Master’s Programme in Mathematics and Operations Research

Towards proper errors!

The errors and their interpretation in automatically assessed exercises in mathematics

Hannu Tiitu

Aalto University

MASTER’S THESIS

Aalto University
MASTER'S THESIS 2017

Towards proper errors!

The errors and their interpretation in automatically assessed exercises in mathematics

Hannu Tiitu

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology.
Otaniemi, 27 Nov 2017

Supervisor: professor Lauri Malmi
Advisor: docent Jarmo Malinen

Aalto University
School of Science
Master's Programme in Mathematics and Operations Research

Aalto University

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Abstract

Author
Hannu Tiitu

Title
Towards proper errors! The errors and their interpretation in automatically assessed exercises in mathematics

School School of Science

Master's programme Mathematics and Operations Research

Major Mathematics **Code** SCI3054

Supervisor professor Lauri Malmi

Advisor docent Jarmo Malinen

Level Master's thesis **Date** 27 Nov 2017 **Pages** 70 **Language** English

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Keywords learning environment, pedagogical usability, error classification, automated assessment, teaching mathematics, STACK

urn http://urn.fi/URN:NBN:fi:aalto-201712188201

A! Aalto-yliopisto

Tiivistelmä

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tekijä
Hannu Tiitu

Työn nimi
Oikeisiin virheisiin! Virheelliset vastaukset ja niiden tulkinta automaattisesti tarkastetuissa matematiikan harjoitustehtävissä

Korkeakoulu Perustieteiden korkeakoulu

Maisteriohjelma Matematiikka ja operaatiotutkimus

Pääaine Matematiikka **Koodi** SCI3054

Valvoja professori Lauri Malmi

Ohjaaja dosentti Jarmo Malinen

Työn laji Diplomityö **Päiväys** 27.11.2017 **Sivuja** 70 **Kieli** englanti

Tiivistelmä

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Avainsanat oppimisympäristö, pedagoginen käytettävyys, virheluokittelu, automaattinen tarkastaminen, matematiikan opetus, Stack

urn http://urn.fi/URN:NBN:fi:aalto-201712188201

Contents

1. Introduction

1.1 Objectives of the Thesis

1. Can player level attribute data improve accuracy of match outcome prediction?

2. How to improve prediction accuracy of the draw class?

3. Is't possible to earn positive returns betting according the model's predictions?

2. Literature Review

1. What methods and how matches are predicted (Win/lose, score) One factor models, Multifactor models, Score vs outcome prediction, data used in multifactor,

2. How simulations are done if any

3. Elo rating

4. beating bookmaker's odds

Groll et al. [1] predicted the FIFA World Cup 2018 match outcomes using three different methods. Data used in their experiments was the as in [2]. This dataset has very limited attributes to describe teams playing styles. Team's ability covariate from the rating method is included in the input vector for the final simulations. Team's ability weights team's recent performance more than FIFA/Coca-Cola World ranking . Including this team's ability parameter improved the accuracy significantly for Random Forest. It's important to notice that the input data used in [1] is not updated during the tournament. This means that ranking stays static even though one team might perform well based on the other attributes and survive far in the tournament. This means that the team's ability is not updated to reflect the real ability based on the prediction results. Groll et al. [1] predicted the match outcomes using regression, not classification. With regression, the feature vector is conditional on match's teams but predicted scores are drawn from independent Poisson distributions. With ranking methods score is dependent. Dixon and Coles [3] identified correlation between scores and are one of the many researchers that have relaxed the strong assumption of conditional independence between the scores.

Leitner et al. [4] simulated UEFA Euro 2008 football tournament using ratings of abilities (such as the Elo rating or FIFA/Coca-Cola World ranking) and Bookmaker's odds. They showed that Bookmarker's odds outperformed Elo, but both are suitable attributes for match outcome prediction. Their models predicted only the probability of a win. This limits the model's accuracy in cases where two or more teams share the same number of points in the group stage. The model which based on the bookmaker's odds was able to predict the teams in the final correctly. One interesting finding in [4] is that FIFA/Coca-Cola World [1] rating has a higher Spearman correlation than Elo rating and could be a good metric if the corresponding winning probabilities could be computed. Results in paper [6] are contrary to these findings. Lasek et al. [6] compared different rating methods and their results imply that Elo rating describes a better team's ability compared to FIFA/Coca-Cola rating.

Simulations [4, 1] are done in a manner which takes into account group draws. Their results show that teams with relatively easier opponents in the group stage have higher probabilities to go further in the tournament.

One clear motivation behind match outcome prediction is the possibility to earn money from betting. The *Efficient market hypothesis* defined by Badarinathi and Kochmann [7] "asserts that investors cannot consistently "beat the market" because stocks reside in perpetual equilibrium" is an assumption used in finance. Weak-form efficiency means that a part of the odds given by bookmakers are mispriced. The efficient market hypothesis has been revoked in finance [8]. Goddard et al. and Badarinathi et al. [9, 7] find that this might be the case in also in betting. However Goddard et al. [9] states that there is some evidence that inefficiencies in the bookmakers' prices have diminished over time.

Kuypers [10] shares some light on how bookmakers calculate their odds. If bookmaker wants to increase its profits while keeping its over-roundness competitive it needs set the odds futher from the market effiecient odds. Also, marketing purposes and heavy one-sided bets can lead to inefficient odds. In the case of heavy one-sided bets bookmaker exposes itself to higher risk exposure if match's outcome is the one that is betted very heavily.

[1] FIFA has changed the ranking algorithm three times since it was introduced. The ranking algorithm will be changed after FIFA World Cup 2018 to resemble Elo rating [5].

3. Data

3.1 Data

In this section, I briefly describe the primary dataset of all international football matches and the secondary dataset of player attributes from the EA Sport's video game series FIFA. I have used both of the datasets to generate new data points; this process will be discussed later in this chapter.

3.1.1 International football match dataset

As the primary source of data I have used results from all international football matches from November 11th, 1872 to June 12th, 2018. This dataset is provided by Kaggle [11] and I have collected the missing games between June 4th, 2018 and June 12th, 2018 from FIFA's website [12]. From this dataset I have used columns: home team, away team, match score and tournament type.

3.1.2 EA Sport's video game series FIFA's player attributes

EA Sport's video game series FIFA describes every player in the game with several different attributes. These attributes are first collected by EA's data reviewers who are made up by coaches, professional scouts and a lot of season ticket holders from around the world. The final value is given by EA editors based on the reviewers' answers [13]. From here onwards EA Sport's video game series FIFA's player attributes are called just player attributes.

Player attributes where available from the year 2007 onwards [14]. I collected this data myself since it was not available as a single dataset.

This dataset has a single data point per player per year. In my research, I assume that the ratings are not changing too much during the season and that the team-level attributes, that are aggregated from player attributes are itself somewhat resistant to minor changes in the player-level ratings. EA Sport's has released updated ratings during the seasons, but the frequency is varying season-wise and data is not easily accessible.

All of the player attributes have a value in the range of 0-99. When two players are compared lower value means that the player's capability regarding that attribute is not as good as the other players. To guarantee similar attributes for every year only 24 player attributes are used that are listed for every year from 2007 to 2018. I have listed these attributes here and the short description is taken from Fifplay [13].

Goalkeeper:

Diving: determines a player's ability to dive as a goalkeeper.

Handling: determines a player's ability to handle the ball and hold onto it using their hands as a goalkeeper.

Kicking: determines a player's ability to kick the ball as a goalkeeper.

Positioning: determines that how well a player is able to perform the positioning on the field as a player or on the goal line as a goalkeeper.

Reflexes: determines a player's ability and speed to react (reflex) for catching/saving the ball as a goalkeeper.

Mental:

Aggression: determines the aggression level of a player on pushing, pulling and tackling.

Heading accuracy: determines a player's accuracy when using the head to pass, shoot or clear the ball.

Marking: determines a player's capability to mark an opposition player or players to prevent them from taking control of the ball.

Physical:

Acceleration: determines the increment of a player's running speed (sprint speed) on the pitch. The acceleration rate specifies how fast a player can reach their maximum sprint speed.

Reactions: determines the acting speed of a player in response to the situations happening around them.

Shot Power: determines the strength of a player's shootings.

Sprint Speed: determines the speed rate of a player's sprinting (running).

Stamina: determines a player's ability to sustain prolonged physical or mental effort in a match.

Strength: determines the quality or state of being physically strong of a player.

Skill:

Ball control: determines the ability of a player to control the ball on the pitch.

Crossing: determines the accuracy and the quality of a player's crosses.

Dribbling: determines a player's ability to carry the ball and past an opponent while being in control.

Finishing: determines the ability of a player to score (ability for finishing - How well they can finish an opportunity with a score).

Free kick accuracy: determines a player's accuracy for taking the Free Kicks.

Long passing: determines a player's accuracy for the long and aerial passes.

Long Shots: determines a player's accuracy for the shots taking from long distances.

Penalties: determine a player's accuracy for the shots taking from the penalty kicks.

Short passing: determines a player's accuracy for the short passes.

Standing tackle: determines the ability of a player to performing standing tackle.

3.1.3 Aggregating team-level attributes

When people predict the end result of a football match they often reason based on the team's attributes. They might say that Italy is more likely to win since it has a stronger defense and its ball control is excellent. This reasoning is of course very subjective, but it introduces an assumption that certain attributes could explain the outcome of a match. In this section, I explain how I have combined player attributes to team-level attributes to describe a football team based on its players' capabilities. Data for starting lineups was not available.

To describe the whole football team I have used the average value from 23 best players. These attributes are: *overall rating, potential, age, height, weight, international reputation* and *weak foot*. As an extra attribute, I calculated the average age from the top 11 players. The idea behind this attribute is to get the average age of the presumed starting lineup.

The other attributes are described by subsection of the players. I have used my knowledge and intuition on football to select the sizes of the subsetions. The goalkeeper attributes are calculated based on the team's two best values for that attributes. The average for the skills required in set piece situations, like the attributes *free kick accuracy* and *penalties* for example, are calculated based on top three ratings, since in most cases very small subset of players handle these situations. Also sprint speed and marking are only calculated based on top three values. *Strength* and *stamina* are calculated using 10 best ratings. Other values are calculated using five best ratings for each attribute.

In cases where the team doesn't have enough players to calculate the attribute's value as mentioned in the previous part I have used this formula $\frac{1}{N}\sum_{i=1}^{N} x_i * \max(N/K, C)$. If N (the number of all available ratings for the attribute), is smaller than K (the required number of ratings) the average is multiplied with a coefficient that has a integer value in the inclusive range from C and 1. I have set C to 0.9.

3.1.4 Historical odds

I have collected odds for the FIFA World Cup 2018 and 2014 from Odds Portal [15]. Odds Portal has collection of odds offered by multiple betting sites. For every match outcome I have used the average value from each availabe odd. Kuypers [10] mentions that football odds are mostly fixed. Based on values that Odds Portal offer, this is not the case anymore. Many odds have changed from the initial opening odd before the match start. I have used the latest value every odd since the data is easier to collect that way.

4. Methods

4.1 Predicting football match outcome with Random Forest Classifier

What is is done. Why random forest? What are the alternative methods

4.1.1 Random Forest

In machine learning supervised learning is the task of learning the relationship between input features and and the target value. Structure that describes this relathionship is called a model. In most cases these models are used to predict the target value based on new input features. There are two types of models: *regression models* and *classifier models*. If the target value is in a real-valued domain the model is called *regression model*. *Classifier models* are used to map the input features to predefined classes. [16]

Decision tree is one of the most popular model type used in classification problems. Decision tree is a directed tree, which means that all of the nodes, except the *root node*, have exactly one incoming edge. Nodes that have outgoing edges are called *internal nodes* and the nodes that have only incoming edge are called the *leave nodes*. Internal nodes in decision tree split the instance space into two or more subspaces according to a certain discrete function of the input's feature values. Usually split is done based on a single feature from the whole feature vector. A single class value is assigned for the *leave nodes*. When new input is given tree is navigated from the *root node* to a *leave node* which determinates the predicted class label. In regression these target values can take continuous values. [16]

Decision trees have many benefits and are very useful "off-the-shelf" predictors. Outliers in the dataset or many irrelevant predictors are not problematic for the trees. Scaling or any other general transformation can be done to the input space since trees are invariant under transformation of the individual predictors. [17] Decision trees have good interpretability if the trees are small.

On main disadvantage of the decision trees is bad prediction accuracy [17]. Decision trees can express the training data well, but have high variance which means that prediction accuracy for unseen data is often worse compared to other models.

Bootstrap aggragating, also called bagging, is a way to improve the prediction accuracy of decision trees by averaging. In bagging the average is taken over the output of a multiple estimators:

$$\hat{f_{bag}}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{*b}(x) \tag{4.1}$$

where B is the number of estimators and f^* is a single estimator. This reduces the high variance of a single tree and makes predictions more accurate.

Bootstrap in bagging means that in the training of a single tree a random sample with replacement is taken from the original sample. Samples used in training come from the same distribution, meaning that the trees are identically distributed (i.d.). This combined with deep trees that have less bias ensures that the variance reduction achieved in bagging comes with a expense of small increase in bias and the loss of interpretability. The loss of interpretability can't be avoided since a single tree can't be used anymore for reasoning. Trees in bagging are only identically distributed. The missing independent property means that the trees in the forest can have pairwise correlation. This is common in cases where input data has one strong predictor which often leads to a situation where all of the tree are splitted similarily. [17]

Amit and Geman's [18] idea of random feature selection inspired Breiman to use bagging in tandem with random feature selection. With this random feature selection correlation between the trees can be reduced since the generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them [19]. Breiman was first to use the name *Random Forest* for algorithms that use bagging and random feature selection with tree predictors [19]. Step-by-step instruction from [17] for Random Forest algorithm are listed in the Algorithm 1.

Main usecases for Random forest are *classification* and *regression*.

Algorithm 1: Random Forest for Regression or Classification.

Here I have some text that I need to have within the algorithm that sets up the problem, and the text stretches over two lines. **Note that the second line has a weird indentation.**

For all the of the seconds in the day:

1. For $b = 1$ to B:
   (a) Draw a bootstrap sample $\boldsymbol{Z}^*$ of size N from the training data.
   (b) Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached.
      i. Select m variables at random from the p variables.
      ii. Pick the best variable/split-point among the m.
      iii. Split the node into two daughter nodes.
2. Output the ensemble of trees $\{T_b\}_1^B$.

To make a prediction at a new point x:
Regression: $\hat{f}_{\text{rf}}^B(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$
Classification: Let $\hat{C}_b(x)$ be the class prediction of the bth random forest tree. Then $\hat{C}_{rf}^B(x)$ = *majority vote* $\left\{\hat{C}_b(x)\right\}_1^B$

4.1.2 Random forest hyperparameter selection

Random forest, as implemented in most of the programming libraries, can be used out-of-the-box with fairly good perfomance. Reason for this is that the default values for certain hyperparameters work well in most of the cases. Lot of effort has been used to study the effects of hyperparamater selection in random forest. Main idea in hyperparameter selection with random forest is to keep the correlation between the trees small while maintaining reasonable strength. In this section I will share some light to this process of hyperparameter selection and introduce an algorithm that will be used to select the most optimal hyperparamters for the prediction model.

One key hyperparameter in random forest is the number of randomly drawn candidate predictors, denoted as K. As mentioned in 4.1.1 this part of random forest algorithm is critical if it's important to achieve low correlation between the trees in the forest. Bernard et al. [20] investigated Breiman's Forest-RI algorithm's performance with different K values using several datasets. Their conclusion was that $K = \sqrt(M)$, where M is the predictor space, is a reasonable setting to induct near-optimal random forest. In cases where there are many or only few relevant predictor variables, choosing the K can have high influence on the results. For example in the case of minuscule K with a dataset that has only small number of important predictors most of the trees are built without the important predictor and have low prediction accuracy. [20] Most of the time while running the algorithm is used to choose the split variables. When number of candidate variables is decreased computation time approximately decreases linearly. [21]

Another hyperparameter that has the correlation lowering effect is the subsample size. Typically the subsample size is the same as the dataset's size, even when subset is formed with replacements. In their emprical analysis Martinez et al. [22] concluded that selecting the subsample size with out-of-bag error can increase the performance of the model and the correct value for subsample size hyperparameter is problem dependent.

Two important hyperparameter related to node splitting are node size and splitting criterion. Segal [23] showed that increasing the amount of noise variables lead to higher optimal node size. For splitting criterion most of the software packages use gini impurity for classification and MSE for regression, which were suggested by Breiman [19]. Probst et al. [24] concluded from multiple studies that not a single criterion has been proven superior to the others regarding the predictive performance.

Since random forest is an ensemble method, the number of estimators (trees) can be changed. For this hyperparameter it's important to set the value as high enough, since after a certain point the variance can't be lowered anymore and the computation time used in training grows beyond resonable [24].

4.1.3 Variable importance in Random Forest

4.1.4 Model specifications

4.2 Forecasting procedure and measuments of model's forecasting accuracy

4.2.1 Description of our forecasting procedure

4.2.2 Forecast accuracy measurements

Precision, recall, F-score, accuracy of probabilistic prediction: Brier score, label ranking loss

4.2.3 Evaluating betting strategy

UNIT, KELLY

5. Results

6. Conclusion

Bibliography

[1] A. Groll, C. Ley, G. Schauberger, and H. Van Eetvelde, "Prediction of the fifa world cup 2018-a random forest approach with an emphasis on estimated team ability parameters," *arXiv preprint arXiv:1806.03208*, 2018.

[2] A. Groll, G. Schauberger, and G. Tutz, "Prediction of major international soccer tournaments based on team-specific regularized poisson regression: An application to the fifa world cup 2014," *Journal of Quantitative Analysis in Sports*, vol. 11, no. 2, pp. 97–115, 2015.

[3] D. M. J. and C. S. G., "Modelling association football scores and inefficiencies in the football betting market," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 46, no. 2, pp. 265–280.

[4] C. Leitner, A. Zeileis, and K. Hornik, "Forecasting sports tournaments by ratings of (prob) abilities: A comparison for the euro 2008," *International Journal of Forecasting*, vol. 26, no. 3, pp. 471–481, 2010.

[5] Wikipedia contributors. (2018) Fifa world rankings — Wikipedia, the free encyclopedia. [Online; accessed 10-July-2018]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Plagiarism&oldid=5139350

[6] J. Lasek, Z. Szlávik, and S. Bhulai, "The predictive power of ranking systems in association football," *International Journal of Applied Pattern Recognition*, vol. 1, no. 1, pp. 27–46, 2013.

[7] R. Badarinathi and L. Kochman, "Football betting and the efficient market hypothesis," *The American Economist*, vol. 40, no. 2, pp. 52–55, 1996.

[8] N. Jegadeesh and S. Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," *The Journal of finance*, vol. 48, no. 1, pp. 65–91, 1993.

[9] J. Goddard and I. Asimakopoulos, "Modelling football match results and the efficiency of fixed-odds betting," Working Paper, Department of Economics, Swansea University, Tech. Rep., 2003.

[10] T. Kuypers, "Information and efficiency: an empirical study of a fixed odds betting market," *Applied Economics*, vol. 32, no. 11, pp. 1353–1363, 2000.

[11] M. Jürisoo. (2018) International football results from 1872 to 2018. [Online; accessed 10-July-2018]. [Online]. Available: https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017

[12] FIFA. (2018) International football results. [Online; accessed 10-July-2018]. [Online]. Available: https://www.fifa.com/live-scores/index.html

[13] F. Play. (2018) Fifa encyclopedia. [Online; accessed 10-July-2018]. [Online]. Available: http://www.fifplay.com/encyclopedia/

[14] Sofifa. (2018) Ea sport's video game series fifa's player attributes. [Online; accessed 10-July-2018]. [Online]. Available: https://sofifa.com

[15] Odds Portal. (2018) World cup 2018 results & historical odds. [Online; accessed 12-July-2018]. [Online]. Available: http://www.oddsportal.com/soccer/world/world-cup/results/

[16] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers-a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 4, pp. 476–487, 2005.

[17] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, NY, USA:, 2001, vol. 1, no. 10.

[18] Y. Amit and D. Geman, "Shape quantization and recognition with randomized trees," *Neural computation*, vol. 9, no. 7, pp. 1545–1588, 1997.

[19] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[20] S. Bernard, L. Heutte, and S. Adam, "Influence of hyperparameters on random forest accuracy," in *International Workshop on Multiple Classifier Systems*. Springer, 2009, pp. 171–180.

[21] M. N. Wright, T. Dankowski, and A. Ziegler, "Unbiased split variable selection for random survival forests using maximally selected rank statistics," *Statistics in medicine*, vol. 36, no. 8, pp. 1272–1284, 2017.

[22] G. Martínez-Muñoz and A. Suárez, "Out-of-bag estimation of the optimal sample size in bagging," *Pattern Recognition*, vol. 43, no. 1, pp. 143–152, 2010.

[23] M. R. Segal, "Machine learning benchmarks and random forest regression," 2004.

[24] P. Probst, M. Wright, and A.-L. Boulesteix, "Hyperparameters and tuning strategies for random forest," *arXiv preprint arXiv:1804.03515*, 2018.

A. Ut purus elit

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

A.1 Fusce mauris

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi.

Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.