

*Leveraging Machine Learning to Assist
Clinicians with Clinical Dementia Rating
(CDR) Scores in Online Dementia Diagnosis*

Albert Weerman, Danish Jain

Paper No: 2023-010

**CESR-SCHAEFFER
WORKING PAPER SERIES**

The Working Papers in this series have not undergone peer review or been edited by USC. The series is intended to make results of CESR and Schaeffer Center research widely available, in preliminary form, to encourage discussion and input from the research community before publication in a formal, peer-reviewed journal. CESR-Schaeffer working papers can be cited without permission of the author so long as the source is clearly referred to as a CESR-Schaeffer working paper.

Leveraging Machine Learning to Assist Clinicians with Clinical Dementia Rating (CDR) Scores in Online Dementia Diagnosis

Albert Weerman¹ and Danish Jain²

¹Center for Social and Economic Research, University of Southern California

²Viterbi, University of Southern California

September 13, 2023

Abstract

To improve dementia diagnostic reliability and accuracy some studies suggest a consensus panel approach for the diagnosis of dementia (1-2). A panel of clinicians meet in person or online to discuss and use the Clinical Dementia Rating (CDR) Score to assess the study participant's cognitive and clinical profile. A few studies including the Longitudinal Aging Study in India (LASI-DAD) and the Health and Aging in Africa: A Longitudinal Study of an INDEPTH Community in South Africa (HAALSI) have moved this rating for their first wave of data collection to an online system where raters enter their diagnosis. The system then filters out cases where no consensus is reached. The raters can go back in the system to update their rating or the case moves to an in-person or online meeting where a moderator discusses the case. To aid the moderator in the second wave of LASI-DAD and HAALSI data collection, we aim to provide a CDR Score prediction based on data collected in wave 1 in combination with the clinician's ratings.

Key words: Clinical Dementia Rating (CDR), dementia, LASI - DAD, HAALSI

1 Introduction

Dementia is a condition where there is a loss in cognitive functioning - remembering, thinking and reasoning. The severity ranges from mild symptoms

where a person's dementia just begins to affect a person's functionality to severe stages where a person has to completely depend on someone else. In recent years, machine learning (ML) models have emerged as powerful tools for predicting dementia and assessing its severity. The aim of this research paper is to investigate the application of machine learning (ML) models for the prediction of Clinical Dementia Rating (CDR) scores. CDR is a global rating scale used to measure the severity of Dementia ranging from 0 (No Dementia) to 3 (severe cognitive impairment).¹⁻³ By utilizing ML models to predict CDR scores, we aim to enhance the precision and speed of dementia assessments, facilitating early intervention and treatment for individuals affected by this condition.

Our research paper details the datasets used in our study, which comprises clinical information from a large sample of individuals with dementia. The core of our study focuses on the application of machine learning models for predicting CDR scores, which are widely utilized for assessing dementia severity.⁴ We discuss the specific models we employed, including decision trees, random forests, support vector machines, XGBoost and ensemble learning models, and provide insight into how these models were trained and validated using our datasets.⁵ Additionally, we analyze and compare the performance of these models in predicting CDR scores.

To conclude, we emphasize the potential benefits of using machine learning models for early detection and personalized treatment of dementia.

2 Datasets

Our study utilized data from the Health and Aging in Africa: A Longitudinal Study of an indepth Community in South Africa (HAALSI) and the Longitudinal Aging Study in India - Diagnostic Assessment of Dementia (LASI-DAD) study, which consisted of adults aged 60 and over who underwent comprehensive neuropsychological testing and informant interviews between 2018 and 2020.⁶⁻⁷ To predict CDR ratings for wave 2, we trained a XGboost model leveraging GridsearchCV its performance. We used clinical consensus ratings from a subset of the sample from wave 1 to train our model, and predicted CDR ratings for the round 2 respondents. After data cleaning, which involved removing records with missing data (‘.x’) and replacing all 0.5 CDR Dementia ratings in the dataset with a 0 rating, we obtained a dataset of 2374 records containing both numeric and categorical data.⁸ We dealt with missing values in each column either by replacing them with the mean or mode of that column, depending on the column type. Categorical values were encoded using label encoders. Finally, we split the dataset into training and testing set in a 3:1 ratio to train and evaluate our model’s performance.

3 System description

Following the completion of the data pre-processing stage, the next step was to split the datasets into training and testing sets, and then train several machine learning (ML) classifiers, including Logistic Regression, KNeighbors, Decision Tree, Support Vector Machine, Random Forest, XGBoost and ensemble models on the training set. To obtain a more accurate assessment of the performance of the classifiers, we utilized k-fold cross-validation technique.

Performance metrics such as accuracy, precision, recall and F1-score were calculated for each fold,

and the results were then averaged across all five folds.⁹ This approach allowed us to obtain a more reliable estimate of the overall performance of each classifier. Based on the results of k-fold cross-validation, we selected the best-performing classifier and further evaluated its performance on the testing set.¹⁰

3.1 Logistic Regression Classifier

Logistic Regression is a supervised machine learning model that can be used as a classifier working on the supposition that the data can be split by a line or a n-dimensional plane.¹¹ Classification is done by calculating the value of a first degree polynomial of the following form:

$$y = w_1 * x_1 + w_2 * x_2 + \dots + w_n * x_n \quad (1)$$

where y is the target variable, x is the input parameter, w is the weight assigned to this parameter, and n is the number of input parameters. After generating the output (y) using logistic regression, the next step is to apply a logistic function, such as the sigmoid to y. This transforms y to a value between 0 and 1.¹² To classify y into class A or B, we applied a decision rule that compared the predicted probability to a threshold value of 0.5. If the predicted probability was greater than or equal to 0.5, the observation was classified as class A or else class B. To perform multiclass classification we make use of the one-vs-rest (“ovr”) option. It involves breaking down a multiclass classification problem into multiple binary classification problems, where each binary classifier distinguishes one class from the rest of the classes. During prediction, the model with the highest confidence score among all binary classifiers is chosen as the predicted class for a given input.

3.2 KNN

K-Nearest Neighbors is a supervised machine learning algorithm which can be utilized for multiclass classification. It works on a foundational assumption that data points that share similar

characteristics are apt to be clustered or located in proximity to one another.¹³ To classify a new data point, the algorithm considers the labels of the k closest data points in the feature space and assigns the majority

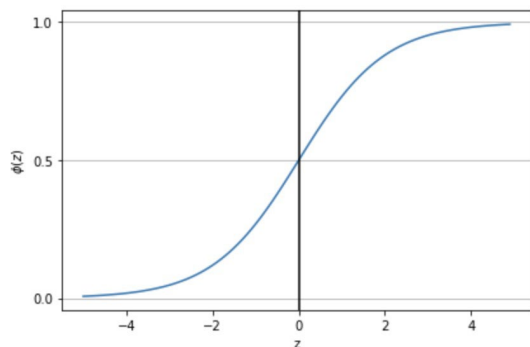


Figure 1: Sigmoid Function

class label among them as the predicted label for the new data point.¹⁴

each class in the node subtracted from 1. We divide the dataset at each stage in such a way that the Gini index is reduced making the feature space pure. This process continues until a stopping criterion is met, such as reaching a maximum tree depth or having too few instances in a node. After the decision tree is constructed, at each node the split rule is used to determine which branch to follow. Once a leaf node is reached, the instance is assigned to one or more classes based on the majority proportion of instances belonging to each class in that node.¹⁸

In case of a multiclass classification, there can be multiple classes assigned to a leaf node, and the majority class is determined based on the proportion of instances belonging to each class. In case of a tie CART assigns a class based on some predefined rule.

Classification and Regression Trees (CART) are supervised machine learning models used for binary and multiclass classification. They work by splitting the dataset into small sections until a stopping condition is met.¹⁵⁻¹⁶ The process begins by splitting the dataset at the root node based on a selected feature and a splitting criterion which can be estimated by calculating the Gini index.¹⁷ Gini index is calculated as the sum of the squared probabilities of

3.4 Support Vector Machines

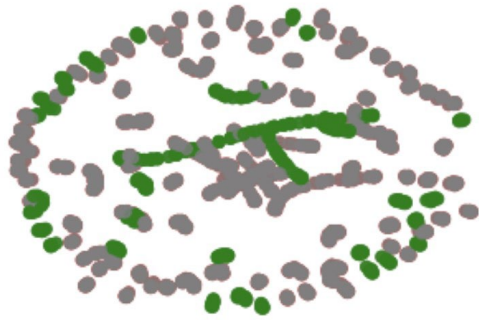


Figure 2: KNN

3.3 Decision Tree Classifier (CART)

Figure 3: KNN

3.5 Random Forest Classifier

Random Forest is an ensemble supervised machine learning algorithm widely used in classification and regression tasks.²⁰ An ensemble model is a model which is created by the combination of multiple models. In a random forest model, different decision trees are constructed with the help of subset of features and labels. Every decision tree constructed has an output generated and the final label is decided by majority voting.²¹

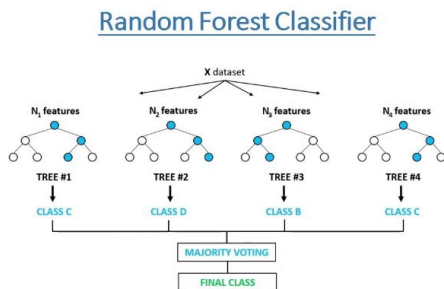
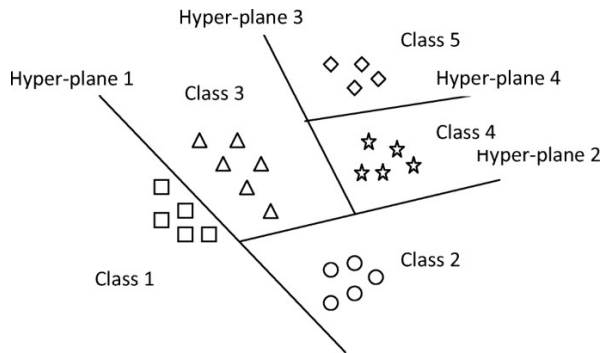


Figure 4: Random Forest Classifier

Support Vector Machine (SVM) is a supervised machine learning algorithm that helps in classification or regression problems by finding an optimal boundary between the labels. The principal foundation of SVM is to find a hyperplane that maximizes the separation of the data points to their potential classes in an n -dimensional space.¹⁹ SVM supports only binary classification. However, to overcome this limitation, SVM makes use of One-to-One approach, which breaks down the multiclass problem into multiple binary classification problems.



3.6 XGBoost

XGBoost is a decision-tree based ensemble machine learning algorithm which makes use of gradient boosting. Gradient boosting is a technique that uses a gradient descent algorithm to add weak models in an iterative manner. This technique sets specific target outcomes for the next model based on the gradient of the error with respect to the prediction, and thus, is called gradient boosting.²² GBDTs (Gradient Boosting Decision Trees) train a set of shallow decision trees in an iterative manner, where each tree is trained using the residual errors of the previous model. The final prediction is a weighted sum of all the tree predictions. GBDTs iteratively train an ensemble of shallow decision trees, with each iteration using the error residuals of the previous model to fit the next model. The final prediction is a weighted sum of all of the tree predictions. Random forest “bagging” minimizes the variance and overfitting, while GBDT “boosting” minimizes the bias and underfitting.

3.7 Ensemble Learning

Ensemble Learning is an approach in which predictions from multiple models are combined to bolster the predictive performance. It combines results from two or more different models to produce a result which has a higher accuracy than any of the models individually. For the purpose of our research, we have combined results from Logistic Regression, Gradient Boosting Classifier and Random Forest Classifier to improve our accuracy.²³⁻²⁴

4 Evaluation

To evaluate the effectiveness of the models, a comparison is made between the models on four different parameters: F1 - Scores, Precision, Recall and Accuracy.²⁵

Accuracy : The base metric used for model evaluation is often Accuracy, describing the number of correct predictions over all predictions:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Precision: Precision is defined as the number of correct positive predictions made.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Recall: Recall is the ratio of number of positive cases the model correctly predicted to the total number of positive cases in the dataset.

$$Recall = \frac{Tp}{Tp + FN} \quad (4)$$

F1 Score: F1-score is defined as the harmonic mean between Precision and Recall.

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (5)$$

Classifier	Accuracy	Precision	Recall	F1 Score
LR	92.09%	84.8%	92.09%	88.30%
KNN	91.71%	85.43%	91.71%	88.19%
CART	92.78%	92.56%	92.35%	92.44%
SVM	92.09%	84.8%	92.09%	88.30%
RF	94.49%	93.71%	94.74%	93.90%
Ensemble	94.74%	94.14%	94.66%	94.24%
XGBoost	95.00%	94.43%	95.00%	94.61%

Table 1: Performance comparison on LASI-DAD Data

Table 1 shows the performance of 7 classifiers: Logistic Regression (LR), K-Nearest Neighbor (KNN), Decision Tree Classifier (CART), Support Vector Machine (SVM), Random Forest (RF), XGBoost and Ensemble model on the LASI-DAD data. From the table, we can see that XGBoost has the highest performance in all the performance metrics with the highest accuracy of 95.58%. Ensemble Model and RF have the second and third-best performance, respectively, while CART, LR, KNN, and SVM have relatively lower performance.

Classifier	Accuracy	Precision	Recall	F1 Score
LR	88.65%	88.55%	88.64%	88.59%
KNN	79.66%	76.34%	79.68%	74.83%
CART	88.17%	87.84%	87.84%	87.84%
SVM	78.54%	61.72%	78.56%	69.13%
RF	92.01%	91.80%	92.00%	91.81%
Ensemble	91.20%	91.30%	91.52%	91.33%
XGBoost	91.36%	91.21%	91.36%	91.26%

Table 2: Performance comparison on HAALSI Data

Table 2 shows the performance of 7 classifiers: Logistic Regression (LR), K-Nearest Neighbor (KNN), Decision Tree Classifier (CART), Support Vector Machine (SVM), Random Forest (RF), XGBoost and Ensemble model on the HAALSI data. From the table, we can see that XGBoost has the highest performance in all the performance metrics with the highest accuracy of 95.00%. Ensemble Model and RF have the second and third-best performance, respectively, while CART, LR, KNN, and SVM have relatively lower performance.

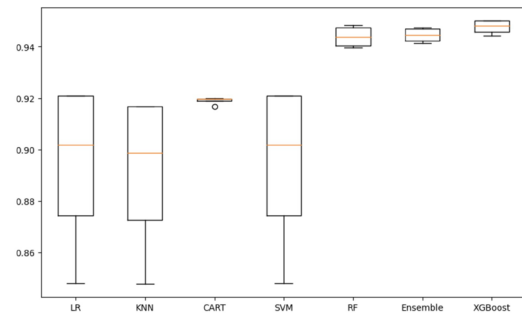


Figure 5: Boxplot comparing the accuracy, precision, recall, and F1-score of different machine learning algorithms on LASI-DAD data

In this specific plot 5, we have the accuracy scores of the following classifiers: LR, KNN, CART, SVM, RF and XGBoost . Each box represents the interquartile range (IQR) of the accuracy scores for a particular model, with the median represented as a horizontal line inside the box. The whiskers show the range of the data, with outliers shown as individual points.

In the plot shown in Figure 6, the accuracy scores of different classifiers are displayed. The classifiers included in the plot are LR (Logistic Regression), KNN (K-Nearest Neighbors), CART (Classifica-

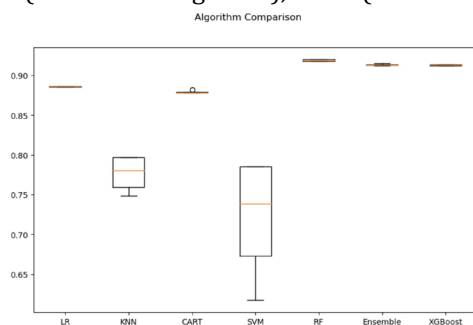


Figure 6: Boxplot comparing the accuracy, precision, recall, and F1-score of different machine learning algorithms on HAALSI data

tion and Regression Trees), SVM (Support Vector Machine), RF (Random Forest), and XGBoost.

The plots use a box-and-whisker representation to show the distribution of accuracy scores for each classifier.²⁶ The boxes represent the interquartile range (IQR) with the median indicated by a horizontal line. Whiskers show the data range, and outliers are displayed as individual points. Analyzing the box plots provide insights into the variation and central tendency of the accuracy scores for each classifier, facilitating comparisons and assessments of their performance on the given task.

We can conclude that the XGBoost model is the best performing model on both datasets based on the observations from the box plots. To further improve the performance of the XGBoost model, we performed hyperparameter tuning using GridSearchCV.²⁷ This technique helps us find the

optimal values for the model’s parameters from a given set of values in a grid. Figure 8 displays the optimal parameter values obtained after applying GridSearchCV on the LASIDAD dataset. Similarly, Figure 8 shows the optimal parameter values for the HAALSI dataset. After hyperparameter tuning, XGBoost model achieved a final accuracy of 95.99% on the LASIDAD dataset and 93.41% on the HAALSI dataset.

```
Best parameters: {'colsample_bytree': 1.0, 'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 2000, 'subsample': 1.0}
```

Figure 7: Optimal parameter values for LASI-DAD Dataset

```
Best parameters: {'colsample_bytree': 0.8, 'learning_rate': 0.01, 'max_depth': 7, 'n_estimators': 1000, 'subsample': 0.5}
```

Figure 8: Optimal parameter values for HAALSI Dataset

5 Results and implementation

The classification report for the XGBoost model used

	precision	recall	f1-score	support
0.0	0.97	0.98	0.98	724
1.0	0.61	0.57	0.59	44
2.0	1.00	0.50	0.67	4
3.0	0.00	0.00	0.00	1
accuracy			0.95	773
macro avg	0.65	0.51	0.56	773
weighted avg	0.95	0.95	0.95	773

Figure9:ClassificationReportonLASI-DADData

Classification Report:

	precision	recall	f1-score	support
0	0.93	0.99	0.96	142
1	0.97	0.72	0.83	40
accuracy			0.93	182
macro avg	0.95	0.86	0.89	182
weighted avg	0.94	0.93	0.93	182

Figure10:ClassificationReportonHAALSIData

The classification report for the LASI-DAD dataset shows that the model performs well for class 0, with high precision and recall. However, it struggles with class 1, exhibiting lower precision and recall. The accuracy of the model is 95.472%. On the other hand, the classification report for the HAALSIdataset indicates that the model has a precision of 93% for class 0 and 97% for class 1. The recall is 99% for class 0 and 72% for class 1. The overall accuracy is 93%.

As a result, these models can be used by clinicians to detect dementia in its earliest stage of development and provide appropriate treatments which might improve a patient's condition. Early detection of dementia can even help clinicians in improved decision-making, such as providing the right treatment and monitoring its progression over time.

Future Scope: This research methodology described above is a general approach that applies to all available data, without focusing on any specific datasets as stated down below.²⁸ cific patient characteristics.²⁹ To enhance the accuracy of the model, personalized models can be developed based on individual patient characteristics such as

age, sex, or medical conditions. By doing so, clinicians can intervene earlier and administer the most suitable treatment before the condition

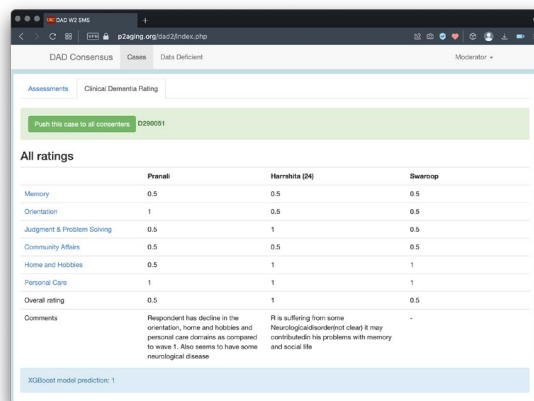
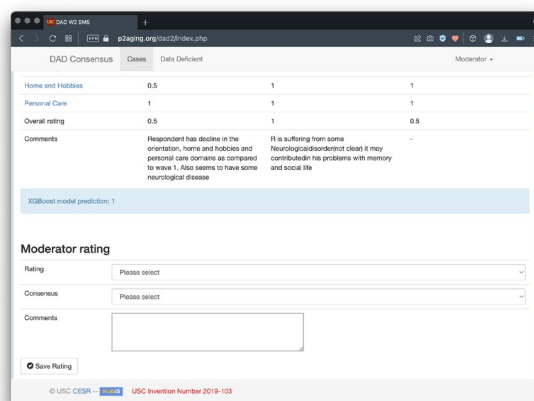


Figure11:LASI-DADOnlineConsensusSystemRaters



worsens, thereby improving patient outcomes.

Figure 12: LASI-DAD Online Consensus System Rating

These screenshots show the suggested CDR rating by the clinicians and the rating calculated by the model. The moderator can take this into consideration at the time of rating.

6 Conclusion

The integration of predictive models into cognitive testing for dementia has significant potential in improving the effectiveness and personalisation of the diagnostic and assessment process. Traditional cognitive tests often have a standardized format that may not fully capture an individual's unique cognitive abilities and the progression of cognitive decline. By leveraging predictive models, it becomes feasible to adapt and tailor the testing based on an individual's predicted risk of dementia.

The incorporation of predictive models provides valuable insights into an individual's likelihood of developing dementia, enabling clinicians to customize the testing approach accordingly. For individuals at a higher risk, more comprehensive and specific cognitive assessments can be administered, facilitating the early detection and monitoring of cognitive impairment. Conversely, for individuals at a lower risk, less intensive testing can be employed, optimizing resources and minimizing unnecessary evaluations.

Moreover, predictive models can play a crucial role in early detection and intervention strategies. By identifying individuals at a higher risk of dementia, proactive measures such as lifestyle modifications, cognitive training, and targeted interventions can be implemented. This proactive approach has the potential to delay or mitigate the onset and progression of dementia, enhancing overall outcomes and quality of life.

The integration of predictive models in cognitive testing not only improves diagnostic accuracy but also enables a more personalized and adaptive approach to cognitive assessment. By tailoring the testing process based on individual risk profiles, clinicians can optimize resources and provide targeted interventions. Ultimately, this integration contributes to advancements in dementia research,

care, and early intervention strategies, leading to improved outcomes for individuals at risk of dementia.

7 References

1. Mendez, M. (2022). General Mental Status Scales, Rating Instruments, and Behavior Inventories. In M. Mendez (Ed.), *The Mental Status Examination Handbook* (pp. 181-199). Elsevier. ISBN 9780323694896.
2. Geldmacher, D. S., & Whitehouse, P. J. (1996). Evaluation of Dementia. *New England Journal of Medicine*, 335(5), 330-336. doi:10.1056/NEJM199608013350507
3. Fagundes Chaves, M'arcia Lorena MD, PhD*; Camozzato, Ana Luiza MD, PhD*; Godinho, Cl'audia MD*; Kochhann, Renata*; Schuh, Artur*; de Almeida, Vanessa Lopes*; Kaye, Jeffrey MD†. Validity of the Clinical Dementia Rating Scale for the Detection and Staging of Dementia in Brazilian Patients. *Alzheimer Disease & Associated Disorders* 21(3):p 210-217, July 2007. — DOI: 10.1097/WAD.0b013e31811ff2b4
4. McCulla MM, Coats M, Van Fleet N, et al. Reliability of clinical nurse specialists in the staging of dementia. *Arch Neurol*. 1989;46:1210-1211.
5. Kavitha, C & Mani, Vinodhini & Srividhya, S & Khalaf, Osamah & Romero, Carlos. (2022). Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. *Frontiers in public health*. 10. 853294. 10.3389/fpubh.2022.853294.
6. Berkman, Lisa F. Health and Aging in Africa: A Longitudinal Study of an INDEPTH Community in South Africa [HAALSI]: Agincourt, South Africa, 2015-2022. Interuniversity Consortium for Political and Social Research [distributor], 2023-03-13. <https://doi.org/10.3886/ICPSR36633.v4>

7. Lee J, Dey AB. Introduction to LASI-DAD: The Longitudinal Aging Study in India Diagnostic Assessment of Dementia. *J Am Geriatr Soc.* 2020 Aug;68 Suppl 3(Suppl 3):S3S4. doi: 10.1111/jgs.16740. PMID: 32815600; PMCID: PMC7513796.
8. Deshmukh, Ratnadeep & Wangikar, Vaishali. (2011). Data Cleaning: Current Approaches and Issues.
9. Miriam Steurer, Robert J. Hill & Norbert Pfeifer (2021) Metrics for evaluating the performance of machine learning based automated valuation models, *Journal of Property Research*, 38:2, 99-129, DOI: 10.1080/09599916.2020.1858937
10. S. Yadav and S. Shukla, "Analysis of kFold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," 2016 IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, India, 2016, pp. 78-83, doi: 10.1109/IACC.2016.25.
11. Peng, Joanne & Lee, Kuk & Ingersoll, Gary. (2002). An Introduction to Logistic Regression Analysis and Reporting. *Journal of Educational Research - J EDUC RES.* 96. 3-14. 10.1080/00220670209598786.
12. Hosmer DW, Taber S, Lemeshow S. The importance of assessing the fit of logistic regression models: a case study. *Am J Public Health.* 1991 Dec;81(12):1630-5. doi: 10.2105/ajph.81.12.1630. PMID: 1746660; PMCID: PMC1405276.
13. Guo, Gongde & Wang, Hui & Bell, David & Bi, Yaxin. (2004). KNN Model-Based Approach in Classification.
14. K. Taunk, S. De, S. Verma and A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 1255-1260, doi: 10.1109/ICCS45141.2019.9065747.
15. Lewis, Roger. (2000). An Introduction to Classification and Regression Tree (CART) Analysis.
16. Farris, Frank. (2010). The Gini Index and Measures of Inequality. *American Mathematical Monthly.* 117. 851-864. 10.4169/000298910X523344.
17. Patel, Harsh & Prajapati, Purvi. (2018). Study and Analysis of Decision Tree Based Classification Algorithms. *International Journal of Computer Sciences and Engineering.* 6. 74-78. 10.26438/ijcse/v6i10.7478.
18. Zhang, Y. (2012). Support Vector Machine Classification Algorithm and Its Application. In: Liu, C., Wang, L., Yang, A. (eds) *Information Computing and Applications.* ICICA 2012. Communications in Computer and Information Science, vol 308. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3642-34041-3_27
19. Breiman, L. Random Forests. *Machine Learning* 45, 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
20. M. El Habib Daho, N. Settouti, M. El Amine Lazouni and M. El Amine Chikh, "Weighted vote for trees aggregation in Random Forest," 2014 International Conference on Multimedia Computing and Systems (ICMCS), Marrakech, Morocco, 2014, pp. 438-443, doi: 10.1109/ICMCS.2014.6911187.
21. Chen, Tianqi & Guestrin, Carlos. (2016). XGBoost: A Scalable Tree Boosting System. 785794. 10.1145/2939672.2939785.
22. Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurobotics*, 7, Article 21. doi:10.3389/fnbot.2013.00021
23. F. Huang, G. Xie and R. Xiao, "Research on Ensemble Learning," 2009 International Conference on Artificial Intelligence and Computational Intelligence, Shanghai, China, 2009, pp. 249-252, doi: 10.1109/AICI.2009.235.

24. Valentini, Giorgio & Masulli, Francesco. (2002). Ensembles of Learning Machines. Neural Nets WIRN Vietri-2002, Series Lecture Notes in Computer Sciences. 2486. 3-22. 10.1007/3-540-45808-5 1.
25. Miriam Steurer, Robert J. Hill & Norbert Pfeifer (2021) Metrics for evaluating the performance of machine learning based automated valuation models, *Journal of Property Research*, 38:2, 99-129, DOI: 10.1080/09599916.2020.1858937
26. Williamson, D & Parker, RA & Kendrick, Juliette. (1989). The box plot: A simple visual method to interpret data. *Annals of internal medicine*. 110. 916-21. 10.1059/0003-4819110-11-916.
27. Pirjatullah, D. Kartini, D. T. Nugrahadi, Muliadi and A. Farmadi, "Hyperparameter Tuning using GridsearchCV on The Comparison of The Activation Function of The ELM Method to The Classification of Pneumonia in Toddlers," 2021 4th International Conference of Computer and Informatics Engineering (IC2IE), Depok, Indonesia, 2021, pp. 390-395, doi: 10.1109/IC2IE53219.2021.9649207.
28. S. Chowdhury and M. P. Schoen, "Research Paper Classification using Supervised Machine Learning Techniques," 2020 Intermountain Engineering, Technology and Computing (IETC), Orem, UT, USA, 2020, pp. 1-6, doi: 10.1109/IETC47856.2020.9249211.
29. Li Y, Xiong C, Aschenbrenner AJ, Chang CH, Weiner MW, Nosheny RL, Mungas D, Bateman RJ, Hassenstab J, Moulder KL, Morris JC. Item response theory analysis of the Clinical Dementia Rating. *Alzheimers Dement*. 2021 Mar;17(3):534-542. doi: 10.1002/alz.12210. Epub 2020 Nov 20. PMID: 33215873; PMCID: PMC8016711.