorite color, age, etc. (see Section 4 for a complete list of the eliminated features and the logic behind each one). In the second feature selection phase, we have used an automated feature selection algorithm in the machine learning literature. Specifically, we have used the recursive feature elimination algorithm (See Section 3.1).

Apart from the two phase feature selection, we have also considered the highest and lowest two flight times for each gender as outlier data and removed them from further analysis. Figure 3.5 shows the scatter plot of the flight time versus gender types. In this figure the outlier samples are indicated with blue color points, while other samples are marked with orange color.
We have used simple linear regression models as a simple and easy to train and analyze models. In order to train and test the model on the provided small dataset, we have employed a leave-one-out train and evaluation strategy. In other words, we have trained multiple models by selecting a single sample as evaluation sample and other samples as train data. Finally, the reported accuracy is measured as the average of all accuracies on the evaluation samples.

In addition to the simple linear regression we have also tried regularized linear regression and the combination of the with the recursive feature elimination algorithm. A detailed discussion of performance of each setup is discussed in 4.

Apart from the linear regression family, we have trained random forest regression models for the flight time prediction problem using the similar leave-one-out fashion. The trained ensembles consists of about 100 regression trees each trained on bootstrapped train datasets. Also, we have selected the mean absolute error (MAE) as the performance measure during the

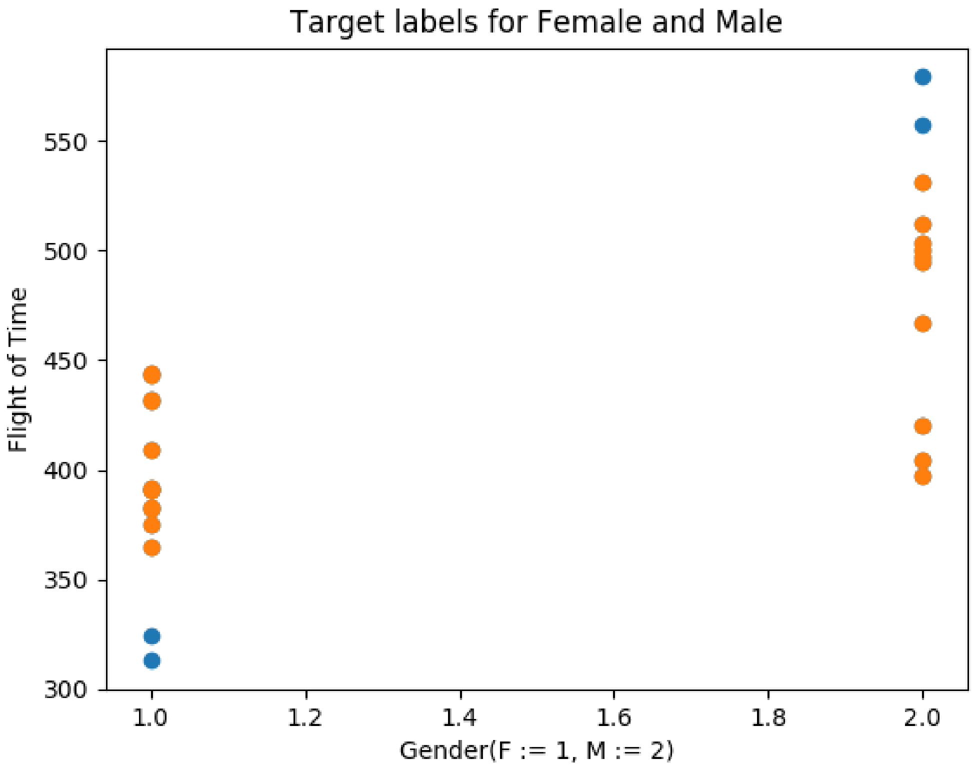

Figure 3.5: The scatter plot of the flight time versus gender types.

tree training. The main reason behind this is that the performance measure that is used in the final competition was defined to be sum of absolute errors.

The final family of learning algorithms that we have used was K nearest neighbor regression. In addition to the mentioned leave-one-out strategy we have used a cross validation loop inside the leave-one-out loop to select the model hyper-parameter K. At last, The K equal to 5 is selected as the optimal number of neighbors to be considered.

After comparing the prediction accuracy of the introduced methods, it turned out that while some of them are creating more accurate results, the prediction accuracy of them is not much significantly different. Accordingly, in order to create a more powerful model we have used the average predicted values of all models (except simple linear regression without recursive feature elimination) as the final prediction. This approach is widely used in machine learning competitions as a way to increase the stability of the final predictions. The main logic behind such a statement is that when outputs if several predictors are averaged together then an unstable output from one of them can only change the final prediction marginally. The final prediction ensemble is consists of about 2700 different predictors each trained using different

This document should only be use for personal use, and not published or distributed.

learning methods and different subsets of train dataset.

4 RESULTS AND DISCUSSION

Our model prediction had a sum absolute error of 346 ms for 11 test cases. On average the error for each test point is $\frac{346}{11} \approx 34$.

The jump height can be calculated using $h = \frac{1}{8}gt^2$. Given our error this translates to $\frac{1}{8}g(x \pm .034)$. Using a upper bound, and lower bound on flight time of 750 ms (pro track athlete) and 100 ms (grandpa), the error can range between $[0.7, 6.5]$ cm. This is a very encouraging result. In the following sections, we will discuss our predictions made and how our different choices in model selection led us to such prediction error.

4.1 COMPARISON BETWEEN THE SELECTED AND PROVIDED FEATURES

The majority questions are almost similar between both the selected and the provided question sets. These includes questions about the age, gender, race, weight, height, amount of exercise, and competitiveness. However, there are a total of three questions that are different. Specifically, we have selected three different questions, based on our literature survey that are considered to be measures of physical readiness. These questions are:

- How fast do you run?
- What is the maximum number of squats that you can do?
- Do you play any sports involving short sprints and jumps?

On the other hand, the final provided questions are designed to get different information about subjects background and preferences. It is noteworthy to mention that while we have considered asking these questions in our suggested questions list, we have finally removed them form this list. The list of these three questions and the reason behind the fact that these are not good questions to ask are summarized in the following.

- Are you suffering or had an injury?
  While this feature makes sense, but based on the population of study which are grad computer science students we do not expect to see many variance across this feature. In fact, the final train dataset confirms this hypothesis. Only one subject had an injury on ankle.
- What is your favorite color?
  It is not really expected to see any correlation between the color and flight time. However, by the feature importance analysis on the final dataset, it turned out that there is a significant correlation between color and the target. However, after further analysis we found that the color feature is correlated with gender and the gender is highly correlated with the target. So while there is a correlation between the favorite color and the flight time, it is not a direct genuine correlation.

- How many hours do you sleep at night?
  This feature is not really considered to be important and the feature importance analysis confirmed our hypothesis.

4.2 Feature Rankings

In our final implementation we have reduced the dimensionality of our input vectors by selecting the most important ones using an initial ridge regression model and throwing out the least important ones (please refer to Section 3.1 for a more formal explanation). On this basis, the following features were selected:

Table 4.1: The features used for training the models.

features	exercise	competitive	height	weight	gender	race

Initially we hypothesized that in this set of attributes the gender, exercise, height would be ranked as top three. To test our hypothesis our models were trained using these feature vectors, and the weights of the different attributes were ranked accordingly. The ranking of these attributes for each algorithm is summarized in the table below.

Table 4.2: Attribute ranking of different models.

features	exercise	competitive	height	weight	gender	race
Ridge	6	5	3	4	2	1
KNN	6	4	5	3	2	1
Random Forest	6	4	5	3	2	1

As expected, race and gender consistently ranked as very important. There is a discrepancy between the importance of weight and height, both seem to be either ranked 3rd or 4/5th. Although both alone is not a strong indicator of fitness and flight time together they could determine the fitness of the person to certain degree. Hence, these feature rankings show the importance of gender, race, weight, height in determining the jump time of a person also cited to be important predictive factors.

4.3 Predictor Performance

Table 4.3 presents a comparison between the prediction accuracy of each predictor in terms of MAE. As it can be deduced from this table, the final prediction accuracy of all these methods is almost the same with only a small marginal difference. It justifies our suggested model combination approach. Figure 4.1 and Figure 4.2 showing the prediction-target regression plots for the random forest and K nearest neighbor regression algorithms, respectively. From the comparison of these plots it can be inferred that while the r value of prediction-target plots is almost similar for the both, the random forest regression is less biased compared to its K nearest neighbor counterpart. In other words, the random forest predictions are covering the

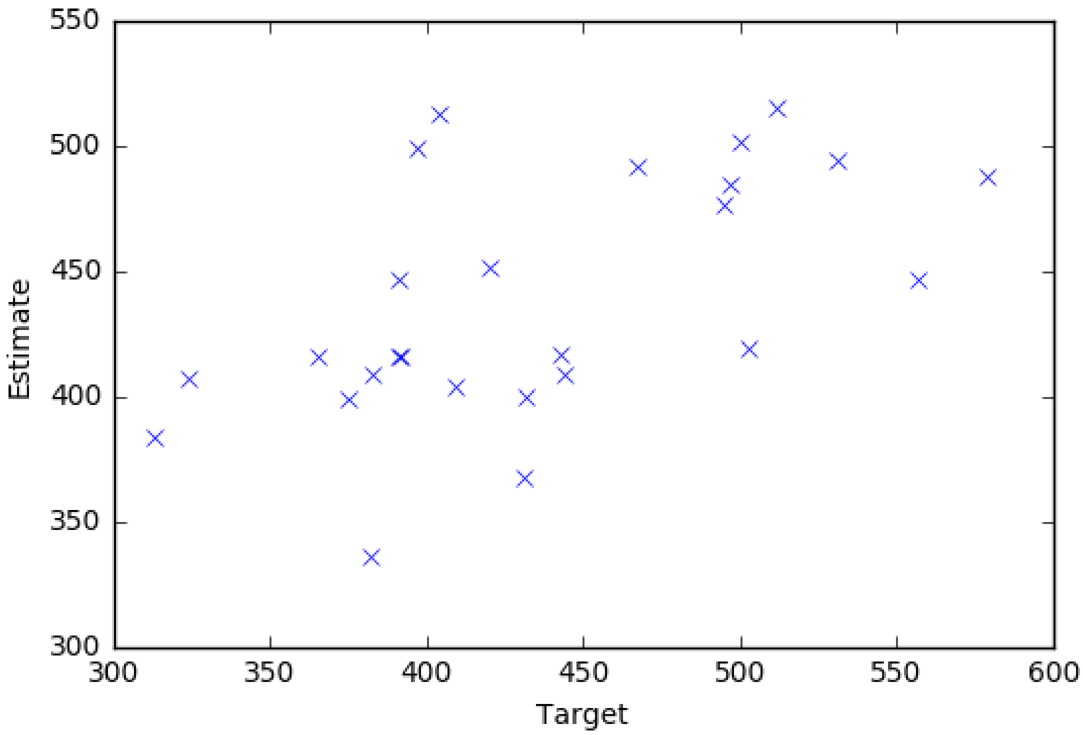

Figure 4.1: The prediction-target regression plot of the random forest regression model.

whole range from about 340 to 575. The K nearest neighbor method; however, predicts either values around 400 ms or values around 500 ms, and hence is more biased toward extremes.

Figure 4.3 shows the histogram of prediction errors using the final prediction ensemble and test phase samples. As it can be seen form this figure, the error values are always less than ±40 ms across all test subjects. It further justifies our reasoning about the conservativeness of our final predictor model. Apart from this, it can also seen from this figure that the error bias is about zero which means that our predictor model was fitted reasonably.

Table 4.3: Predictor Accuracy

Algorithm	Accuracy (ms)
Ridge	38.79
KNN	37.92
Random Forest	39.95

4.4 Per Subject Analysis

In this part, we discuss our prediction for each subject in this section. Our own qualitative analysis on the correcness of these predictions is explained briefly on a case by case basis.

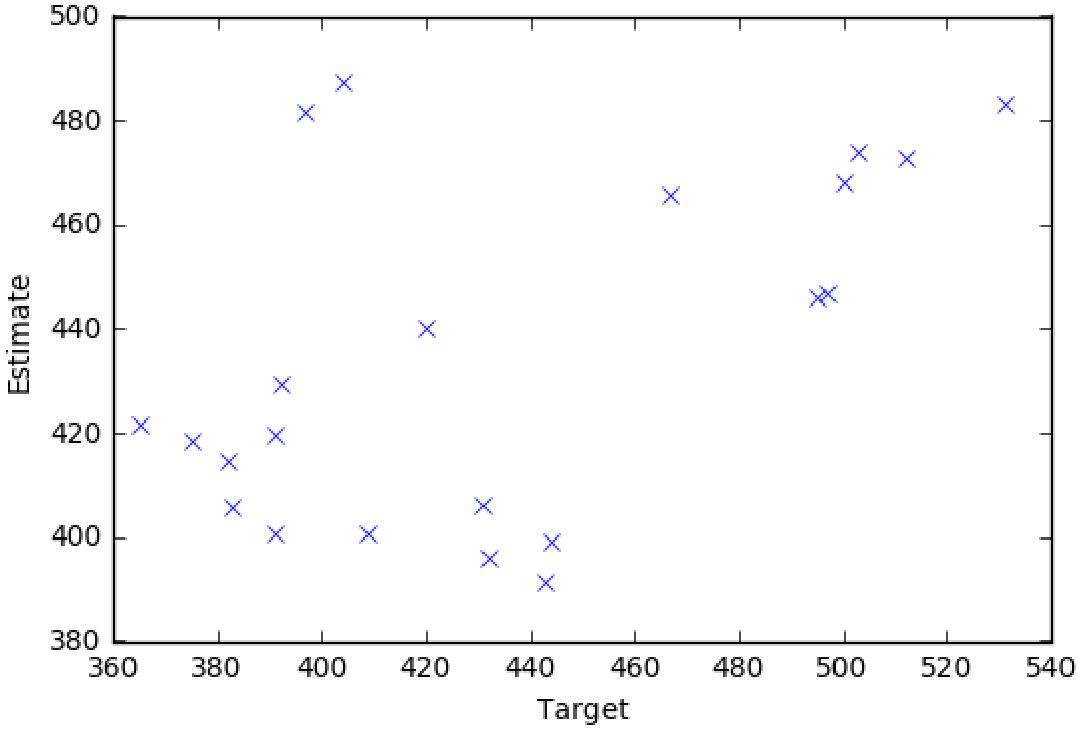

Figure 4.2: The prediction-target regression plot of the K nearest neighbors regression (K=5) .

TEST CASE 1

This subject is a white female. As this subject classified as a female, a prediction near the average of 380 ms is reasonable. The average of our predictors, predicted a value of 392 ms. The target value for this subject was 434 ms. This could be due to her fitness level, and our model did not rank this highly as discussed the feature selection. Further, she is tall and low weight which is correlated to higher jump height. On this subject we accumulated an absolute error of $434 - 392 = 42$ ms. This error results in 4.2 cm jump height difference.

TEST CASE 2

This subject is an Asian Indian male. He also indicated that he does not exercise and hence his fitness level is low. We would expect him to jump a below average for males. We predicted a value of 442 ms for this test case. The target value was 398 ms. We accumulated an error of $442 - 398 = 44$ ms. This error results in 4.5 cm jump height difference.

TEST CASE 3

Similar to test case 1, but is an Indian Asian. The gender directly translates to the different average predictions. Our model predicted a value of 391 ms. This makes sense as as she is not

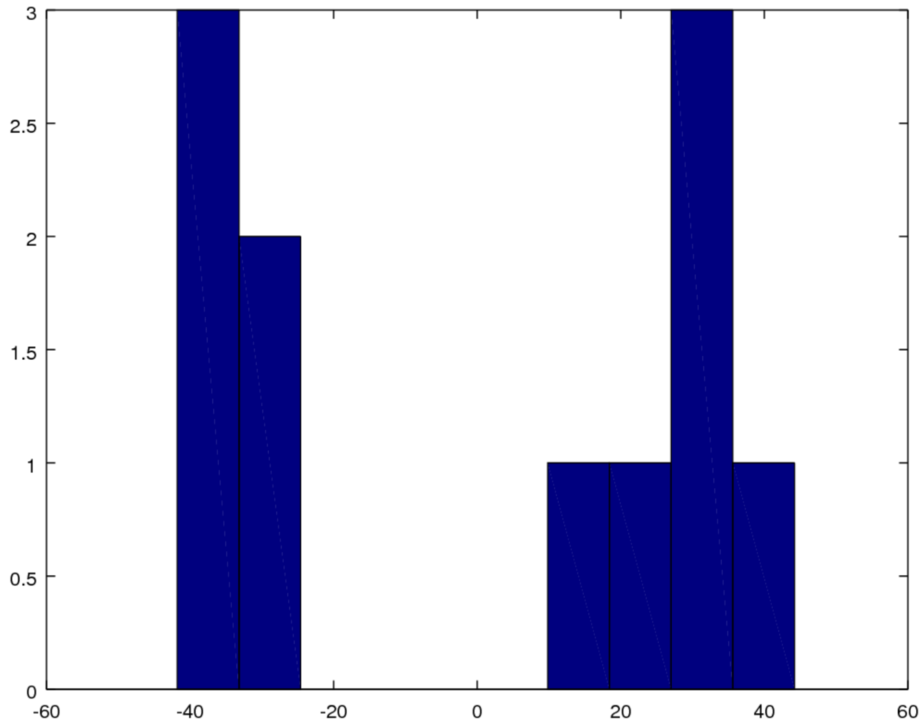

Figure 4.3: The histogram of prediction errors using the final prediction ensemble and test phase samples.

as physically active as the other high jumpers in female category. Also, she is short and over weight. The target value is 356 ms. It seems our prediction is high once again, this could be due to dropping the 2 lower values that were outliers for the females, hence on average we will predict a higher value. This error results in $391 - 356 = 35$ ms and a jump height difference of 3.2 cm.

Test Case 4

This subject is very similar to test case 1, though she identified as an Asian Indian. Although, she has reported similar frame (height, weight) as test case 1 as well as her fitness levels are comparable she had a lower flight time of 360 ms. This has difference of 74 ms. It could be attributed to the fact that she is Indian, and not white. But test case 3 was also an Indian with a much larger frame, and lower fitness levels but nearly jumped the same with a 4 ms difference. We hypothesize that she could have potentially not given her maximal effort on the jump performance.

Test Case 5

This test case is a white male, and is relatively skinny weight (height = 67, weight = 120lb). Our models prediction for this test case was 439 ms which is below average. Our model underestimated the flight time of the person by 36 ms, making a 4 cm jump height error. We attribute this underestimation to removing the top two outliers from the data set prior to training the models. If we were to re-train our model with the original data set, our prediction would increase to 463 ms resulting in a much lower error.

Test Case 6

This test case is an Asian Indian female with similar frame as test case 3. She considers herself competitive and trains more than once a week. Our model predicted 410 ms, and her actual jump value was 392 ms. Our model over estimated in this case by 18 ms, which translates to 1.8 cm.

Test Case 7

This test case is an Asian Indian female with a large frame 26.6 BMI, overweight. She does not exercise often and also does not consider herself a very competitive person. Our model overestimated (395 ms) for this test subject by 29 ms. This results in an

Test Case 8

Our prediction for test case 8 is 475 ms, which is again underestimating the jump height by 36 ms. This test case was a Chinese male athlete with reasonable exercise aptitude. Although, this is a relatively small prediction error we attribute this underestimation to the fact that we removed the two top outliers in the data set.

Test Case 9

Similarly for this test case our model underestimated the flight time. Our prediction was 403 ms, and the target value was 431 ms. The subject is female and exercise about once a week. The subject was expected to jump as much as the average female subjects. However, it turned out that she jumped a little higher. One possible explanation could be the weight of the subject which is about 106 pounds and potentially makes jumping easier for her.

Test Case 10

Our model predicted 485 ms, and the target value was 516 ms. This is a male subject with normal height and weight. Moreover, he exercises once a week which makes him like a typical male in our study. However, the predicted value is about 30 ms lower than the actual jump. Authors believe that based on the provided features it is really hard to distinguish this subject from a typical male subject in the dataset.

Summary of Case 1 to 10

In general our final model underestimated the flight time for male athletes, and overestimated female athletes. This is evident from the fact that we tried to fit a better model by removing outliers that seemed to not represent the crowd. It turned out that this resulted in a slight over fit to the training data. Though we consistently see that our average prediction error is within ≈ 30 ms of the actual value, and this results in $[1.2 - 6]$ cm jump height error prediction. This indicates that it is possible to predict reasonably well flight times of individuals based on their gender, race, height, and weight.

4.5 Comparison with Other Works

Figure 4.4 presents a comparison between the results of the proposed approach and the results of other groups for each test subject. From this figure, the performance of different groups can be classified into three different categories that are results with sum of absolute error (SAE) of around 200 ms, 300 ms, and 500 ms. The proposed method in this work, with the SAE of 346 ms, belongs to the second category. Apart from this, from the presentation that the top three groups made during the lecture, it can be concluded that using simple linear models on a manually selected subset of features results in the superior performance (hypothetically, on the provided test dataset). The main reason behind this could be attributed to the fact that compared to models with more complexity, simple linear regression models can be fitted more appropriately on the provided small training dataset and hence are able to make stronger predictions.

NAME of team members	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	SAE
	434	398	356	360	475	392	366	386	511	431	516	
Shangzhao Qiu, Yan Yan	418	435	365	367	497	367	363	414	492	416	497	200
Huey-ru Tsai and Yue Xin	412	432	358	348	510	367	358	426	501	417	514	204
Angela Chen , Wei-Ting Chen	451	421	361	322	459	392	363	415	521	379	455	253
Ning Xin, Aozhu Chen	442	425	378	350	450	413	367	416	523	387	459	257
Kaicheng Wang & Yubin Xiao	409	507	395	389	469	402	378	380	513	393	510	282
Shirya Sasank, Swathi	440	451	344	344	281	342	407	347	354	355	496	667
Dong Joon Kim & Migyeong Gwak	392	455	376	401	468	389	389	384	493	394	462	303
Yichao Zhou, Manni Chen	375	394	410	395	490	340	387	418	500	431	494	305
Atishay Aggarwal; Anshita Mehrotra	432	421	405	414	491	422	392	395	490	400	461	316
Anahita Hosseini, Xi Han	398	451	401	389	439	403	398	407	483	408	488	342
Aanchal Dalmia, Nikita Tulpule	421	463	383	365	410	398	372	421	507	402	595	335
S. Kauchuee, S. Darabi	392	442	391	388	439	410	395	405	475	403	485	346
Xirui Zhu, Jianing Liu	376	571	392	340	457	394	392	429	512	431	564	426
Wadekar, Gayathri	339	423	357	377	375	353	377	350	387	483	386	629
Jason La, Wesley Minner	372	488	412	401	435	415	415	434	513	423	514	421
Isha Verma; Shuktika Jain	369	422	409	429	453	370	445	448	563	434	513	455
Orpaz Goldstein, Timothy Portfolio	350	409	372	353	507	332	381	434	546	419	653	456
Ameya Kabre, Shikhar Malhotra	406	404	407	359	395	376	455	451	531	399	433	472

Figure 4.4: The histogram of prediction errors using the final prediction ensemble and test phase samples.

This document should only be use for personal use, and not published or distributed.

5 Conclusion

In this paper we have presented a flight time prediction approach as well as the analysis and discussion of its performance. In summary, the project consisted of three phases. In the first phase, 10 questions were selected which are considered to be highly correlated with flight time among individuals. In the second phase, a set of 10 questions were asked from about 24 subjects and their flight time were measured using an electronic jump mat. At this phase, a prediction model is created using the collected training data which is basically an ensemble of about 2700 different learners. At last, in the third phase, about 10 new subjects were asked to fill-up the same questionnaire and do the same physical activity.

The final prediction model which is proposed in this work is able to predict the individuals time of flight reliably. Specifically, we have achieved the mean absolute error (MAE) of about 38 ms on the train data using a leave-one-out strategy. Also, regarding the accuracy of the test data the MAE and sum of absolute error of predictions were 31 ms and 346 ms respectively.

Regarding the points of improvement, it seems that while in many cases combining the prediction from and ensemble of predictors makes the final prediction more stable, at the same time it inclines the predictions towards more conservative predictions. Accordingly, the final predictions are slightly biased toward either 400 ms or 480 ms. From the statistical point of view, so doing is sensible because the number of train data where relatively small compared to the number of features. Therefore, we expect that our prediction model can potentially perform much better given more training data provided. It is also worth mentioning that in the train phase, we have considered the highest and lowest flight times for each gender as outliers and excluded them from the further analysis. While doing so looked reasonable considering the data distribution of the train dataset, in the test phase it turned out that we have test subjects with values similar to what we previously considered as outliers.

References

[1] L. F. Aragón-Vargas and M. Melissa Gross, "Kinesiological factors in vertical jump performance: Differences among individuals," *Journal of Applied Biomechanics*, vol. 13, no. 1, pp. 24–44, 1997.

[2] N. Z. Abidin and M. B. Adam, "Prediction of vertical jump height from anthropometric factors in male and female martial arts athletes," *The Malaysian journal of medical sciences: MJMS*, vol. 20, no. 1, p. 39, 2013.

[3] S. E. Genuario and F. A. Dolgener, "The relationship of isokinetic torque at two speeds to the vertical jump," *Research quarterly for exercise and sport*, vol. 51, no. 4, pp. 593–598, 1980.

[4] L. M. Guth and S. M. Roth, "Genetic influence on athletic performance," *Current opinion in pediatrics*, vol. 25, no. 6, p. 653, 2013.

[5] D. S. Davis, D. A. Briscoe, C. T. Markowski, S. E. Saville, and C. L. Taylor, "Physical characteristics that predict vertical jump performance in recreational male athletes," *Physical Therapy in Sport*, 2003.

[6] B. E. Saelens, R. J. Seeley, K. Van Schaick, L. F. Donnelly, and K. J. O'Brien, "Visceral abdominal fat is correlated with whole-body fat and physical activity among 8-y-old children at risk of obesity," *American Journal of Clinical Nutrition*, 2007.

[7] J. B. Cronin 'and and K. T. Hansenˆ, "STRENGTH AND POWER PREDICTORS OF SPORTS SPEED," *Journal of Strength and Conditioning Research*, vol. 19, no. 2, pp. 349–357, 2005.

[8] M. C. Marques, H. Gil, R. J. Ramos, A. M. Costa, and D. A. Marinho, "Section III âĂŞ Sport, Physical Education & Recreation Relationships Between Vertical Jump Strength Metrics and 5 Meters Sprint Time," *Journal of Human Kinetics*, vol. 29, no. 29, pp. 115–122, 2011.

[9] U. Wisløff, C. Castagna, J. Helgerud, R. Jones, and J. Hoff, "Strong correlation of maximal squat strength with sprint performance and vertical jump height in elite soccer players," *Br J Sports Med*, vol. 38, pp. 285–288, 2004.

[10] P. Maulder and J. Cronin, "Horizontal and vertical jump assessment: Reliability, symmetry, discriminative and predictive ability," *Physical Therapy in Sport*, 2005.

[11] G. Markovic, "Does plyometric training improve vertical jump height? A meta-analytical review," *Br J Sports Med*, vol. 41, pp. 349–355, 2007.

[12] W. D. McLeod, S. C. Hunter, and B. Etchison, "Performance measurement and percent body fat in the high school athlete," *Am J Sports Med*, vol. 11, no. 6, pp. 390–397, 1983. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve{&}db=PubMed{&}dopt=Citation{&}list{_}uids=6650716

[13] T. K. Ho, "Random decision forests," in *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, vol. 1, 1995, pp. 278–282.

This document should only be use for personal use, and not published or distributed.

[14] ——, “The random subspace method for constructing decision forests,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, no. 8, pp. 832–844, 1998.

[15] C. Strobl and A. Boulesteix, “Bias in random forest variable importance measures: Illustrations, sources and a solution,” *BMC*, 2007.

[16] K. Archer and R. Kimes, “Empirical characterization of random forest variable importance measures,” *Computational Statistics & Data Analysis*, 2008.

[17] M. Khalilia, S. Chakraborty, and M. Popescu, “Predicting disease risks from highly imbalanced data using random forest,” *BMC medical informatics and decision making*, vol. 11, no. 1, p. 1, 2011.

[18] A. D. Flaxman, A. Vahdatpour, S. Green, S. L. James, and C. J. L. Murray, “Random forests for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards,” *Population health metrics*, vol. 9, no. 1, p. 1, 2011.

[19] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert, A. D. N. Initiative, and Others, “Random forest-based similarity measures for multi-modal classification of Alzheimer’s disease,” *NeuroImage*, vol. 65, pp. 167–175, 2013.

[20] M. Boucekine, A. Loundou, K. Baumstarck, P. Minaya-Flores, J. Pelletier, B. Ghattas, and P. Auquier, “Using the random forest method to detect a response shift in the quality of life of multiple sclerosis patients: a cohort study,” *BMC medical research methodology*, vol. 13, no. 1, p. 1, 2013.

[21] R. Rojas, “AdaBoost and the super bowl of classifiers a tutorial introduction to adaptive boosting,” *Freie University, Berlin, Tech. Rep*, 2009.