11th International Conference on Higher Education Advances (HEAd'25)
Universitat Politècnica de València, València, 2025
DOI: https://doi.org/10.4995/HEAd25.2025.20273

Interpreting Decision Tree model using ELI5 for modelling student performance: Case of a structural analysis graduate attribute-based teaching course

Masikini Lugoma[1], Masengo Ilunga[2]
[1]Department of Mining Engineering & Geomatics, University of South Africa, South Africa [2]Department of Civil & Environmental Engineering, & Building Science, University of South Africa, South Africa.

How to cite: Lugoma, M.; Ilunga, M. (2025). Interpreting Decision Tree model using ELI5 for modelling student performance: Case of a structural analysis graduate attribute-based teaching course. In: 11th International Conference on Higher Education Advances (HEAd'25). Valencia, 17–20 June 2025. https://doi.org/10.4995/HEAd25.2025.20273

Abstract
Decision tree is used in this study as a classification machine learning to model year mark, using mainly formative assessments. Binary digits are used to classify the different marks, which illustrate performing versus non-performing students when considering four assessments (e.g. 1 to 4). A structural analysis course is used for the application of the decision tree model and is part of the Advanced Diploma programme of engineering technologists. This new programme is designed and developed on engineering competency known as graduate attributes. The results revealed higher accuracy, precision and F1-score during the testing period. Although decision tree model can be easily interpreted through its tree structure, the interpretability of the model can be enhanced through ELI5, which could reveal the ranking of assessments as influencing factors for the whole dataset (globally) and for each student performance prediction (locally), during testing. Hence, the decision tree algorithm was proven elegantly to be a suitable candidate for modelling the competence level of students and could guide teaching staff on student fate, in an academic environment.

Keywords: *Interpretability, decision tree, modelling, prediction, year mark, assessments.*

1. Introduction

Institutions of higher learning use passrate as an indicator of student performance or competence in each course, whereas the throughput rate measures the degree to which students exit a specific programme. In either case, assessments play a major role. Formative assessments can be instrumental in monitoring the progress of the student by assessing her/him on specific learning outcomes, while summative assessment can be administered to evaluate the student at the close of the course by determining whether the student can move to the next level. Assessment criteria are critical to define the parameters that guide students on what different tasks should be covered in assessments. Hence, this should be communicated transparently to students, from the inception of the course. For instance, inappropriate assessment settings and the lack of timely and no-detailed feedback to students may contribute to the low-grade point average (Daka, et al, 2021). For a specific course, the year mark is often an aggregate of the weighted assessments, which is ultimately used for deciding on the student competence in the course. Although the grade of the student can be affected by external factors such as anger, anxiety or tiredness, assessments are acknowledged to play a very important role for achieving learning outcomes (Matzavela & Alepis, 2021). Artificial Intelligence (AI) has been embraced in many real word applications. Within AI in particular, Machine learning (ML) techniques have been used for measuring student performance (Rastrollo-Guerrerom et al., 2020, Hashim et al., 2020). The issue of interpretability or explainability has been the focus in recent years due to the nature of ML, which is perceived as a black box. Attempts have been made on utilising eXplainable AI (XAI) tools to carry out model interpretability and feature importance for student performance (Wang & Luo, 2024; Islam et al., 2024). However, the literature on student performance based solely on assessments by using interpretability techniques is very rare. This current research deals preliminarily with decision tree model as a popular algorithm to establish a relationship between assessments and year mark, which can be binarised to capture the two categories "pass" or "fail". Moreover, ELI5 (Explain like I am a 5-year-old) is used as an enhancing aspect in the interpretability of student performance outcome from the decision tree. ELI5 fits most algorithms (Matzavela & Alepis, 2021), particularly optimising sparse logical models such as decision trees (Rudin et al., 2022), by presenting trustworthy, accountable and interpretable ML algorithms. This study focuses primarily on assessments of a specific course as inputs to ML since they are the only factors that contribute to the year mark of the student and ultimately, they are the only barometric variables to decide on student competence. The teaching module is assessed based on graduate attributes (GA) to decide on the student competence, particularly for new engineering technology programmes that have been introduced four years ago in the South African education. There should be a distinction between the weight of each assessment towards 100 % of the year mark and the weight derived from the interpretability of the decision tree algorithm, through ELI5. In what follows, course will mean teaching module. Algorithm, model, and technique will have synonymous meaning.

2. Decision Tree classifier

2.1. Fundamentals of a decision tree model

A decision tree algorithm deals with both supervised classification and regression problems for ML. For classification problems, decision tree uses a structural tree, where each attribute is represented by a node and the value of the attribute is represented by its branch, while the class is represented by the leaf (Sulistiani & Aldino, 2020). The root constitutes the top node of this decision tree. Figure 1 illustrates a simple structure of decision tree.

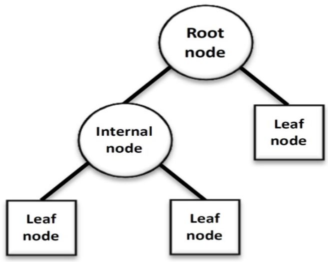

Figure 1. Simple form of decision tree model. Source: https://www.researchgate.net/figure/Basic-structure-of-a-decision-tree-All-decision-trees-are-built-through-recursion_fig3_295860754

A dependent variable is depicted by each leaf node for classification problems. The decision tree algorithm operates recursively, enabling the dataset to be partitioned into subsets, using the independent variable. For each iteration, the independent variable is selected if it has split the dataset into homogeneous subsets (in terms of the dependent variable). Conditions are set for the operations to be carried out. Usually, these conditions are for example maximum depth, or minimum number of samples for a given node. Entropy shows the purity of a subset of a system and helps to measure the split. It is given by the following expression

$$H = \sum_{i=1}^{n} p_i Log(p_i) \quad (1)$$

Where p_i: proportion of labels present in the subset and n: number of dependent variables the subset. The splitting process is associated with information gain and the parent node gives birth to a child node. In this way, child nodes whose information gain is maximised are searched by the decision tree. The information gain is given by

$$\mathrm{G} = \mathrm{H_P} - \sum_{\mathrm{i}=1}^{\mathrm{q}} \beta_{\mathrm{i}}(\mathrm{Hc_i}) \quad (2)$$

Where H_P: entropy for the parent node; β_j: fraction of points in the child node concerning points in the parent node, q: number of child nodes and Hc_i: entropy for each child node.

The Gini index called the Gini impurity, which should be low for a better split, is given by Equation (3) below.

$$\mathrm{Gini} = 1 - \sum_{\mathrm{i}=1}^{\mathrm{q}} (p_i)^2 \quad (3)$$

2.2. Metrics for model performance assessment

The model performance of the classification can be done both for training and testing parts. Most importantly, the interest is more on the testing part to measure the prediction capability of the ML algorithm. Therefore, the usual metrics to assess the model performance for classification problems are accuracy, recall, precision and f1-score (e.g. Ouedraogo, 2021; Prabaswara et al., 2023). These metrics are derived from the confusion matrix,

$$Confusion\,matrix = \begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix} \tag{4}$$

Where TP: True Positive assignments associated with positive classes which are correctly predicted; TN: True Negative assignments associated with negative classes which are correctly predicted, FP: False Positive assignments associated with positive classes which are incorrectly predicted; and FN: False negatives associated with negative classes which are incorrectly predicted. Therefore, the performance metrics are given by the following equations:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{5}$$

$$Recall = \frac{TP}{TP+FP} \tag{6}$$

$$Precision = \frac{TP}{TP+FN} \tag{7}$$

$$F1 - score = 2\,\frac{Precision \times Recall}{Precision + Recal} \tag{8}$$

In the above equations, the values of TP, TN should be far higher than FP and FN for good model performance, which is associated with high values of these metrics. It should be noted that F1-score in Equation (8) also called balanced F-score, is defined by the harmonic mean value derived from the Precision and the Recall.

2.3. Interpretability and explainability of machine learning

The black box nature of ML has been attributed to the lack of understanding how ML arrives at its conclusion during modelling process (Ouedraogo, 2021). Interpretation capability of ML may ensure unbiasedness during modelling process; by providing trust and transparency, thus reducing model complexity (Ouedraogo, 2021). Trust is not necessarily created by using explainable models; however, it is possible that distrust could be possible, hence they enable the model users to make a decision of trust or not (Rudin et al., 2022). When the inner operations of systems are intelligible, interpretable systems are said to be explainable, hence understandable by a human (Yildirim et al., 2024). In the literature, interpretable machine learning seems to be popular, therefore this study will use quite often interpretability as opposed to explainability. Generally, interpretability and explainability of ML fall under the category of the eXplainable AI (XAI) field, which is widely recognised to be very important for the AI models deployment (Arrieta et al., 2020). There are various interpretable models; for example, the ELI5 XAI was used in conjunction with LIME (Local Interpretable Model agnostic Explanations) and SHAP (SHapley Additive exPlanations) algorithmic frameworks (Vishwarupe et al., 2022). It is reiterated that this preliminary study focuses on ELI5 since it is one of the simplest and popular interpretability metrics. ELI5 can be used as local based explanation method, with the characteristics of approximating a complex model locally to a linear model, where the output is close to a regression model, which includes coefficients and bias (Ouedraogo, 2021). ELI5 as other methods can fit to both global interpretation where model's parameters are observed for the purpose of pointing out how the model works at global and local interpretation, where features leading to a specific single prediction are identified (Huilgol, 2020). In applications other than student performance, it is noted that decision tree model has been interpreted with ELI5 (Kawakura et al., 2022; Shivaprasad et al., 2024).

3. Data availability and methods

3.1. Data availability

Structural Analysis 3 course is taught at Advanced Diploma level. The course constitutes 120 notional hours, which is equivalent to 12 credits. This module is offered by the Department of Civil and Environmental Engineering, and Building Science, at the University of South Africa. Graduate attribute 1 covers essentially aspects to develop student's capability for identifying, formulating, analysing and solving broadly defined engineering problems. For a student to be declared competent in this specific GA, the student should score at least 50% for the year mark. Additionally, the student should achieve at least 50% for assessment 2, which is a project assignment in the form of a portfolio. Besides, assessments 1, 3 and 4 were considered. All assessments counted 30% each, except assessment 1, which weighed 10%. The number of attributes used in the ML were 146 for the year 2023 semester 2 and were complete, hence did not contain missing values.

3.2. Methods

The attributes were composed of independent variables (features) and the dependent variable namely (label). These attributes were used in the model training and testing of the decision tree. Independent variables were the four assessments, and the dependent variable (year mark) was translated in a binary class. The binary digit was a conversion of a "fail" or "pass" final score used for student performance in the module under investigation. This conversion enabled easy learning of the model. The dataset was split into the ratio 8:2 for training and testing respectively. The decision tree algorithm was implemented elegantly using scikit-learn library in Python. Google Collaboratory programming environment was used to implement the algorithm. Hence, the DecisionTreeClassifier() function was used during training. The randomised state was set to 42 to ensure the results were reproducible. The performance metrics were then used to evaluate the model. The model interpretation was globally and locally done by using the following code: eli.explain_weights(classifier_dtc) and eli.show_prediction(classifier_dtc, X_test.iloc[1], feature_names=list(X.columns), show_feature_values=True)

4. Results and discussion

Figure 2 below shows the degree of association among the features, which was assessed to avoid any duplication of information for strongly correlated features. The features were noted as Ass1, 2, 3 and 4 to describe assessments 1, 2, 3 and 4 respectively. Generally, the correlation values were shown relatively weak. Hence, none of the features could be left out.

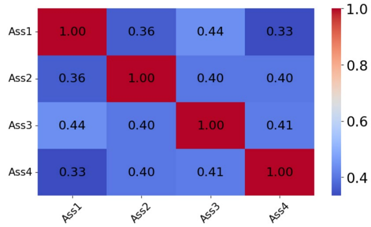

Figure 2. Pearson correlation values between assignment and year mark.

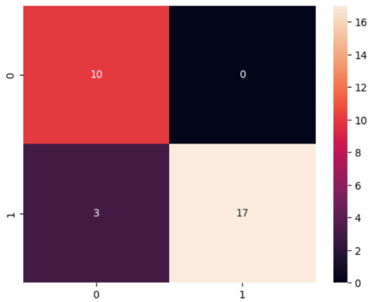

Figure 3. Confusion matrix emanating from the student performance in the structural analysis graduate attribute-based module.

From the confusion matric displayed in Figure 3, FN and FP were 3 and 0 respectively whereas the TN and TP were 10 and 17 respectively. The sum of TP and TN outweighs the sum of their counterparts FN and FP. Hence, the model was believed to perform well. However, the prediction of TP dominated. This could justify the use of decision tree model in predicting student performance from assessments. The Accuracy, Precision, Recall and F1-score of the model for structural analysis course were 0.9, 0.92, 0.90 and 0.9 respectively. Thus, the model demonstrated good performance in identifying performing students versus non-performing students, as shown in Figure 4. In the tree model Ass1, 2, 3 and 4 are equivalent to x0, x1, x2 and x3 respectively. In this figure, the decision model with its root at the top, is depicted with all nodes, leaves and branches and shows transparently the decision rule during the splitting process. If a mark for Ass2(x1) < 46.5 is true, the splitting process is good at 9.5% with a high degree of purity and the 90.5% undergoes a check on Ass3 = x2 and so on. Hyperparameter tuning led to a maximum depth of 5 for the model performance.

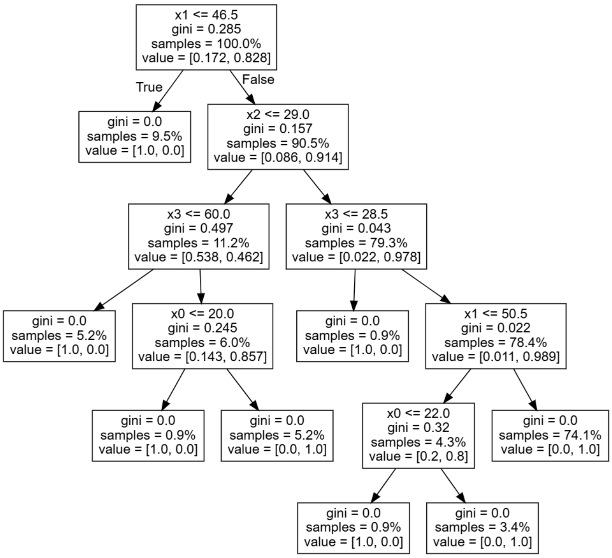

Figure 4. Decision tree model for the structural graduate attribute-based course.

Figure 5 displays the permutation importance by visualising feature weights in the decision tree model and their level of contribution as a way of global interpretability derived from ELI5 library. In the descending order, the feature importance visualization is Ass2 = x1, Ass4 = x3, Ass3 = x2 and Ass1 = x0. Therefore, academics should pay attention primarily to the project, and the 2 tests. These assessments have 30% in weight for contribution to the year mark. Furthermore, model inspection on the prediction level is supported, which uses similar visualization with weights adding up to either probability of a class in classification models or predicted value in case of regression models (Gashi et al., 2022). In Figure 6, ELI5 shows only the prediction of the first record in the training dataset (local interpretability). Hence, using this interpretability technique, the top prediction can be easily explained and a good fashion of visualising the feature impacts plus a bias term is displayed. The gradient of green and red color indicates the positive and negative impact on the model decisions and no interactive options exist (Gashi et al, 2022). For the local level at the first two features, Ass1 and Ass3 have a strong impact on the first student (record). ELI5 is not a fully flagged model-agnostic explanations technique, but it is seen as an enhancer of the tree-based model. This situation may improve trust in the final score computation and decision-making about the student's competence in a specific course. This is of great importance in the educational system that requires transparency and insights into prediction results to set effective strategic measures (Tong & Li, 2025).

Weight	Feature
0.5143	x1
0.2019	x3
0.1837	x2
0.1001	x0

Figure 5. Global interpretability for the structural graduate attribute-based course.

y=0 (probability **1.000**) top features

Contribution?	Feature	Value
+0.857	Ass1	12.000
+0.453	Ass3	0.000
+0.172	<BIAS>	1.000
-0.087	Ass2	71.000
-0.396	Ass4	87.000

Figure 6. Local interpretability for the structural graduate attribute-based course.

5. Conclusion

The decision tree machine learning has been evaluated preliminarily and showed good accuracy for students' performance, based on graded continuous assessment tasks. The ML learning hereto referred has demonstrated its capability to predict student competence for a structural analysis module taught in engineering. Although the model displayed the complex characteristic of the relationship between the independent variables and the dependent variable in the form of a binary class, the use of ELI5 XAI technique enabled local interpretability, by clarifying transparency the prediction of the final score of each student. It also displayed the impacts of each assessment on the overall model performance, hence helped for global interpretability. The versatility of the decision tree method to deal with various problems, including student performance problems, does not mean an end on its own, but requires accountability and transparency in the modeling outputs. This preliminary study has demonstrated that course leaders, may enhance their decision on student performance, by focusing on assessments that may contribute impactfully on the students' fate. In this way ML can be applied in strengthening best practices of teaching and learning to assist struggling students, who are not successful in the graduate attribute assessed courses. Thus, compelling attention could be drawn to the students who were declared not competent. These preliminary results present some limitations due to the dataset size and ELI5 has been applied as a monolithic interpretability method. A re-evaluation of the decision tree model could be done as the dataset increases over time. It is suggested that future research could include more ML algorithms other that decision tree. Hence, interpretability methods other than ELI5 could be explored.

References

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, *58*, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Daka, H., Mulenga-Hagane, M. L., Mukalula-Kalumbi, M., & Lisulo, S. (2021). Making summative assessment effective. *European Modern Studies Journal*, *5*(4), 224-237.

Gashi, M., Vuković, M., Jekic, N., Thalmann, S., Holzinger, A., Jean-Quartier, C., & Jeanquartier, F. (2022). State-of-the-art explainability methods with focus on visual analytics showcased by glioma classification. *BioMedInformatics*, *2*(1), 139-158. https://doi.org/10.3390/biomedinformatics2010009

Islam, M. R., Nitu, A. M., Marjan, M. A., Uddin, M. P., Afjal, M. I., & Mamun, M. A. A. (2024). Enhancing tertiary students' programming skills with an explainable Educational Data Mining approach. *PloS one*, *19*(9), e0307536. https://doi.org/10.1371/journal.pone.0307536

Hashim, A.S. Awadh, W.A. & Hamoud, A. K. (2020) Student performance prediction model based on supervised machine learning algorithms. *In IOP conference series: materials science and engineer*ing 928(3), p. 032019). IOP Publishing.

Huilgol, P. (2020). 6 Python Libraries to Interpret Machine Learning Models and Build Trust https://www.analyticsvidhya.com/blog/2020/03/6-python-libraries-interpret-machine-learning-models/

Kawakura, S., Hirafuji, M., Ninomiya, S., & Shibasaki, R. (2022). Adaptations of explainable artificial intelligence (XAI) to agricultural data models with ELI5, PDPbox, and skater using diverse agricultural worker data. *European Journal of Artificial Intelligence and Machine Learning*, *1*(3), 27-34

Matzavela, V., & Alepis, E. (2021). Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments. *Computers and Education: Artificial Intelligence*, *2*, 100035.

Ouedraogo, D. N. (2021). Interpretable Machine Learning Model Selection for Breast Cancer Diagnosis Based on K-means Clustering. *Applied Medical Informatics*, *43*(3), 91-102.

Prabaswara, R., Lemantara, J., & Jusak, J. (2023). Classification of Secondary School Destination for Inclusive Students using Decision Tree Algorithm. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *7*(5), 1009-1018. https://doi.org/10.29207/resti.v7i5.5081

Rastrollo-Guerrero, J.L., Gómez-Pulido, J.A. & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review, *Applied Sciences*, *10*(3), 1042. doi:10.3390/app10031042

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistic Surveys*, *16*, 1-85. https://doi.org/10.1214/21-SS133.

Shivaprasad, S., Chadaga, K., Dias, C. C., Sampathila, N., & Prabhu, S. (2024). An interpretable schizophrenia diagnosis framework using machine learning and explainable artificial intelligence. *Systems Science & Control Engineering*, *12*(1). https://doi.org/10.1080/21642583.2024.2364033

Sulistiani, H.; Aldino, A.A. (2020). Decision tree c4.5 algorithm for tuition aid grant program classification (case study: department of information system, universitas teknokrat Indonesia). Jurnal Ilmiah Edutic/Vol.7, No.1, November 2020, 40-50. doi:10.21107/edutic.v7i1.8849

Tong, T., & Li, Z. (2025). Predicting learning achievement using ensemble learning with result explanation. *PloS one*, *20*(1), e0312124. https://doi.org/10.1371/journal.pone.0312124

Vishwarupe, V., Joshi, P. M., Mathias, N., Maheshwari, S., Mhaisalkar, S., & Pawar, V. (2022). Explainable AI and interpretable machine learning: A case study in perspective. *Procedia Computer Science*, *204*, 869-876. 10.1016/j.procs.2022.08.105

Wang, S., & Luo, B. (2024). Academic achievement prediction in higher education through interpretable modeling. *Plos one*, *19*(9), e0309838. https://doi.org/10.1371/journal.pone.0309838

Yildirim, M., Okay, F. Y., & Özdemir, S. (2024). A comparative analysis on the reliability of interpretable machine learning. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, *30*(4), 494-508. doi: 10.5505/pajes.2023.49473