Supplementary Data
The prediction of caregiver burden in Amyotrophic Lateral Sclerosis: A machine learning approach using Random Forests applied to a cohort study

Anna Markella Antoniadi, Miriam Galvin, Mark Heverin, Orla Hardiman and Catherine Mooney

1 Methods

1.1 Data Collection

The McGill Quality of Life (McGill) questionnaire was used to measure the quality of life on a scale of 0 ("very bad") to 10 ("excellent") for each question.[1] The Hospital Anxiety and Depression Scale (HADS) is a self-rating scale that was used in order to measure the symptoms of anxiety and depression.[2] From the aforementioned questionnaires, we included both the total score as well as each individual score (from each individual question) in our analysis. Caregiver burden was determined from the Zarit Burden Interview (ZBI), that is used to self-assess the frequency of several positive or negative feelings.[3] The level of burden according to the ZBI can range from 0 to 88.

Regarding the stage of the disease for the patients, both the Milano-Torino (MiToS) functional staging,[4] and King's clinical staging,[5] systems will be considered, as recommended,[6]. The MiToS system measures the loss of independent function in four key domains: walking/self-care, swallowing, communicating and breathing. Stage 0 indicates no loss of independence in any domain despite functional involvement, stages 1-4 indicate the number of domains with loss of independence and stage 5 is death. In King's staging system, stages 1-3 are reached according to the number of the affected regions as per the El Escorial,[7] criteria (bulbar, upper limbs, lower limbs) and stage 4 is reached in the case of feeding or respiratory failure secondary to ALS. The Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised,[8] (ALSFRS-R) was used as a measure of the progression of ALS. As opposed to the staging scale, the higher the ALSFRS-R score is, the better the condition of the patient. The total score ranges between 0 and 48 and the individual questions quantify symptoms of the regions that are used in the El Escorial criteria. Cognitive and behavioural impairment were assessed using the Edinburgh Cognitive and Behavioural ALS Screen (ECAS),[9, 10] and the Beaumont Behavioural Inventory (BBI),[11] respectively.

1.2 Data Analysis

We used LASSO,[12] (Least Absolute Shrinkage and Selection Operator) regression and gradient tree boosting methods to create predictive models and compare the results with the Random Forest models. LASSO was used instead of a standard logistic regression because of the large number of variables in our dataset (as a feature selection method). The lamda parameter that was used was identified using 10-fold cross validation with the method cv.glmnet from the "glmnet",[13] package in R version 3.0-2, and its value was 0.05395605. To create the model we used the function "glmnet" from the same package. XGBoost,[14] (Extreme Gradient Boosting) is a popular implementation of the gradient tree boosting method,[15] that is freely and readily available, easy to use, and accurate and used in many ML competitions. The R package that was used was the "xgboost" version 0.90.0.2. In terms of parameter tuning, the maximum depth of each decision tree was 3 (we tried 3 and 4), the number of boosting rounds was 10 (among 5, 10 and 15), the early-stopping rounds in case of no improvement were 3 (among 3 and 5).

The measures that were used for the evaluation process were sensitivity, specificity and Matthews correlation coefficient (MCC):

$$Sensitivity = \frac{TP}{P}$$

$$Specificity = \frac{TN}{N}$$

$$MCC = \frac{(\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN})}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

where P stands for Positives, TP for True Positives, FN for False Positives, N for Negatives and TN and FN for True and False Negatives respectively.

Seventy-five percent of the caregivers were assigned into the training dataset and the remaining were used in the independent test set. The raw numbers of this split are shown in Table S1.

Table S1: Distribution of caregivers into burden classes.

Data	*LowBurden*(0)	*HighBurden*(1)
Full	74	103
Training	56	78
Test	18	25

2 Results

2.1 Study Participants

One hundred patients or patient-carer dyads were recruited for the study but not all were eligible, and many were lost to follow-up. The main consideration around eligibility for the current analysis was that we could only use patient/caregiver dyads and could not use data where the patient had participated without a caregiver. Apart from this eligibility issue, the reasons for the drop in numbers between each time point included: the death of the patient, the patient was too sick to continue in the study, the patient or the caregiver did not wish to complete a further follow up. The numbers at each stage are shown in a flowchart in Figure S1.

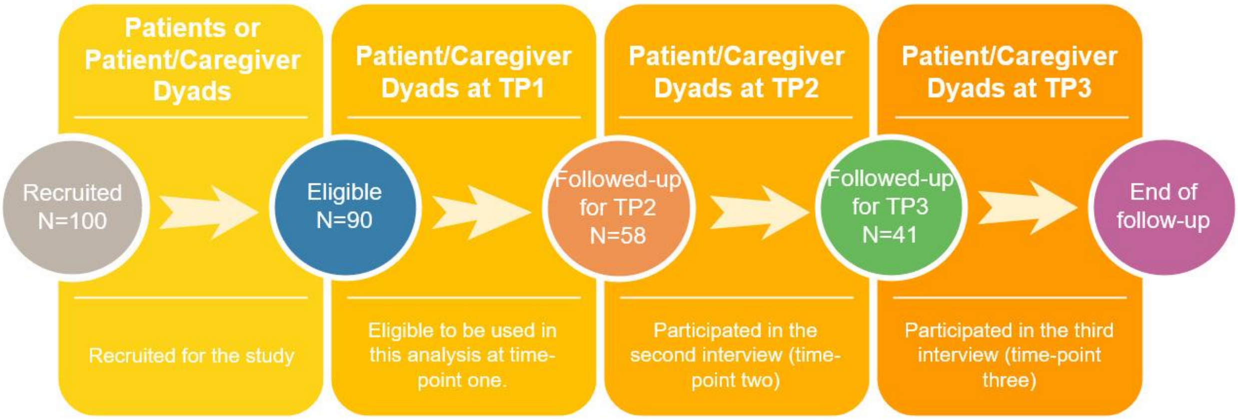


Figure S1: Flowchart of participation at each stage

.

2.2 Patient Demographics

The average ALSFRS-R scores of the 73 patients that had a score recorded in the ALS Registry (17 missing) can be found in Table S2. The bulbar and respiratory scores can range between 0 and 12, while the motor scores can range between 0 and 24.

Table S2: ALSFRS-R scores of patients at baseline.

Score	*Mean*	*SD*	*Min*	*Max*
ALSFRS − R Total	33.04	8.10	14	46
ALSFRS − R Bulbar	9.21	2.80	1	12
ALSFRS − R Motor	13.77	5.84	0	24
ALSFRS − R Respiratory	10.07	2.59	3	12

Antoniadi AM, *et al. BMJ Open* 2020; 10:e033109. doi: 10.1136/bmjopen-2019-033109

King's staging: From the patients that were interviewed at the second time-point (n=58) and that had a recorded King's Stage at that time (n=55), 16 (29.1%) had a change of stage since baseline. Also, from the patients that participated in the third interview (n=41), 5 (12.1%) had a change of stage since T1 (all patients had a recorded King's stage at this point). It is important to note that 49 (54.4%) of the initially interviewed patients (n=90) were already assigned to Stage 4 (Using NIV or gastrostomy) and could not progress to another stage of the disease. From the patients that were in stage 4 at baseline, 26 also participated in the second interview and 17 participated in all interviews.

MiToS staging: From the patients that participated in the second interview (n=58) and that had a recorded MiToS Stage at that time (n=52), 12 (23.1%) had a change of stage since. Also, from the patients that participated in all interviews (n=41) and that had a recorded MiToS Stage at that time (n=34), 8 (23.5%) had a change of stage since the first interview. The importance of using both staging systems can be reflected on the considerable differences in the distribution of patients across stages at the same time-point.

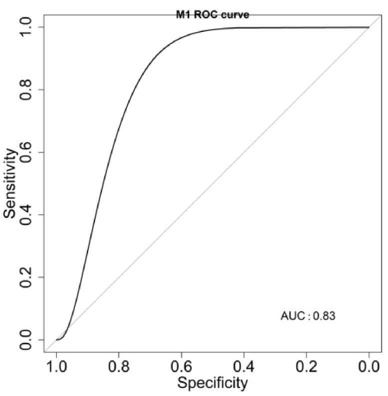
Table S3: Stages of patients at each timepoint.

Staging System	*Stage*	*Time*1	*Time*2	*Time*3
King's		n=88	n=55	n=41
	1	10 (11.4%)	5 (9%)	3 (7.3%)
	2	14 (15.9%)	3 (5.5%)	4 (9.8%)
	3	15 (17.0%)	11 (20%)	6 (14.6%)
	4	49 (55.7%)	36 (65.5%)	28 (68.3%)
MiToS		n=87	n=52	n=34
	0	49 (56.3%)	23 (44.2%)	14 (41.2%)
	1	26 (30.0%)	23 (44.2%)	15 (44.1%)
	2	11 (12.6%)	6 (11.6%)	5 (14.7%)
	3	1 (1.1%)	0	0
	4	0	0	0

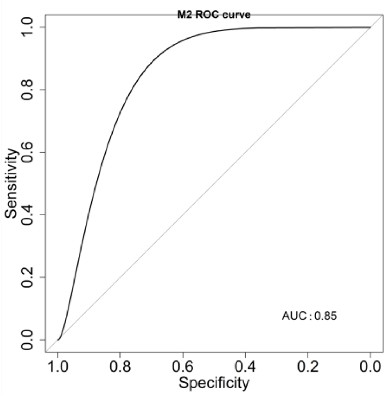
2.3 Machine Learning

The results of the experiment using LASSO were: 0.64 sensitivity, 0.83 specificity and 0.47 MCC, while the XGBoost results were 0.76 sensitivity, 0.78 specificity and 0.53 MCC.

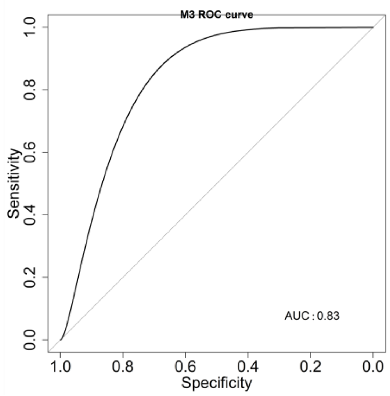
The results of the first set of models (M1-M6) are presented below. Figure S2 shows the ROC Curves for the models, to provide additional visual information to the AUC that is shown in the table in the main text. Figure S3 contains the importance of the variables in the models that were not presented in the main text (M1, M3, M4, M5, M6).



(a) M1: model created using missForest-imputed full dataset.

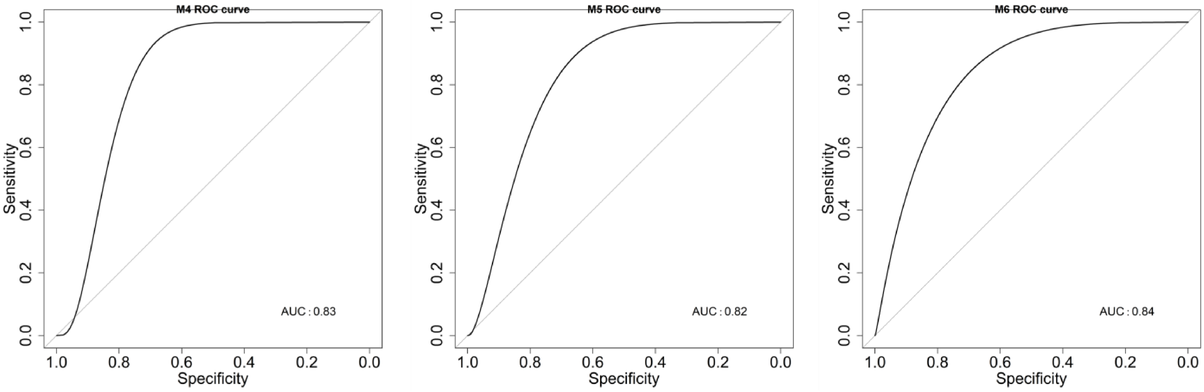


(b) M2: model that includes the 25 variables mostly used in M1.



(c) M3: model that includes the 15 variables mostly used in M1.

Antoniadi AM, *et al. BMJ Open* 2020; 10:e033109. doi: 10.1136/bmjopen-2019-033109



(d) M4: model created using the Median-imputed full dataset.

(e) M5: model that includes the 25 variables mostly used in M4.

(f) M6: model that includes the 15 variables mostly used in M4.

Figure S2: ROC Curves of models M1-M6.

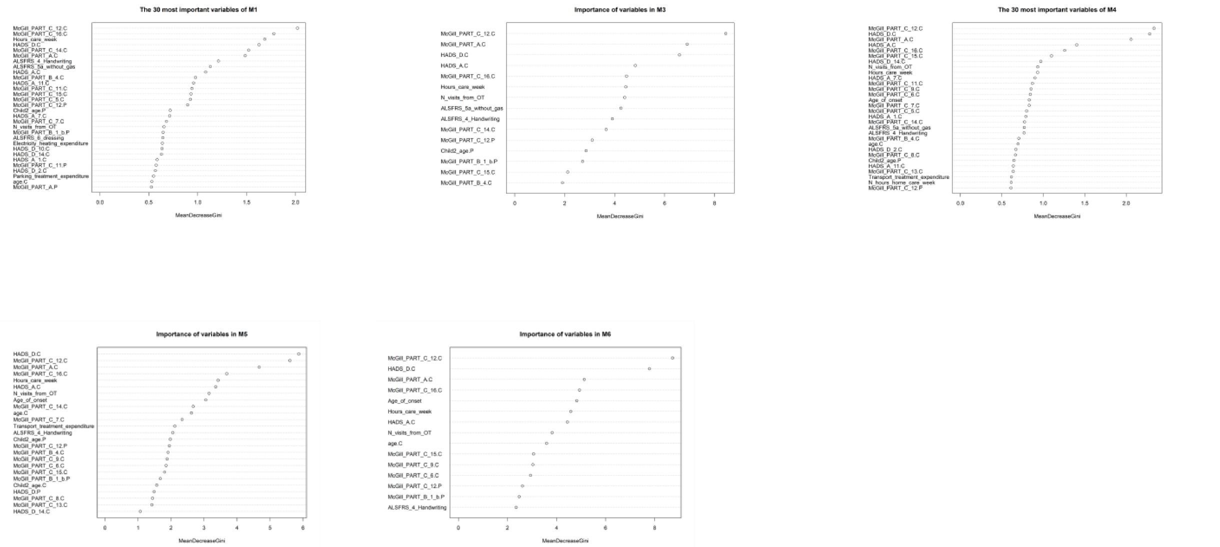


Figure S3: Most important variables of models M1-M6 according to Mean Decrease of the Gini index.

A list of the subset of 76 variables that were selected for the development of a CDSS (the creation of models M7-M12):

1. Demographic and financial information:
   - Hours of care per week provided by the primary caregiver.
   - Whether the patient and caregiver live together.
   - Primary caregiver's age, current and pre-caregiving activity status, carer's monthly benefit and allowance, health condition (self-assessed), presence of any long-term illness/disability, change of employment status since caregiver duties, respite care offer.
   - Patient's monthly disability allowance, principal status, housing tenure, county, medical card ownership and condition, primary medical certificate, IMNDA (Irish Motor Neurone Disease Association) grant, invalidity pension per month, palliative services use, activity status before symptoms, visits to and from the healthcare professionals.

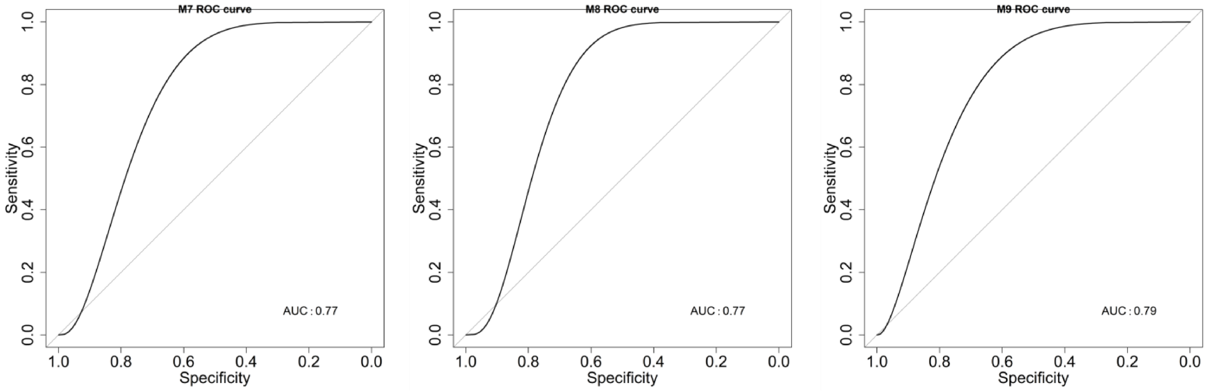
- The patient's and primary caregiver's marital status, sex, number of children, living arrangements, education, car access, employment status, health insurance coverage.

2. Patient's clinical information.

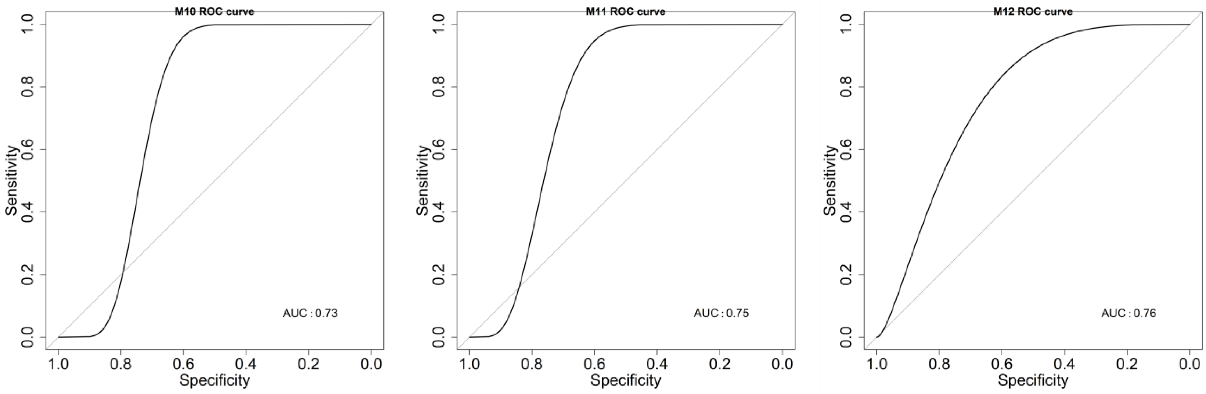
- Age and site of onset, first symptom (lower or upper limb), El Escorial Diagnosis
- Visit to the Accident and Emergency (A&E) department due to fall
- ALSFRS-R individual scores, cognitive and behavioural impairment, MITOS and King's stage
- Number of walking, posture, and home aids and appliances used.

The results of models that were created using the aforementioned subset of variables (M7-M12) are presented below. Figure S4 shows the ROC Curves for the models and Figure S5 contains the importance of the variables in the models that were not presented in the main text (M7, M8, M10, M11, M12).

Antoniadi AM, *et al. BMJ Open* 2020; 10:e033109. doi: 10.1136/bmjopen-2019-033109

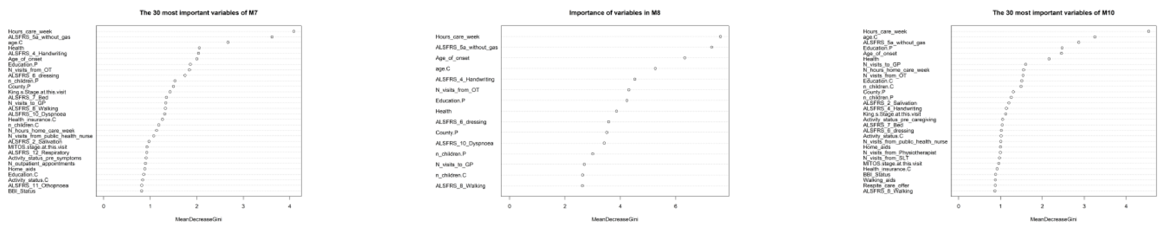


(a) M7: model created using missForest-imputed full dataset.
(b) M8: model that includes the 25 variables mostly used in M7.
(c) M9: model that includes the 15 variables mostly used in M7.



(d) M10: model created using the Median-imputed full dataset.
(e) M11: model that includes the 25 variables mostly used in M10.
(f) M12: model that includes the 15 variables mostly used in M10.

Figure S4: ROC Curves of models M7-M12.



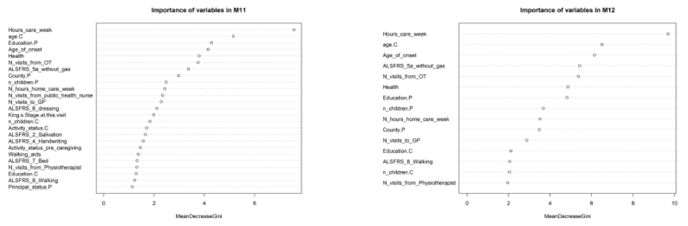


Figure S5: Most important variables of models M7 and M10 according to Mean Decrease of the Gini index.

Antoniadi AM, *et al. BMJ Open* 2020; 10:e033109. doi: 10.1136/bmjopen-2019-033109

References

[1] Cohen SR, Sawatzky R, Russell LB, et al. Measuring the quality of life of people at the end of life: The McGill Quality of Life Questionnaire–Revised. *Palliative Medicine*. 2017;31(2):120– 129.

[2] Moorey S, Greer S, Watson M, et al. The Factor Structure and Factor Stability of the Hospital Anxiety and Depression Scale in Patients with Cancer. *British Journal of Psychiatry*. 1991;158(2):255–259.

[3] Zarit SH, Reever KE, Bach-Peterson J. Relatives of the Impaired Elderly: Correlates of Feelings of Burden. *The Gerontologist*. 1980;20(6):649–655.

[4] Chiò A, Hammond ER, Mora G, et al. Development and evaluation of a clinical staging system for amyotrophic lateral sclerosis. Journal of Neurology, *Neurosurgery & Psychiatry*. 2015;86(1):38–44.

[5] Roche JC, Rojas-Garcia R, Scott KM, et al. A proposed staging system for amyotrophic lateral sclerosis. *Brain*. 2012;135(3):847–852.

[6] Fang T, Khleifat AA, Stahl DR, et al. Comparison of the King's and MiToS staging systems for ALS. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*. 2017;18(3-4):227–232.

[7] Brooks BR. El escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*. 1994;124:96 – 107.

[8] Cedarbaum JM, Stambler N, Malta E, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. Journal of the Neurological Sciences. 1999;169(1):13 – 21.

[9] Abrahams S, Newton J, Niven E, et al. Screening for cognition and behaviour changes in ALS. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*. 2014;15(1-2):9–14. PMID: 23781974. Available from: https://doi.org/10.3109/21678421.2013.805784.

[10] Niven E, Newton J, Foley J, et al. Validation of the Edinburgh Cognitive and Behavioural Amyotrophic Lateral Sclerosis Screen (ECAS): A cognitive tool for motor disorders. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*. 2015;16(3-4):172–179.Available from: https://doi.org/10.3109/21678421.2015.1030430.

[11] Elamin M, Pinto-Grau M, Burke T, et al. Identifying behavioural changes in ALS: Validation of the Beaumont Behavioural Inventory (BBI). *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*. 2017;18(1-2):68–73. PMID: 27894191. Available from: https://doi.org/10.1080/21678421.2016.1248976.

[12] Tibshirani R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B (Methodological).* 1996;58(1):267–288. Available from: http://www.jstor.org/stable/2346178.

[13] Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, Articles. 2010;33(1):1–22. Available from: https://www.jstatsoft.org/v033/i01.

[14] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* KDD '16. New York, NY, USA: Association for Computing Machinery; 2016. p. 785–794. Available from: https://doi.org/10.1145/2939672.2939785.

[15] Friedman J. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*. 2001 10;29:1189–1232.