JOIV : Int. J. Inform. Visualization, 8(2) - May 2024 652-662

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Exploring Classification Algorithms for Detecting Learning Loss in Islamic Religious Education: A Comparative Study

Rohmat Mulyana Sapdi [a,*], Dian Sa'adillah Maylawati [b], Diena Rauda Ramdania [b], Ichsan Budiman [b], Muhammad Insan Al-Amin [b], Mi'raj Fuadi [c]

[a] *Department of Islamic Education, UIN Sunan Gunung Djati Bandung, Cibiru, Bandung, 40614, Indonesia*
[b] *Department of Informatics, UIN Sunan Gunung Djati Bandung, Cibiru, Bandung, 40614, Indonesia*
[c] *Agricultural Engineering, Universitas Mataram, Selaparang, Mataram, 83115, Indonesia*

Corresponding author: [*]rohmat.mulyana@uinsgd.ac.id

***Abstract*— This study investigates the detection of learning loss in Islamic religious education subjects in Indonesia. Focusing on the effectiveness of multiple classification algorithms, the research assesses learning loss across literacy, numeracy, writing, and science domains. While education traditionally involves knowledge transmission, it also seeks to instill values. Given Indonesia's predominantly Islamic demographic, Islamic Religious Education (IRE) is pivotal in disseminating moral and cultural values, encompassing teachings from the Koran, Hadith, Aqedah, morality, Fiqh, and Islamic history. The study's central aim is to discern learning loss in IRE in Islamic schools, utilizing the Gradient Boosting Classifier as its primary analytical tool. Various classification algorithms, including the Cat Boost Classifier, Light Gradient Boosting Machine, Extreme Gradient Boosting, and others, were tested. The study engaged a sample of 38,326 Islamic Elementary school students, 29,350 Islamic Junior High school students, and 13,474 Islamic High school students across Indonesia. The findings revealed that the Light Gradient Boosting Machine was the most effective model for Islamic Elementary and High school data, while the Cat Boost Classifier excelled for Islamic Junior High school data. These results highlight the extent of learning loss in IRE and offer invaluable perspectives for education stakeholders. Future studies are encouraged to further explore the root causes of this learning loss and devise specific interventions to tackle these issues effectively.**

***Keywords*—Education; gradient boosting classifier; Islamic religious education; learning loss; values.**

Manuscript received 17 May. 2023; revised 10 Sep. 2023; accepted 23 Oct. 2023. Date of publication 31 May. 2024.

I. Introduction

Technology can be a medium for the transfer of knowledge. Still, more is needed to transfer values and skills, requiring direct face-to-face communication and practical approaches. The condition of facilities gaps and technology use is a big challenge for the world of education during the COVID-19 pandemic. This fact became one of the triggers for the emergence of learning loss conditions during the COVID-19 pandemic. Learning loss is a condition where knowledge and skills are lost in general and precisely due to a decline in the processed academic [1], [2]. Learning loss is caused by extended holiday periods, students dropping out of school, and closing face-to-face schools due to critical or urgent conditions such as a pandemic. Learning loss in various fields of education due to the COVID-19 pandemic has become a global educational problem [3]–[10].

Based on research released by INOVASI & the Center for Standards and Education Policy, Ministry of Education, Culture, Research, and Technology, from 18,368 elementary school students from grade 1 to grade 3 in Indonesia, there is an indication of a learning loss of 40% for literacy and 56% for numeracy during the COVID-19 pandemic [11]. This condition needs to be taken seriously. Learning difficulties such as literacy and numeracy at the elementary school level due to learning loss permanently impact the nation's next generation, especially at the youth level [12]–[18].

The Ministry of Religion, through the Directorate General of Islamic Education, considers it essential to overcome the learning loss that has occurred in the world of education, especially in an Islamic state school, because of the COVID-19 pandemic. Within the framework of Islamic Education, to achieve well-being in 2030, compiled by the Organization for Economic Co-operation and Development (OECD) [19], it is necessary to add religious literacy skills to students. Religious literacy can be obtained early through Islamic Religious Education (IRE). IRE is a step and pattern of guidance that provides knowledge and shapes students' personalities, attitudes, and skills in practicing Islamic religious teachings and values in daily life [20]–[22]. Moreover, IRE is in line with the mandate of the Law of the Republic of Indonesia Number 20 of 2003 concerning the National Education System and the function of education in the transfer of value. Therefore, the impact of learning loss during the COVID-19 pandemic on Islamic Religious Education (IRE) needs to be identified and appropriate mitigation carried out.

Indonesia has 9,681,284 Islamic School students from all levels and classes (Kindergarten, Elementary, Junior High, and High) spread across 83,551 institutions in the Odd semester of 2020/2021 [23]. This big data can be processed through a data science approach with machine learning technology. IRE learning loss conditions in Islamic School students can be detected to give mitigation recommendations for student needs. Data science is a modern technique that collaborates statistics by utilizing machine learning technology and artificial intelligence to find interesting patterns, insight knowledge, or meaningful information from big data to help make business decisions [24]–[28].

Various machine learning approaches can be used in data science according to business needs and targets to be produced, one of which is the semi-supervised learning approach. Semi-supervised learning is a machine learning approach that combines supervised and unsupervised learning approaches [29]–[33]. The unsupervised model is used to look at data groups that indicate the presence of IRE learning loss for further use as a reference for labeling data. Meanwhile, the supervised learning approach is used to build a learning loss detection model for IRE in Islamic Schools. This study aims to find the best model using the Gradient Boosting Classifier.

II. Material and Method

Classification is a widely used supervised learning machine learning approach [34]–[40] that aims to categorize data into predefined classes or categories. This study employs various classification algorithms in the model-building process to detect learning loss in IRE within Islamic schools. The following classification algorithms are utilized:

1. CatBoost Classifier.
2. Light Gradient Boosting Machine (LightGBM).
3. Extreme Gradient Boosting (XGBoost).
4. Gradient Boosting Classifier.
5. Random Forest Classifier.
6. Ada Boost Classifier.
7. Extra Trees Classifier.
8. Logistic Regression.
9. Ridge Classifier.
10. Linear Discriminant Analysis.
11. Decision Tree Classifier.
12. K-Neighbours Classifier.
13. Naive Bayes.
14. SVM - Linear Kernel.
15. Dummy Classifier.
16. Quadratic Discriminant Analysis.

III. Results and Discussion

In the initial stage, the reading of the labeled dataset is carried out. Fig 1 shows the process. Fig 2 is the determination of the training data configuration. Its purpose is to configure the training process and training flow transformation functions. The parameters needed are the dataset and labels. The description shows the configuration name of the experiment being performed, and the value shows the value of the configuration.

```
dataset = pd.read_csv('./mi_num_labeled.csv')

shuffled_dataset = dataset.iloc[:, [2,3,4,5,6,7,8,9,10,11,12,13,14,15,21]].sample(frac=1)

shuffled_dataset
```

	G01Q01	G01Q02	G01Q04	G02Q03	G02Q05	G02Q06	G02Q07	G02Q08	G02Q09	G02Q10	G02Q11	G01Q33	G03Q12	G03Q14	clusters
32898	2	19	1	1	3	1	1	1	1	2	1	2	1	1	0
19557	1	6	1	1	3	2	1	1	1	2	1	1	1	1	0
15890	1	4	1	1	2	1	1	1	1	2	1	2	1	1	1
30406	1	34	1	1	1	2	1	1	1	2	1	2	1	1	0
25219	2	17	2	1	1	1	1	1	1	2	1	1	1	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
28368	2	6	1	1	2	3	2	2	1	2	1	2	1	1	0
34803	2	2	1	1	1	1	1	1	1	1	1	2	1	1	0
19629	1	2	1	1	3	1	1	1	1	2	1	2	1	1	0
20215	1	4	1	1	3	1	1	1	1	2	1	2	1	1	0
964	1	6	1	1	3	2	1	1	1	2	1	2	1	1	1

38326 rows × 15 columns

Fig. 1 The process of reading a labeled dataset

```
s = setup(shuffled_dataset, target = 'clusters')
```

	Description	Value
0	session_id	6962
1	Target	clusters
2	Target Type	Binary
3	Label Encoded	None
4	Original Data	(38326, 15)
5	Missing Values	False
6	Numeric Features	1
7	Categorical Features	13
8	Ordinal Features	False
9	High Cardinality Features	False
10	High Cardinality Method	None
11	Transformed Train Set	(26828, 24)
12	Transformed Test Set	(11498, 24)
13	Shuffle Train-Test	True
14	Stratify Train-Test	False
15	Fold Generator	StratifiedKFold
16	Fold Number	10
17	CPU Jobs	-1
18	Use GPU	False
19	Log Experiment	False
20	Experiment Name	clf-default-name
21	USI	81ab
22	Imputation Type	simple
23	Iterative Imputation Iteration	None
24	Numeric Imputer	mean
25	Iterative Imputation Numeric Model	None
26	Categorical Imputer	constant
27	Iterative Imputation Categorical Model	None
28	Unknown Categoricals Handling	least_frequent

Fig. 2 The training configuration process

The following process is training and evaluating various classification models through predetermined configurations. This training process is carried out using the cross-validation method, where the output is the performance matrix of each model. Then a comparison of the performance of each model because of the training is carried out. Fig. 2 is a display of the configuration of a data analysis process. This process is conducted on a dataset with a binary "clusters" target. The original dataset has a size of (38326, 15) with no missing values and consists of 1 numeric feature and 13 categorical features. The data division for training and testing has been randomized. Additionally, various other parameters such as the type of imputation, the number of cross-validation folds, and GPU usage are also shown. The following are the results of the classification and evaluation of models for each level of education, starting from Islamic elementary, Junior High, and High school.

A. Classification Model for Islamic Elementary School Levels

Based on the model comparison results in Table 1, the model with the best Accuracy for Islamic Elementary School data is the Light Gradient Boosting Machine, with an accuracy value of 0.6102. The model evaluation process is described through the following processes: learning curve, validation curve, confusion matrix, AUC-ROC curve, Prediction error, Classification report, Precision, Recall, and Threshold.

TABLE I
MODEL COMPARISON RESULT FOR ISLAMIC ELEMENTARY SCHOOL DATA

Model	Accuracy	AUC	Recall	Proc.	F1	Kappa	MCC	TT (Sec)
Lightgbm	0.6102	0.6527	0.6842	0.5952	0.6366	0.2207	0.2232	0.0940
Catboost	0.6075	0.6518	0.6790	0.5932	0.6332	0.2153	0.2177	10.0130
xgboost	0.6060	0.6485	0.6840	0.5909	0.6339	0.2123	0.2150	1.2540
Gbc	0.6025	0.6416	0.6641	0.5904	0.6250	0.2052	0.2069	0.3990
Rf	0.5936	0.6239	0.6528	0.5828	0.6157	0.1875	0.1889	0.6080
Ada	0.5932	0.6307	0.6489	0.5831	0.6141	0.1866	0.1878	0.1630
Lr	0.5861	0.6245	0.6506	0.5755	0.6107	0.1725	0.1741	0.4030
Lda	0.5861	0.6245	0.6523	0.5752	0.6113	0.1724	0.1740	0.0550
Ridge	0.5860	0.0000	0.6522	0.5751	0.6112	0.1722	0.1738	0.0240
Et	0.5853	0.6130	0.6150	0.5795	0.5967	0.1707	0.1710	0.6150
Dt	0.5813	0.5958	0.6145	0.5753	0.5942	0.1626	0.1631	0.0430
Knn	0.5742	0.5990	0.6216	0.5668	0.5929	0.1485	0.1493	0.5300
Svm	0.5665	0.0000	0.7110	0.5628	0.6100	0.1336	0.1530	0.2410
Nb	0.5516	0.6094	0.8713	0.5309	0.6598	0.1045	0.1356	0.0270
Qda	0.5078	0.5078	0.7097	0.5071	0.5709	0.0164	0.0257	0.0330
Dummy	0.5011	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0180

Description:
1. **Lightgbm:** Light Gradient Boosting Machine
2. **Catboost:** CatBoost Classifier
3. **Xgboost**: Extreme Gradient Boosting
4. **Gbc:** Gradient Boosting Classifier
5. **Rf:** Random Forest Classifier
6. **Ada:** Ada Boost Classifier
7. **Lr:** Logistic Regression
8. **Lda:** Linear Discriminant Analysis
9. **Ridge:** Ridge Classifier
10. **Et:** Extra Trees Classifier
11. **Dt:** Decision Tree Classifier
12. **Knn:** K Neighbours Classifier
13. **SVM:** SVM - Linear Kernel
14. **Nb:** Naive Bayes
15. **Qda:** Quadratic Discriminant Analysis
16. **Dummy:** Dummy Classifier

Fig. 3 shows the learning curve's result, which shows the effect of adding more samples during the training process. The effect is illustrated by examining the statistical performance of the model in terms of training and validation scores. The validation curve is an essential diagnostic tool that shows the sensitivity between changes in the Accuracy of a Machine Learning model and changes in some model parameters. The function of the validation curve is to evaluate existing models based on hyperparameters. On the validation curve, there are training scores and cross-validation scores. The model will likely be a poor fit if both scores are low. This means the model must be more complex and informed by more features. If the training curve reaches a high score relatively quickly and the validation curve lags, then the model is overfitting. This condition indicates that the model is very complex, and the data is too small or can also mean too little data. The ideal condition is when the parameter values where the training and validation curves are closest to each other.

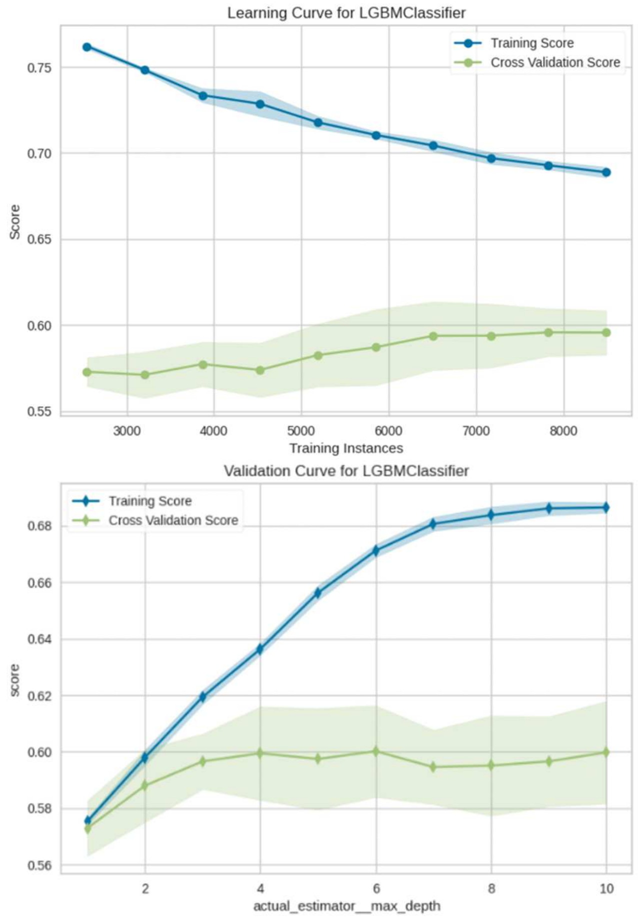

Fig. 3 LGBM learning curve and validation curve for Islamic Elementary school data

The two graphs on Fig 3 display evaluation curves for the LGBM Classifier. The first graph is a validation curve that shows the comparison between the training score and the cross-validation score against the maximum depth of the estimator (`actual_estimator_max_depth`). It can be observed that as the depth increases, the training score rises while the cross-validation score tends to stabilize after reaching a certain point. The second graph is a learning curve that illustrates the model's performance based on the number of training instances. As the number of instances increases, the training score gradually decreases, while the cross-validation score increases until both converge at a certain point, indicating that the model has reached a stable point in its learning.

The Confusion Matrix is a matrix that presents a summary of all the prediction results generated by comparing the predicted results with the expected results. Based on the Confusion Matrix, the following matrices can be identified as follows:

- Precision: the ratio of the number of items identified as positive correctly to the number identified as positive.
- Recall: Comparison of the number of relevant items identified correctly with all correct items.
- Accuracy: comparison of the number of items predicted correctly and incorrectly with the total of all predictions made.

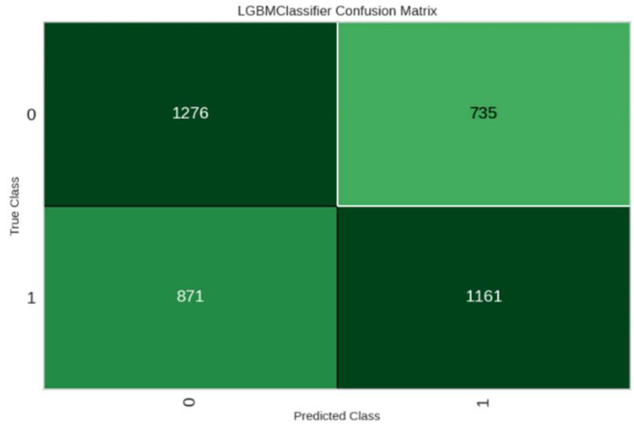

Fig. 4 Confusion Matrix result

Fig 4 shows that the Precision value: is 0,6435, the Recall value: is 0,5943, and the Accuracy is 0,602. Fig 5 shows the ROC (Receiver Operating Characteristic) curve for the LGBM Classifier. The ROC curve is used to evaluate the classification performance of a model across all classification thresholds. The AUC (Area Under the Curve) value reflects how well the model differentiates between positive and negative classes. In this case, the AUC values for class 0, class 1, the micro-average ROC curve, and the macro-average ROC curve are all 0.64. This indicates that the model has a moderate discriminative ability. The closer the AUC value is to 1, the better the model's performance in distinguishing between positive and negative classes.

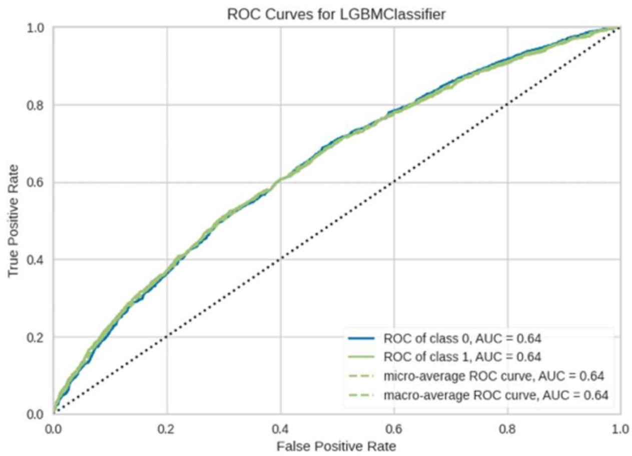

Fig. 5 AUC-ROC result

The AUC-ROC curve performs the classification problem at various threshold settings. ROC is a probability curve, and AUC represents the degree or measure of separateness. It provides information on how much the model can differentiate between classes. The higher the AUC, the better the model predicts 0 as 0 and 1 as 1. An excellent model has an AUC close to 1, meaning it has a good separability measure. Conversely, the closer to 0, the model has the worst separability measure. The AUC result obtained in Islamic Elementary School modeling is 0.64. This result indicates that the model has a good measure of separability (Fig 5).

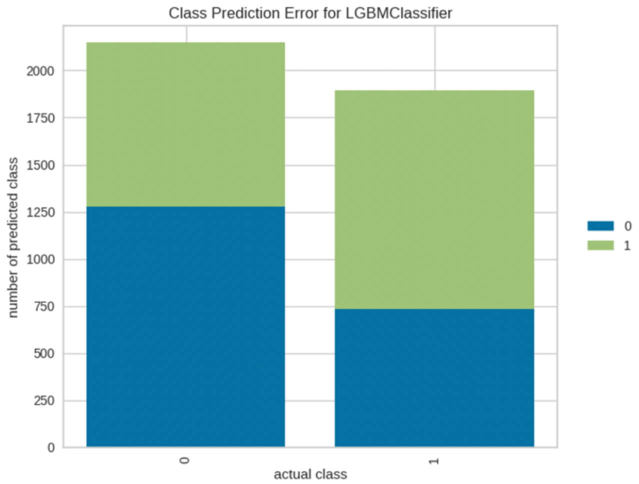

Fig. 6 Prediction error for Islamic Elementary School data

Fig. 6 illustrates the class prediction errors for the LGBM Classifier model. The vertical axis displays the number of predicted classes, while the horizontal axis represents the actual classes. For the actual class '0', most are correctly predicted as '0' (the blue section), but some are mispredicted as class '1' (the upper green section). Conversely, for the actual class '1', many are accurately predicted as '1', but a portion is mispredicted as class '0' (the lower blue section). This highlights the accuracy and error levels of the model's predictions for each class.

Fig 7 is a heatmap detailing performance metrics for two classes: '0' and '1'. For class '1', the Precision, Recall, F1-score, and Support are 0.612, 0.571, 0.591, and 2032, respectively. Conversely, for class '0', these values are 0.594, 0.635, 0.614, and 2011. Darker shades in the heatmap correspond to higher metric values, indicating better performance.

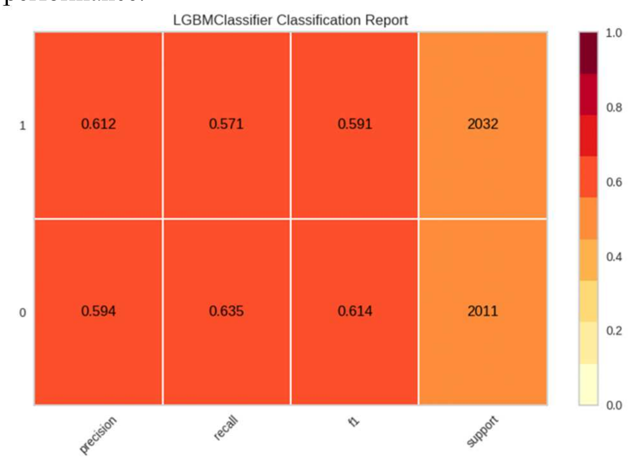

Fig. 7 Classification report data result

The graph on Fig. 8 depicts the threshold plot for the LGBM Classifier model, showcasing how different metrics (precision, recall, f1, and queue rate) change as the discrimination threshold is adjusted. The vertical axis represents the score of each metric, ranging from 0 to 1, while the horizontal axis indicates the discrimination threshold, also from 0 to 1. The dashed line marked "t=0.30" highlights a specific threshold value. Notably, as the threshold increases, precision rises while recall drops. The f1 score, which is the harmonic mean of precision and recall, peaks around the 0.3 threshold. The queue rate, represented by the shaded area, decreases as the threshold approaches 1, emphasizing the trade-offs between these metrics at different threshold levels.

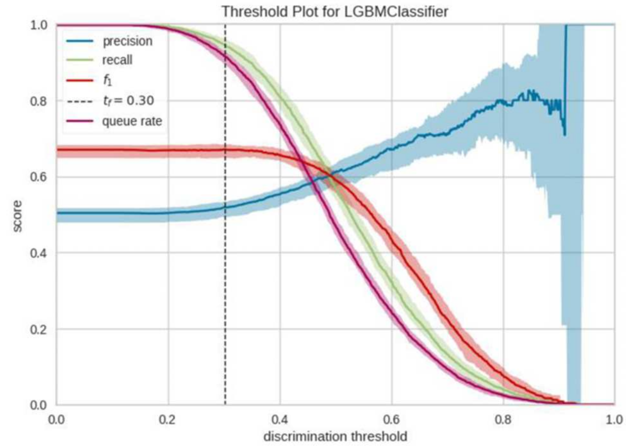

Fig. 8 Threshold Plot of LGBM Islamic Elementary School Model

Fig. 9 presents a "Feature Importance Plot" which visualizes the significance of various features in a predictive model. The vertical axis lists the feature names, while the horizontal axis quantifies their importance, ranging from 0 to roughly 1000. Each dot corresponds to the importance score of a particular feature. Notably, the feature "G01Q02" stands out as having the highest importance score, far surpassing the other features. This indicates that "G01Q02" plays a more critical role in the model's predictions compared to the other listed features.

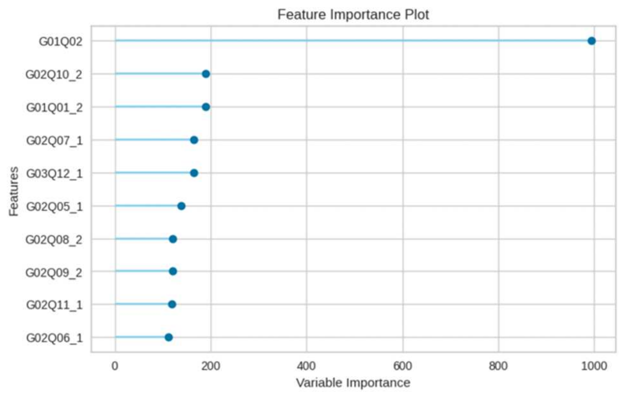

Fig. 9 Feature importance data Islamic Elementary School

B. Classification Model for Islamic Junior High School Levels

Table 2 shows the best model for Islamic Junior High school data. Unlike the best model in Islamic Elementary schools, Cat Boost Classifier has the highest accuracy value of 0.6380. Fig. 10 displays the Islamic Junior High School Learning Curve data results.

TABLE II
MODEL COMPARISON RESULT FOR ISLAMIC JUNIOR HIGH SCHOOL DATA

Model	Accuracy	AUC	Recall	Proc.	F1	Kappa	MCC	TT (Sec)
xgboost	0.5993	0.6353	0.6194	0.5923	0.6055	0.1987	0.1990	0.3120
Catboost	0.5975	0.6380	0.6183	0.5906	0.6041	0.1953	0.1956	5.2260
Lightgbm	0.5958	0.6378	0.6133	0.5896	0.6010	0.1919	0.1922	0.0440
Gbc	0.5905	0.6279	0.6057	0.5848	0.5949	0.1812	0.1814	0.1680
Ada	0.5837	0.6181	0.5968	0.5784	0.5874	0.1675	0.1676	0.0720
Ridge	0.5708	0.0000	0.5884	0.5652	0.5764	0.1418	0.1420	0.0150
Lda	0.5708	0.6005	0.5884	0.5652	0.5764	0.1418	0.1420	0.0270
Lr	0.5704	0.6005	0.5872	0.5649	0.5757	0.1411	0.1412	0.2980
Rf	0.5686	0.5962	0.5712	0.5650	0.5680	0.1372	0.1372	0.2490
Knn	0.5618	0.5865	0.5769	0.5568	0.5666	0.1237	0.1238	0.2290
Dt	0.5606	0.5722	0.5331	0.5606	0.5464	0.1209	0.1210	0.0210
Et	0.5603	0.5829	0.5331	0.5602	0.5463	0.1203	0.1205	0.2770
Svm	0.5354	0.0000	0.6381	0.5442	0.5416	0.0723	0.0922	0.0630
Nb	0.5348	0.5874	0.8843	0.5186	0.6537	0.0740	0.1032	0.0150
Qda	0.5052	0.5045	0.5039	0.5018	0.4654	0.0103	0.0130	0.0230
Dummy	0.5034	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0030

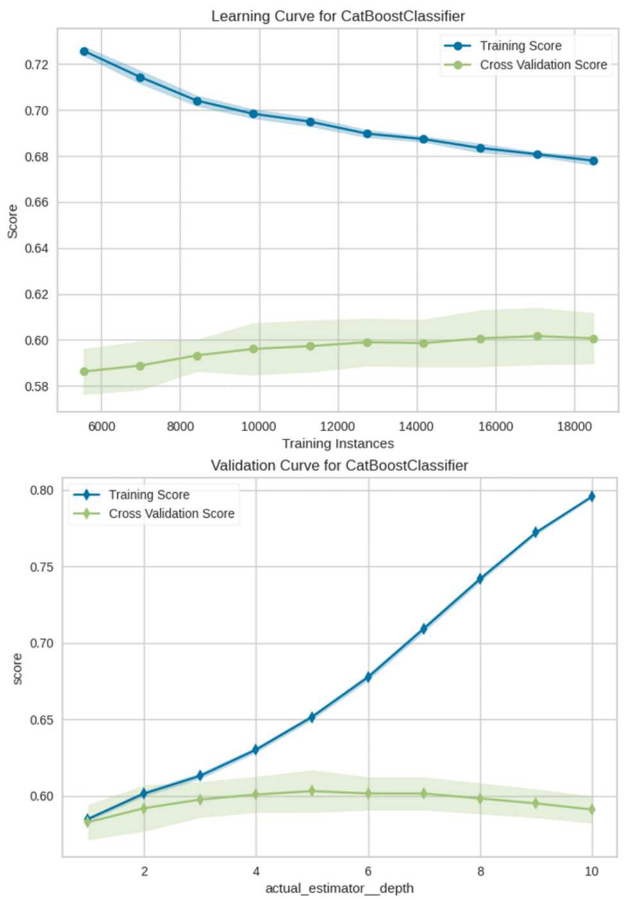

Fig. 10 Learning curve and Validation curve CatBoost Classifier Islamic Junior High school data

The first graph on Fig 10 depicts a "Validation Curve for CatBoostClassifier", illustrating how both the training and cross-validation scores evolve as the "actual_estimator_depth" increases. As the depth grows, the training score consistently rises, indicating potential overfitting, while the cross-validation score remains relatively stagnant. The second graph showcases a "Learning Curve for CatBoostClassifier", presenting the relationship between the number of training instances and model performance. As more training instances are utilized, the training score gradually decreases, and the cross-validation score slightly increases before plateauing. This suggests that adding more training data may not significantly improve the model's validation performance. Additionally, Fig 10 shows that the training score obtained is 0.8, while the cross-validation score is 0.59. With a difference of 0.21 points, these results indicate that the model applied is ideal.

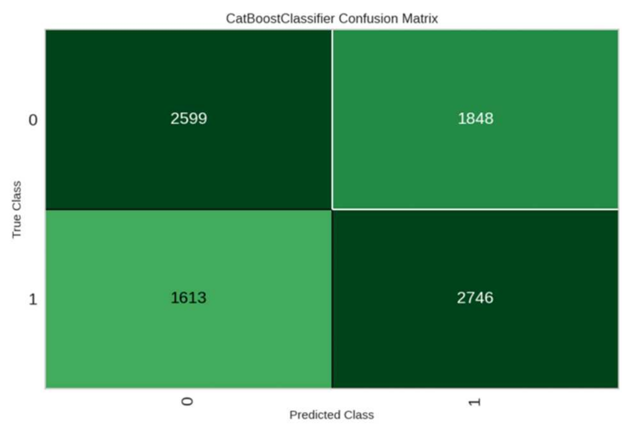

Fig. 11 Confusion matrix of Islamic Junior High school data

The Fig 11 displayed image represents a "Confusion Matrix" for the CatBoostClassifier. It shows the performance of the classifier in terms of its predictions. The top-left value (2599) represents the number of true negatives, meaning 2599 instances were correctly predicted as class 0. Conversely, the bottom-right value (2746) denotes the true positives, indicating 2746 instances were accurately predicted as class 1. The top-right value (1848) represents the false positives, where 1848 instances were incorrectly predicted as class 1. Lastly, the bottom-left value (1613) signifies false negatives, meaning 1613 instances were mistakenly predicted as class 0. This matrix provides a comprehensive view of the model's performance and its prediction accuracy for both classes.

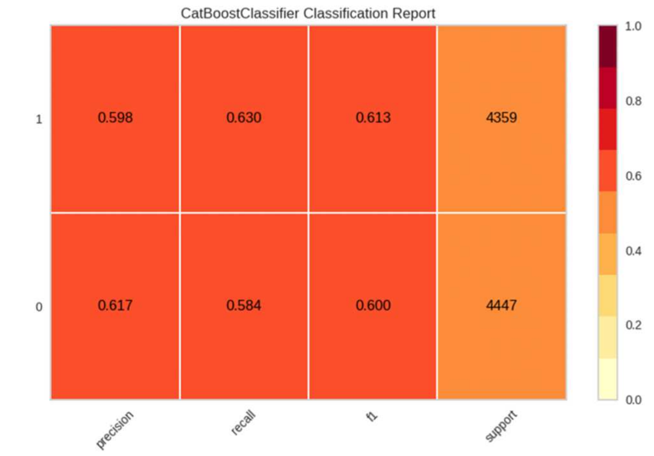

Fig. 12 Classification Report

Fig. 12 showcases a classification report for the CatBoostClassifier. For class 1, the precision stands at 0.598, recall at 0.630, and the F1 score is 0.613, with the support being 4,359 instances. On the other hand, for class 0, the classifier has a precision of 0.617, a recall of 0.584, and an F1 score of 0.600, based on a support of 4,447 instances. Precision evaluates the accuracy of the positive predictions, recall indicates the proportion of actual positives that were correctly classified, and the F1 score offers the harmonic mean of precision and recall. The 'support' metric reflects the total occurrences of each respective class in the dataset. This information provides a holistic assessment of the classifier's effectiveness for both class categories.

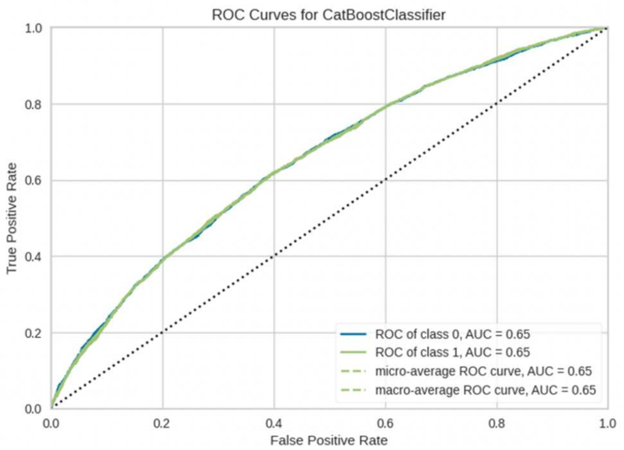

Fig. 13 ROC-AUC results for Islamic Junior High school data

Fig. 13 illustrates the ROC (Receiver Operating Characteristic) curves for the CatBoostClassifier. The ROC curve is a graphical representation of a classifier's performance, plotting the true positive rate against the false positive rate. From the provided legend, both classes (0 and 1) have an AUC (Area Under the Curve) value of 0.65. The AUC value, ranging from 0 to 1, is an indicator of the model's ability to distinguish between the classes – the closer the value is to 1, the better the model's predictive capability. An AUC of 0.65 suggests a moderately good classifier performance. Additionally, both the micro-average and macro-average ROC curves also have an AUC of 0.65, further reinforcing the consistent performance of the model across different evaluation metrics.

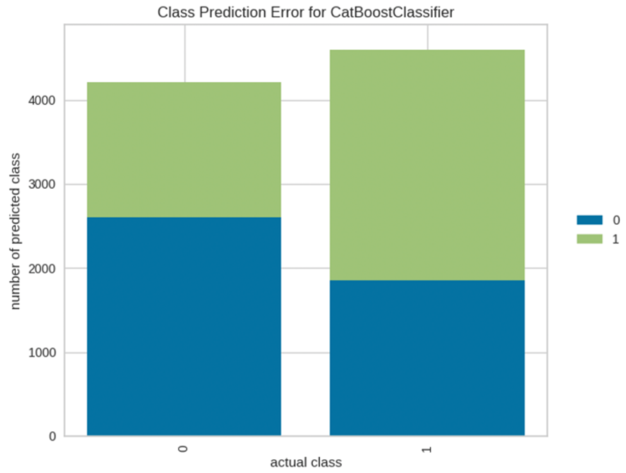

Fig. 14 Prediction Error Islamic Junior High School data

The graphic on Fig. 14 presents the Class Prediction Error for the CatBoostClassifier. It contrasts the actual class with the predicted class. For actual class 0, a significant portion was correctly predicted as class 0 (blue segment), but there's also a part misclassified as class 1 (green segment). Similarly, for actual class 1, while a large part was correctly identified as class 1 (green), there's a notable portion that was erroneously predicted as class 0 (blue). The exact numbers are not provided in the image, but it's evident that both classes have prediction errors, with actual class 1 having a higher proportion of misclassifications.

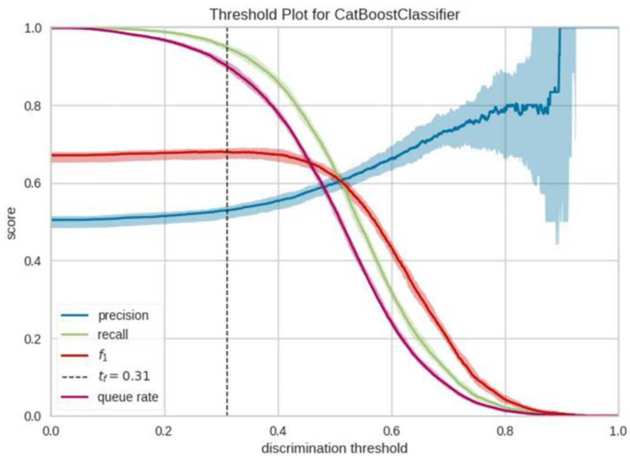

Fig. 15 Threshold Islamic Junior High School data

The graph displays on Fig. 15 the Threshold Plot for the CatBoostClassifier, showcasing the relationship between various evaluation metrics and the discrimination threshold. As the threshold increases, precision (blue line) decreases, while recall (red line) rises. The F1 score (green line), which harmonizes precision and recall, peaks around the threshold of 0.31, as indicated by the dotted line. The queue rate (purple line) represents the proportion of positive cases at each threshold. The shaded region highlights the variability in the threshold. The intersection of these curves can assist in choosing an optimal threshold for classification based on the desired trade-off between precision and recall.

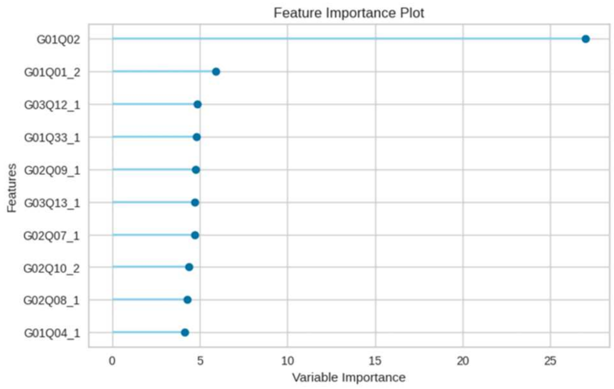

Fig. 16 Feature Importance of Islamic Junior High School Data

The graph on Fig. 16 presents a Feature Importance Plot, illustrating the significance of various features in a predictive model. Each feature is denoted by its unique code on the y-axis, and its importance is represented on the x-axis. Feature "G01Q02" stands out as the most influential, with an importance value exceeding 25. The rest of the features display lesser importance, with most of them clustered below the 10-mark. This visualization aids in understanding which features are crucial in making predictions and which ones have a minor impact on the model's outcome.

C. Classification Model for Islamic High School Levels

In Table 3, after comparing various models, the best model for High school is obtained, namely the Light Gradient Boosting machine with an accuracy value of 0.5959. This model is the same as the best model used in Islamic Elementary school data.

TABLE III
RESULT OF MODEL COMPARISON FOR ISLAMIC HIGH SCHOOL DATA

Model	Accuracy	AUC	Recall	Proc.	F1	Kappa	MCC	TT (Sec)
Lightgbm	0.5959	0.6348	0.5681	0.5999	0.5834	0.1917	0.1920	0.0220
Gbc	0.5944	0.6305	0.5653	0.5991	0.5815	0.1887	0.1891	0.0720
Catboost	0.5906	0.6268	0.5710	0.5928	0.5816	0.1811	0.1813	2.9300
xgboost	0.5884	0.6250	0.5742	0.5896	0.5817	0.1767	0.1768	0.1530
Ada	0.5862	0.6097	0.5457	0.5922	0.5679	0.1721	0.1727	0.0350
Knn	0.5625	0.5842	0.5308	0.5652	0.5472	0.1249	0.1252	0.1470
Rf	0.5622	0.5895	0.5527	0.5620	0.5573	0.1243	0.1244	0.1090
Lr	0.5621	0.5837	0.5325	0.5645	0.5480	0.1240	0.1242	0.2750
Ridge	0.5619	0.0000	0.5315	0.5644	0.5474	0.1236	0.1238	0.0040
Lda	0.5619	0.5838	0.5315	0.5644	0.5474	0.1236	0.1238	0.0070
Et	0.5565	0.5766	0.5072	0.5611	0.5326	0.1127	0.1132	0.1200
Dt	0.5537	0.5606	0.4934	0.5595	0.5242	0.1071	0.1079	0.0060
Nb	0.5311	0.5771	0.1369	0.6364	0.2249	0.0601	0.0962	0.0040
Svm	0.5185	0.0000	0.3940	0.5667	0.3519	0.0365	0.0618	0.0230
Dummy	0.5014	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0030
Qda	0.4939	0.4941	0.5575	0.4914	0.5132	-0.0118	-0.0131	0.0060

```
save_model(best, 'ma_classification_pipeline')

Transformation Pipeline and Model Successfully Saved
(Pipeline(memory=None,
          steps=[('dtypes',
                  DataTypes_Auto_infer(categorical_features=[],
                                       display_types=True, features_todrop=[],
                                       id_columns=[],
                                       ml_usecase='classification',
                                       numerical_features=[], target='clusters',
                                       time_features=[])),
                 ('imputer',
                  Simple_Imputer(categorical_strategy='not_available',
                                 fill_value_categorical=None,
                                 fill_value_numerical=None,
                                 numeric_str...
                  LGBMClassifier(boosting_type='gbdt', class_weight=None,
                                 colsample_bytree=1.0, importance_type='split',
                                 learning_rate=0.1, max_depth=-1,
                                 min_child_samples=20, min_child_weight=0.001,
                                 min_split_gain=0.0, n_estimators=100, n_jobs=-1,
                                 num_leaves=31, objective=None, random_state=615,
                                 reg_alpha=0.0, reg_lambda=0.0, silent='warn',
                                 subsample=1.0, subsample_for_bin=200000,
                                 subsample_freq=0)]],
          verbose=False),
```

Fig. 17 Best Islamic High school data storage model

Fig 17 showcases the process of saving a machine learning model using the `save_model` function. The model, named 'best', is being saved with the identifier "ma_classification_pipeline". This model consists of a transformation pipeline followed by an LGBM Classifier. The transformation includes automatic data type inference and a simple imputation strategy where unavailable categorical values are denoted as 'not_available'. As for the LGBM Classifier, it employs gradient boosting ('gdbt') with notable parameters such as a learning rate of 0.1, 100 estimators, and a maximum depth of -1 (indicating no limit). The random state is set to 615, ensuring reproducibility. The other parameters offer further insights into the model's fine-tuning and complexity, ensuring optimal performance.

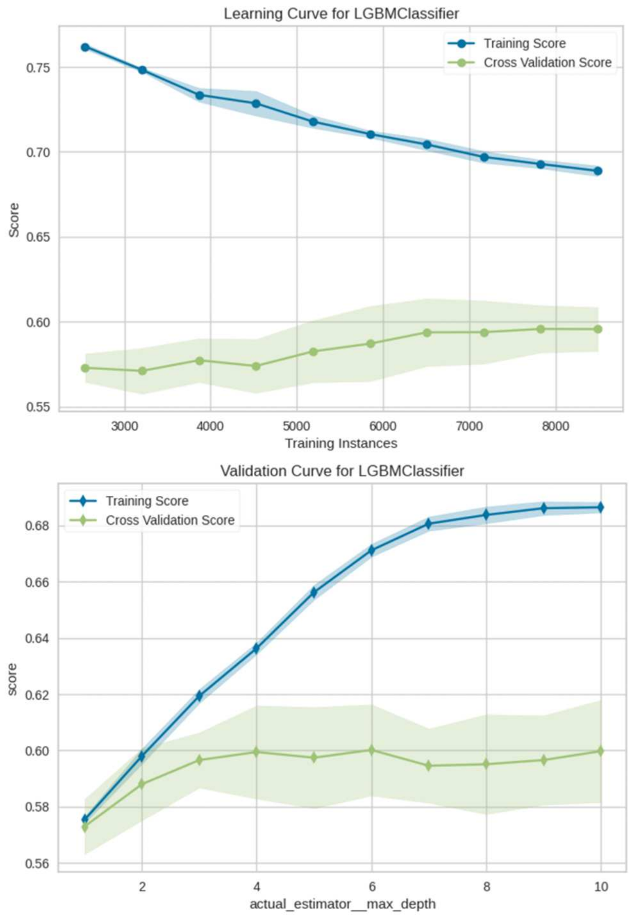

Fig. 18 Learning and Validation Curve Islamic High School data

The two graphs on Fig 18 represent performance evaluation of an LGBM Classifier. The first, titled "Learning Curve," shows that as the number of training instances increases, both the training and cross-validation scores tend to converge, hovering just above a score of 0.70. This suggests the model stabilizes with more data, achieving consistent performance. The second graph, "Validation Curve," plots the scores against varying maximum depths of the classifier, ranging from 2 to 10. As the depth increases, the training score rises sharply to nearly 0.68, while the cross-validation score remains relatively steady, slightly above 0.60. This might indicate the model begins to overfit as depth increases, given the widening gap between the training and cross-validation scores.

Fig 19 displays a confusion matrix for the LGBM Classifier. The matrix provides a breakdown of the classifier's predictions compared to the true outcomes. From the matrix, 1,276 instances were correctly predicted as class 0 (True Negatives), while 735 instances were falsely predicted as class 1 when they belong to class 0 (False Positives). Conversely, 1,161 instances were accurately predicted as class 1 (True Positives), and 871 instances were mistakenly predicted as class 0 when they were in fact class 1 (False Negatives). This matrix is essential in evaluating the performance of the classifier, highlighting areas where it performs well and where improvements may be needed.

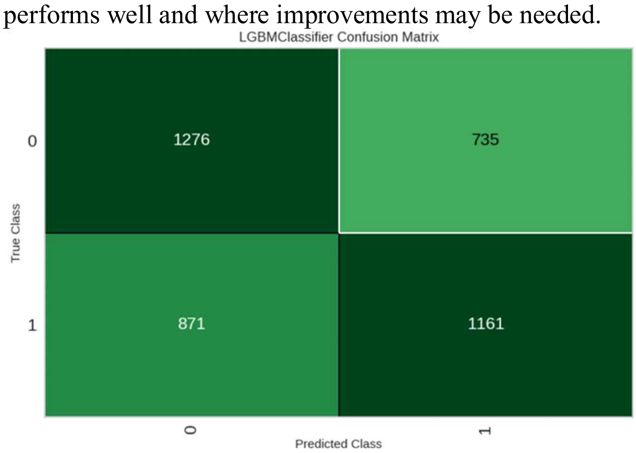

Fig. 19 Confusion Matrix Islamic High School data

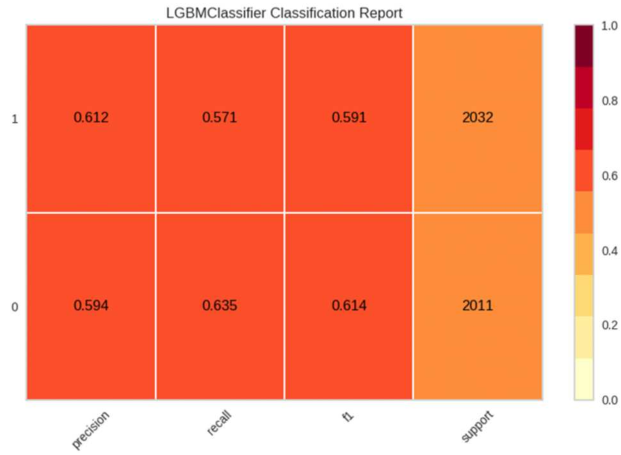

Fig. 20 Classification report data Islamic High School

Fig. 20 presents a classification report for the LGBM Classifier. For class 1, the classifier achieved a precision of 0.612, a recall of 0.571, and an F1 score of 0.591, with a support of 2,032 instances. Meanwhile, for class 0, the classifier posted a precision of 0.594, a recall of 0.635, and an F1 score of 0.614, based on a support of 2,011 instances. Precision assesses the accuracy of the positive predictions, recall measures the proportion of actual positives correctly identified, and the F1 score provides the harmonic mean of precision and recall. The 'support' indicates the number of actual occurrences of each class in the dataset. This report gives a comprehensive evaluation of the classifier's performance for both classes.

The image presents in Fig 21 is the ROC (Receiver Operating Characteristic) curves for the LGBM Classifier. Both the curves for class 0 and class 1 closely follow each other, each yielding an AUC (Area Under the Curve) score of 0.64. Additionally, the micro-average and macro-average ROC curves also have identical AUC values of 0.64. The AUC value quantifies the overall ability of the model to distinguish between the positive and negative classes. An AUC of 0.64 indicates a reasonable classification performance, which is better than random guessing but might benefit from further optimization. The diagonal dotted line represents the ROC curve for a random classifier; hence, the LGBM Classifier performs notably better than a random guess.

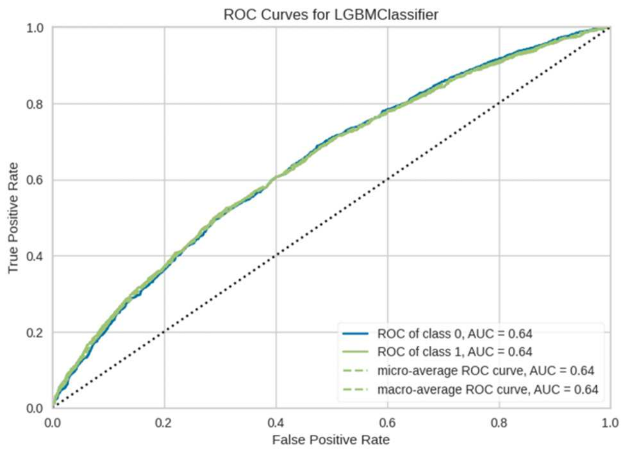

Fig. 21 ROC-AUC curve for Islamic High school data

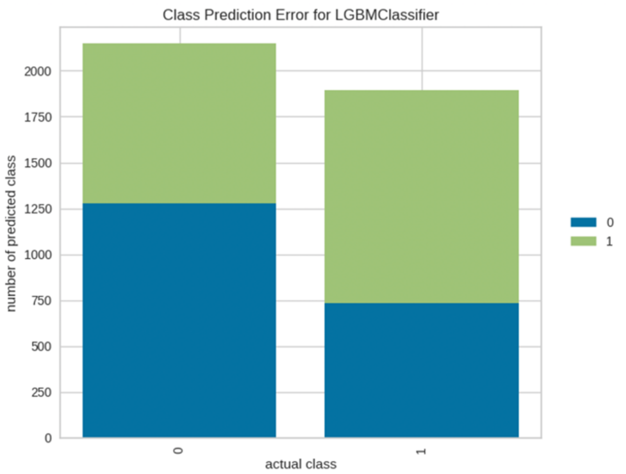

Fig. 22 Prediction Error Islamic High School data

The graph showcases the class prediction error for the LGBM Classifier. For the actual class "0", approximately 1250 instances were correctly classified as "0" (blue), but about 750 instances were mistakenly classified as "1" (green). Conversely, for the actual class "1", roughly 1250 cases were accurately classified as "1" (green), while about 750 cases were incorrectly classified as "0" (blue). The misclassification is relatively balanced for both classes, indicating potential areas for refining the model.

The graph in Fig 23 presents a threshold plot for the LGBM Classifier. It illustrates how precision, recall, F1 score, and queue rate metrics evolve based on different discrimination thresholds. At a threshold of approximately 0.30, denoted by the dashed line, the F1 score reaches its peak. Beyond this point, as the threshold increases, precision grows while recall decreases. The queue rate, which represents the proportion of instances predicted as positive, steadily declines as the threshold elevates. The histogram in the background indicates the distribution of scores, with a concentration of instances around the 0.6 to 1.0 range. This visualization aids in determining the optimal threshold for model classification based on desired metrics.

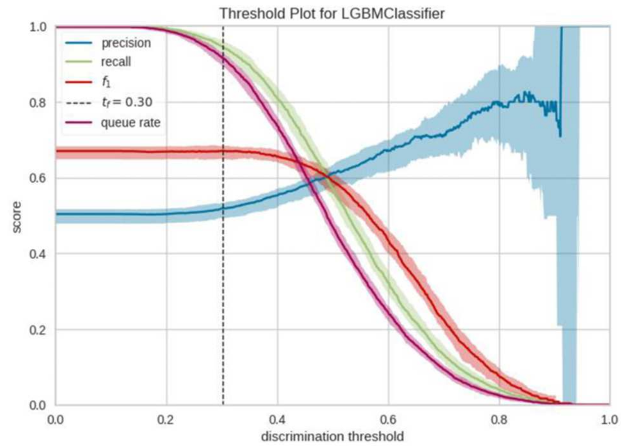

Fig. 23 Threshold of Islamic High school data

Fig. 24 illustrates the Feature Importance Plot derived from the Islamic High School questionnaire data. Each dot on the plot signifies the importance of a particular feature, with its horizontal position indicating the magnitude of its significance. The feature "G01Q02" stands out as the most impactful, possessing the highest value on the importance scale. This feature corresponds to a question about the province of origin of the students who participated in the questionnaire. Meanwhile, other features like "G02Q06_1" and "G02Q11_1" register comparatively lower importance values. This representation offers valuable insights into which features considerably influence the model's predictions, serving as a guide for potential feature engineering or model enhancements.

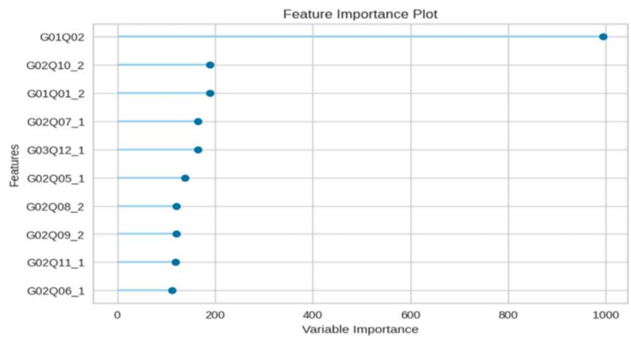

Fig. 24 Feature Importance of Islamic High School Data

IV. Conclusion

This study successfully pioneered a system that employs the Gradient Boosting Classifier to identify learning loss in Islamic religious education subjects within Indonesia. Tailored to accommodate students from diverse educational stages—elementary, junior high, and high school—this system integrates an expansive set of educational questions. These span essential domains such as the Koran, Hadith, Aqedah, morality, Fiqh, and the history of Islamic culture and serve as the foundational testing material.

Our investigation unearthed pronounced disparities in the best-performing models contingent on the educational levels. The Light Gradient Boosting Machine was discerned as the most effective for Islamic Elementary and High school students, while the CatBoost Classifier was paramount for Islamic Junior High school students. Such outcomes offer consequential insights for future investigations and enhancements in Islamic religious education. Subsequent research endeavors should delve into the root causes behind these variations in optimal models for each educational tier. Probing into specific hurdles and knowledge gaps in Islamic religious education that lead to learning loss can be particularly enlightening. Such insights will be instrumental in crafting bespoke educational strategies, ensuring the preservation and enhancement of learning in this crucial subject area.

ACKNOWLEDGMENT

The authors thank the Directorate General of Islamic Education, Ministry of Religion of the Republic of Indonesia, for supporting the implementation of this research.

REFERENCES

[1] I. Muhtarom, “Apa Itu Learning Loss yang Ditakutkan Nadiem Makarim?,” *Tempo.co*, 2021.
[2] R. D. Adiputri, “”Learning Loss” di Masa Pandemi,” *kompas.id*, 2022.
[3] W. Durongkaveroj, “Learning loss due to university closures during the COVID-19 pandemic: Evidence from Thailand’s largest public university,” *Thail. World Econ.*, vol. 41, no. 2, 2023.
[4] J. P. Azevedo, A. Hasan, D. Goldemberg, K. Geven, and S. A. Iqbal, “Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates,” *World Bank Res. Obs.*, vol. 36, no. 1, pp. 1–40, 2021.
[5] S. Budi, I. S. Utami, R. N. Jannah, N. L. Wulandari, N. A. Ani, and W. Saputri, “Deteksi Potensi Learning Loss pada Siswa Berkebutuhan Khusus Selama Pembelajaran Daring Masa Pandemi Covid-19 di Sekolah Inklusif,” *J. Basicedu*, vol. 5, no. 5, pp. 3607–3613, 2021.
[6] J. Schult, N. Mahler, B. Fauth, and M. A. Lindner, “Did students learn less during the COVID-19 pandemic? Reading and mathematics competencies before and after the first pandemic wave,” *Sch. Eff. Sch. Improv.*, vol. 33, no. 4, pp. 544–563, 2022.
[7] M. A. Maulyda, M. Erfan, and V. R. Hidayati, “Analisis situasi pembelajaran selama pandemi covid-19 di sdn senurus: kemungkinan terjadinya learning loss,” *COLLASE (Creative Learn. Students Elem. Educ.*, vol. 4, no. 3, pp. 328–336, 2021.
[8] J. Hoofman and E. Secord, “The effect of COVID-19 on education,” *Pediatr. Clin.*, vol. 68, no. 5, pp. 1071–1079, 2021.
[9] M. A. Callejon-Leblic *et al.*, “Loss of smell and taste can accurately predict COVID-19 infection: a machine-learning approach,” *J. Clin. Med.*, vol. 10, no. 4, p. 570, 2021.
[10] C. Ardington, G. Wills, and J. Kotze, “COVID-19 learning losses: Early grade reading in South Africa,” *Int. J. Educ. Dev.*, vol. 86, p. 102480, 2021.
[11] INOVASI & Pusat Standar dan Kebijakan Pendidikan, I.N.O.V.A.S.I., and P. S. Kebijakan Pendidikan, *Studi Analisa Situasi Pembelajaran dalam Masa Pandemi*. Jakarta, 2022.
[12] A. Widodo, P. D. Angga, M. Syazali, and U. Umar, “Mainstreaming Parental Involvement in Post-Pandemic: Resolving Learning Loss with the Partnership Model in Elementary Schools,” *J. Kependidikan J. Has. Penelit. dan Kaji. Kepustakaan di Bid. Pendidikan, Pengajaran dan Pembelajaran*, vol. 9, no. 2, p. 365, 2023.
[13] A. Banawi, A. Latuconsina, and S. Latuconsina, “Exploring the Students’ Reading, Writing, and Numeracy Skills in Southeast Maluku Regency Coastal Elementary Schools,” *Al Ibtida J. Pendidik. Guru MI*, vol. 9, no. 2, pp. 252–264, 2022.
[14] F. Firdaus and A. Septiady, “Peningkatan Kemampuan Literasi Dan Numerasi Di Sekolah 3t (Tertinggal, Terluar, Terdepan) Di Masa Pandemi Covid-19 Melalui Program Kampus Mengajar,” *SKYLANDSEA Prof. J. Ekon. Bisnis dan Teknol.*, vol. 1, no. 2, pp. 213–220, 2021.
[15] R. Oostdam, M. van Diepen, B. Zijlstra, and R. Fukkink, “Effects of the COVID-19 school lockdowns on language and math performance of students in elementary schools: implications for educational practice and reducing inequality,” *Eur. J. Psychol. Educ.*, pp. 1–21, 2023.
[16] M. Krywult-Albańska and Ł. Albański, “Gender and educational inequalities during the COVID-19 pandemic: Preliminary insights from Poland,” *Sustainability*, vol. 13, no. 22, p. 12403, 2021.
[17] C. Boruchowicz, S. W. Parker, and L. Robbins, “Time use of youth during a pandemic: evidence from Mexico,” *World Dev.*, vol. 149, p. 105687, 2022.
[18] C. F. Drane, L. Vernon, and S. O’Shea, “Vulnerable learners in the age of COVID-19: A scoping review,” *Aust. Educ. Res.*, vol. 48, no. 4, pp. 585–604, 2021.
[19] Organisation for Economic Co-operation and Development (OECD), “The Future of Education and Skills (Education 2023),” 2018.
[20] Z. Ali, *Pendidikan Agama Islam*. Jakarta: Bumi Aksara, 2007.
[21] S. Arifin, *Pendidikan Agama Islam*. Deepublish, 2018.
[22] A. R. Nurjaman, *Pendidikan Agama Islam*. Bumi Aksara, 2020.
[23] Direktorat Jendral Pendidikan Islam Kementrian Agama RI, “Rekapitulasi Data Pokok Pendidikan Islam (Madrasah Semester Ganjil 2020/2021),” 2021.
[24] F. Amalina *et al.*, “Blending big data analytics: Review on challenges and a recent study,” *Ieee Access*, vol. 8, pp. 3629–3645, 2019.
[25] J. Wang, Y. Zhao, P. Balamurugan, and P. Selvaraj, “Managerial decision support system using an integrated model of AI and big data analytics,” *Ann. Oper. Res.*, pp. 1–18, 2022.
[26] P. Ghavami, “Big data analytics methods,” in *Big Data Analytics Methods*, de Gruyter, 2019.
[27] X. Shu and Y. Ye, “Knowledge Discovery: Methods from data mining and machine learning,” *Soc. Sci. Res.*, vol. 110, p. 102817, 2023.
[28] H. M. Shah, B. B. Gardas, V. S. Narwane, and H. S. Mehta, “The contemporary state of big data analytics and artificial intelligence towards intelligent supply chain risk management: A comprehensive review,” *Kybernetes*, vol. 52, no. 5, pp. 1643–1697, 2023.
[29] X. Zhu and A. B. Goldberg, “Introduction to semi-supervised learning,” *Synth. Lect. Artif. Intell. Mach. Learn.*, vol. 3, no. 1, pp. 1–130, 2009.
[30] J. E. Van Engelen, H. H. Hoos, J. E. Engelen, and H. H. Hoos, “A survey on semi-supervised learning,” *Mach. Learn*, vol. 109, no. 2, pp. 373–440, 2020.
[31] M. F. A. Hady and F. Schwenker, “Semi-supervised learning,” *Handb. Neural Inf. Process.*, pp. 215–239, 2013.
[32] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, “Mixmatch: A holistic approach to semi-supervised learning,” *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.
[33] J. Ran, Y. Ji, and B. Tang, “A semi-supervised learning approach to ieee 802.11 network anomaly detection,” in *2019 IEEE 89th vehicular technology conference (VTC2019-Spring)*, 2019, pp. 1–5.
[34] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, “A survey on contrastive self-supervised learning,” *Technologies*, vol. 9, no. 1, p. 2, 2020.
[35] H. Di, X. Ke, Z. Peng, and Z. Dongdong, “Surface defect classification of steels with a new semi-supervised learning method,” *Opt. Lasers Eng.*, vol. 117, pp. 40–48, 2019.
[36] C. Yin and Z. Chen, “Developing sustainable classification of diseases via deep learning and semi-supervised learning,” in *Healthcare*, 2020, vol. 8, no. 3, p. 291.
[37] T. Jiang, J. L. Gradus, and A. J. Rosellini, “Supervised machine learning: a brief primer,” *Behav. Ther.*, vol. 51, no. 5, pp. 675–687, 2020.
[38] M. Binkhonain and L. Zhao, “A review of machine learning algorithms for identification and classification of non-functional requirements,” *Expert Syst. with Appl. X*, vol. 1, p. 100001, 2019.
[39] P. C. Sen, M. Hajra, and M. Ghosh, “Supervised classification algorithms in machine learning: A survey and review,” in *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018*, 2020, pp. 99–111.
[40] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, “Medical image segmentation using deep learning: A survey,” *IET Image Process.*, vol. 16, no. 5, pp. 1243–1267, 2022.