Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi

Bilecik Şeyh Edebali University Journal of Science

e-ISSN: 2458-7575 (https://dergipark.org.tr/tr/pub/bseufbd)

Araştırma Makalesi – Research Article

Geliş Tarihi / Received: 08/02/2025

Kabul Tarihi / Accepted: 14/04/2025

Yayın Tarihi / Published: 30/11/2025

Predicting Graduation of Associate Degree Students: An Ensemble Learning Approach

Ön Lisans Öğrencilerinin Mezuniyetinin Tahmini: Bir Topluluk Öğrenme Yaklaşımı

Ahmet Kala[1*], Cem Özkurt[2], Yasin Ok[3]

[1*]Sakarya University of Applied Sciences/ Department of Information Technologies/ Sakarya, Türkiye/ ahmetkala@subu.edu.tr / https://orcid.org/0000-0002-0598-1181

[2]Sakarya University of Applied Sciences/ Department of Computer Engineering/ Sakarya, Türkiye/ cemozkurt@subu.edu.tr / https://orcid.org/0000-0002-1251-7715

[3]Sakarya University of Applied Sciences/ Department of Information Technologies/ Sakarya, Türkiye/ yasinok@subu.edu.tr / https://orcid.org/0009-0006-5918-1979

Etik Beyan: Bu çalışmanın hazırlanma sürecinde bilimsel ve etik ilkelere uyulduğu ve yararlanılan tüm çalışmaların kaynakçada belirtildiği beyan olunur.

Yapay Zeka Etik Beyanı: Yazar bu makalenin hazırlanma sürecinin hiç bir aşamasında yapay zekadan faydalanılmadığını; bu konuda tüm sorumluluğun kendisine(kendilerine) ait olduğunu beyan etmektedir.

Çıkar Çatışması: Çıkar çatışması beyan edilmemiştir.

Finansman: Bu araştırmayı desteklemek için dış fon kullanılmamıştır.

Ethical Statement: It is declared that scientific and ethical principles were followed during the preparation of this study and that all studies used are stated in the bibliography.

Artificial Intelligence Ethical Statement: The author declares that artificial intelligence was not utilized at any stage of the preparation process of this article and accepts full responsibility in this regard.

Conflicts of Interest: The author(s) has no conflict of interest to declare.

Grant Support: The author(s) acknowledge that they received no external funding to support this research.

Predicting Graduation of Associate Degree Students: An Ensemble Learning Approach

ABSTRACT

On-time graduation is an important indicator of success in associate degree programs that aim to provide students with basic knowledge and skills in a specific profession or field and enable them to quickly enter the workforce. However, graduation delays increase the economic burden and cause psychological difficulties by delaying the process of starting a professional career. This study aims to estimate students' on-time graduation predictions at an early stage, that is, at the beginning of the semester, and to identify at-risk students who may not graduate on time. Thus, time will be gained for intervention for these students at risk. Ensemble learning methods and classical machine learning models, which were created by combining multiple models, were used for early prediction. The effectiveness of the models was examined with demographic information, high school achievements, university entrance exam results and academic performance data obtained from an associate degree program at a state university. Classification performance estimates were made in three academic stages: the beginning of the first semester, the end of the first semester and the end of the second semester. The results were evaluated according to the F1 score performance metric and it was seen that the LR model performed better at the beginning of the semester and the ensemble learning methods performed better in the other two periods. Additionally, cumulatively adding within-term academic performance data increased the classification performances of all models.

Keywords- On-Time Graduation, Educational Data Mining, Ensemble Learning, Machine Learning

Highlights

- This study demonstrates that ensemble learning methods can accurately predict associate degree students' on-time graduation at early stages.
- Ensemble learning methods, particularly CatBoost and XGBoost, achieved the highest F1-scores in later prediction stages (end of first and second semesters), outperforming traditional models.
- Logistic Regression provided the most balanced performance with limited data at the beginning of the semester.
- Adding within-term academic performance data cumulatively improved the performance of all models.

Ön Lisans Öğrencilerinin Mezuniyetinin Tahmini: Bir Topluluk Öğrenme Yaklaşımı

ÖZ

Zamanında mezuniyet, öğrencilere belirli bir meslek veya alanda temel bilgi ve beceriler kazandırmayı ve onları hızla iş gücüne katılmalarını sağlamayı amaçlayan önlisans programlarında başarıyı gösteren önemli bir göstergedir. Ancak, mezuniyet gecikmeleri profesyonel kariyere başlama sürecini erteleyerek ekonomik yükü artırmakta ve psikolojik zorluklara yol açmaktadır. Bu çalışma, öğrencilerin zamanında mezuniyet tahminlerini erken aşamada, yani dönem başında tahmin etmeyi ve zamanında mezun olamayacak risk altındaki öğrencileri belirlemeyi amaçlamaktadır. Böylece, bu risk altındaki öğrenciler için müdahale için zaman kazanılmış olacaktır. Birden fazla modeli birleştirerek oluşturulan topluluk öğrenme yöntemleri ve klasik makine öğrenme modelleri erken tahmin için kullanılmıştır. Modellerin etkinliği, bir devlet üniversitesinin önlisans programından alınan demografik bilgiler, lise başarıları, üniversite giriş sınavı sonuçları ve akademik performans verileri ile incelenmiştir. Sınıflandırma performansı tahminleri üç akademik aşamada yapılmıştır: ilk dönemin başı, ilk dönemin sonu ve ikinci dönemin sonu. Sonuçlar, F1 skoru performans metriğine göre değerlendirilmiş ve dönem başında LR modelinin, diğer iki dönemde ise topluluk öğrenme yöntemlerinin daha iyi performans sergilediği görülmüştür. Ayrıca, dönem içi akademik performans verilerinin kümülatif olarak eklenmesi, tüm modellerin sınıflandırma performanslarını artırmıştır.

Anahtar Kelimeler- Zamanında Mezuniyet, Eğitim Veri Madenciliği, Topluluk Öğrenme, Makine Öğrenimi

Öne Çıkanlar

- Bu çalışma, topluluk öğrenme yöntemlerinin ön lisans öğrencilerinin zamanında mezuniyetini erken aşamalarda yüksek doğrulukla tahmin edebileceğini göstermektedir.
- Topluluk öğrenme yöntemleri, özellikle CatBoost ve XGBoost, sonraki tahmin aşamalarında (birinci ve ikinci dönem sonları) en yüksek F1-skorlarına ulaşarak geleneksel modellerden daha iyi performans göstermiştir.
- Dönem başında ise Lojistik Regresyon, sınırlı veriyle en dengeli performansı sağlamıştır.
- Dönem içi akademik performans verilerinin eklenmesi, tüm modellerin başarısını artırmıştır.

I.INTRODUCTION

In associate degree programs designed to prepare students for the workforce by providing them with basic skills and knowledge in specific areas, on-time graduation rate is a critical performance measure. These programs aim to equip students with basic expertise and ensure a smooth transition to professional life after graduation. However, graduation rates or dropout rates are often affected by a wide variety of factors such as demographic, academic, and institutional.

The prolonged graduation period of associate degree students can cause many negative effects such as decreased academic success, loss of motivation, prolonged education, increased education costs, financial pressure on the student and their family, late entry into the labor market, missing career opportunities, and stress, anxiety, and loss of self-confidence in students. Therefore, it is very important to predict students who are at risk, i.e. who will not graduate on time, at an early stage to allow timely interventions.

In recent years, various academic studies have explored the use of machine learning and ensemble learning models to predict on-time graduation using educational data. The existing literature predominantly focuses on undergraduate and graduate-level students, with limited research targeting high school, doctoral, and particularly associate degree students. Commonly used models include ensemble methods such as AdaBoost and Random Forest [1–5], while more advanced techniques such as Gradient Boosting, XGBoost, and LightGBM have seen relatively limited application [6]. These studies typically utilize demographic attributes and academic indicators such as GPA.

This study differs from prior work by specifically focusing on associate degree students—a group that has received comparatively less attention in the literature. Furthermore, it employs advanced ensemble learning techniques, including CatBoost, LightGBM, Gradient Boosting, and XGBoost, to assess their effectiveness in predicting on-time graduation. The objective is to enhance institutional efficiency and increase the rate of on-time graduation by enabling early identification of students at risk.

The study aims to predict whether associate degree students will graduate on time using ensemble learning methods and traditional machine learning techniques. In addition, it evaluates and compares the performance of these methods in predicting on-time graduation. The dataset used includes demographic information, high school academic performance, national university entrance exam scores, and academic performance in higher education, collected from an associate degree program at a public university. Predictions were conducted at three critical academic milestones: the beginning of the first semester, the end of the first semester, and the end of the second semester.

The remainder of this paper is structured as follows: Section II presents a review of related work in the literature; Section III describes the methodology used in the study; Section IV reports and discusses the experimental findings; and Section V concludes the paper with a summary of key insights and implications.

II.STUDIES ON STUDENT GRADUATION PREDICTION IN THE LITERATURE

In recent years, various academic studies have been conducted on machine learning models and ensemble methods to predict timely graduation using educational data. These studies have been examined in terms of the models used, datasets, and data content, focusing on different educational levels (bachelor's, master's, doctorate) and various datasets. A comparative analysis of these studies is presented in Table 1.

At the bachelor's level, Ananto (2024) achieved an accuracy of 87.755% using bagging and boosting methods with academic data such as GPA, attendance, and credit hours [7]. Anggrawan et al. (2023) achieved 77% accuracy using an SVM model supported by SMOTE [8]. Ariska et al. (2022) achieved 96.85% training and 93.72% testing accuracy using the C5.0 algorithm, emphasizing the importance of factors such as GPA, gender, and study program [9]. Willy Hanafi et al. achieved 90.32% accuracy using the C4.5 decision tree algorithm on a dataset that included academic and psychometric test scores [10]. Kurniawan et al. achieved 93.10% accuracy using the C4.5 algorithm on student data from 2010-2012 [11]. Ace C. Lagman et al. [12] achieved 86.82% accuracy using logistic regression, while Law et al. [6] achieved 87.24% accuracy by combining ensemble-SMOTE with logistic regression. Mehta achieved 85% accuracy using Naive Bayes on 2018 bachelor's student data [13]. Sonia Sri Muliani et al. [14] achieved 69.84% accuracy using artificial neural networks (ANN), while Devi Ayu Rahmawati et al. [15] achieved 91% testing accuracy using the random forest method. Samuel et al. achieved 84.015% accuracy by combining the C4.5 decision tree algorithm with SMOTE [16]. Tholib et al. achieved 96.12% accuracy using the Naive Bayes model [17].

At the master's level, Hafsat S. Bako et al. achieved 83.74% accuracy using the random forest method, incorporating factors such as age, gender, marital status, and CGPA [1]. Riadi et al. achieved 80% accuracy using the Naive Bayes model, emphasizing the importance of variables such as graduate GPA, TOEFL scores, and study duration [18].

At the doctorate level, Allen et al. (2016) achieved 10.2% explained variability using logistic regression, examining the role of preadmission variables (prior degree, GPA, unsatisfactory grades, etc.) in predicting graduation outcomes for doctoral students [19].

Literature review reveals that ensemble learning methods, particularly AdaBoost and Random Forest [1–5] have been extensively employed in existing studies. In contrast, Gradient Boosting, XGBoost, and LightGBM approaches have received relatively less attention [6]. The majority of research has concentrated on undergraduate and master's students, whereas investigations focusing on high school [20], doctoral [1], and associate degree populations remain scarce. Commonly utilized predictors include demographic variables and academic achievement metrics such as GPA.

Table 1. Comparative analysis of machine learning and ensemble models for on-time graduation prediction.

Ref.	Ensemble Models	Other ML Models	Best Model	Best Accuracy or F1-Score	Education Level	Dataset Content
[7]	CART with Bagging, CART with Boosting	CART	Boosting with CART	87.755% (accuracy)	Bachelor’s Degree	GPAs, attendance, credit hours, length of study
[8]	None	SVM, SVM with SMOTE	SVM with SMOTE	77% (accuracy)	Bachelor’s Degree	GPAs, gender, graduation status
[9]	None	C5.0 DT, Linear Regression	C5.0 Algorithm	93.72% (accuracy)	Bachelor’s Degree	GPAs, gender, study program, length of study
[10]	None	C4.5 DT	C4.5 DT	90.32% (accuracy)	Bachelor’s Degree	Academic scores, psychotest scores, EQ data
[11]	None	C4.5 DT	C4.5 DT	93.10% (accuracy)	Bachelor’s Degree	GPAs, semester performance, and gender
[12]	Bagging, Boosting, Stacking	LR, ANN, DT, NB	LR	86.82% (accuracy)	Bachelor’s Degree	Student graduation data
[6]	Gradient Boosting, RF, Adaptive Boosting, XGBoost, LightGBM, Categorical Boosting	LR, LDA, NB, KNN, SVM, DT	LR	87.24% (accuracy)	Bachelor’s Degree	Student profiles, registration details, grades, and GPAs
[13]	None	NB	NB	85% (accuracy)	Bachelor’s Degree	GPAs, semester performance, and demographics
[14]	None	ANN	ANN	69.84% (accuracy)	Bachelor’s Degree	Gender, Work, Age, GPS GPAs, Graduation Status
[15]	RF	None	RF	91% (accuracy)	Bachelor’s Degree	GPAs, Graduation Status, etc.
[16]	None	C4.5 DT	C4.5 DT (without SMOTE)	84.015% (accuracy)	Bachelor’s Degree	GPAs, repeating courses, study leave, gender, religion, , etc.
[17]	None	C4.5 DT, NB	NB	96.12% (accuracy)	Bachelor’s Degree	Student data including study program, GPAs, etc.
[1]	RF, Bootstrap Aggregating, AdaBoost	DT, SMO, Random Tree	RF	83,74% (accuracy)	Master’s Degree	Age, gender, marital status, CGPA, year of entry, year of graduation
[18]	None	NB, KNN	NB	80% (accuracy)	Master’s Degree	GPAs, TOEFL scores, study duration
[19]	None	LR	LR	10.2% (explained variability)	Doctorate	Preadmission variables

[2]	RF, AdaBoost	LR, DT, Gaussian, SVM, KNN, MLP	MLP	91.87% (accuracy)	Bachelor's Degree	GPAs
[3]	AdaBoost, Bagging, RF, Rotation Forest	None	Rotation Forest	75.95% (accuracy)	Bachelor's Degree	Student grades
[4]	Stacking Ensemble, RF, AdaBoost	LDA, LR, CART,KNN, SVM, ANN, NB	Stacking Ensemble (LR, NB, KNN, NN)	74.82% (accuracy)	Bachelor's Degree	Demographic attributes, first-year academic performance, financial status
[5]	AdaBoost	LR, DT, SVM	AdaBoost (DT)	82% (F1-Score)	Bachelor's Degree	Name, gender, student status, enrolment year, parents' education, GPAs, etc.
[20]	None	Multilevel LR	Multilevel LR	None	High School	Student engagement survey data, attendance, behavior, and academic performance

III.METHOD

This study used ensemble and traditional machine learning models to predict whether students enrolled in associate degree programs at a public university would graduate on time across three different academic time periods. The three academic time periods are as follows:

- Beginning of the First Semester: Using demographic data, high school performance, and entrance exam results.
- End of the First Semester: Including academic performance data from the first semester.
- End of the Second Semester: Adding the second semester's academic data for the most updated predictions.

Figure 1 visually presents the workflow followed in this study. The methodology consists of key stages, including data collection, data preprocessing, data splitting, model development, and evaluation. Each of these stages will be discussed in detail in the following sections.

A. Data Collection

The dataset used in this study was collected from 3,171 students who enrolled in associate degree programs at Sakarya University of Applied Sciences, a public university in Türkiye, during the 2022–2023 academic year. These students were expected to graduate by the end of the 2023–2024 academic year under the standard program duration. It includes demographic information, high school performance data, Higher Education Institutions Examination (YKS) results, and higher education academic performance data, as detailed in Table 2.

Table 2. Associate Degree Dataset.

Group	Attribute	Description	Type
Demographic	Gender	The gender of the student (Female/Male)	Categorical
	Age	The age of the student	Numeric
	AddressRegion	The region of the province where the student resides	Categorical
High School and University Entrance	PreferenceOrder	The order of the student's placement preference	Numeric
	AdmissionType	The type of exam application (YKS/YKS_EK)	Categorical
	PlacementRank	The placement rank of the student's chosen program	Numeric
	PlacementScore	The placement score of the student's chosen program	Numeric
	HighSchoolGPA	The high school GPA of the student	Numeric
	HighSchoolTopStudentStatus	Is the student the top student of the high school (YES/NO)	Categorical
	HighSchoolStatus	The educational status of the high school the student graduated from	Categorical
	HighSchoolGroup	The group of the type of high school the student graduated from	Categorical
Academic Performance	Department	The program where the student studied	Categorical
	FirstSemester	GPA in the first semester	Numeric
	SecondSemester	GPA in the second semester	Numeric
Output	GraduationStatus	The graduation status of the student (GRADUATED/NOT GRADUATED)	Categorical

B. Data Preprocessing

Data preprocessing consists of three main steps: data integration, data cleaning, and data transformation. In the first step, higher education academic performance data provided by the Student Information System (SIS) were integrated with high school and Higher Education Institutions Exam (YKS) results data offered to universities via the Presidency of Measurement, Selection and Placement Center (ÖSYM) web service. During the second step of data cleaning, the "Is the student the top student of the high school" variable, which had only a single unique value, was removed from the dataset. There are no missing, erroneous, or inconsistent data that could negatively affect the modeling process. Finally, in the data transformation step, since machine learning models operate exclusively on numerical data, categorical variables in the dataset were converted into an appropriate format. LabelEncoder was applied to categorical variables, assigning each category a unique numerical value. Numerical variables were normalized to ensure fair evaluation and weighting of variables at different scales throughout the modeling process. Standardization was performed using mean and standard deviation, each feature was rescaled to have a mean of 0 and a standard deviation of 1. The standardized data was then normalized between 0 and 1 using the Min-Max scaling method.

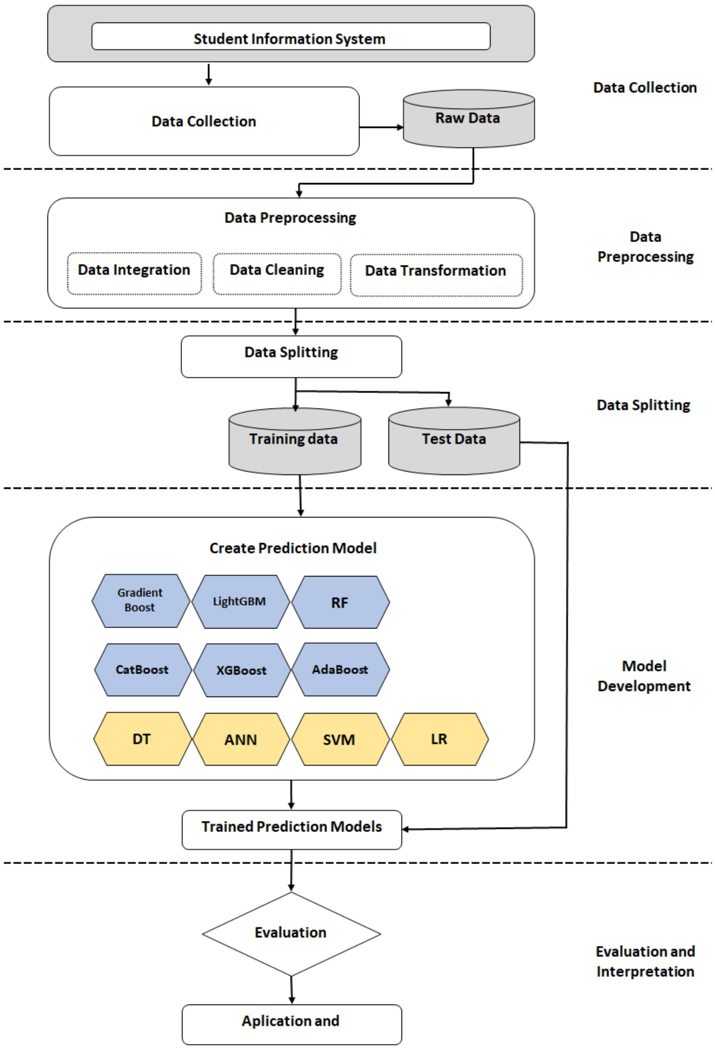

Figure 1. Workflow for predicting the graduation of associate degree students.

C. Data Splitting

The data was divided into two groups, 70% for training and the rest for testing. The training set was used for the model to learn the relationships and patterns in the data. The remaining 30% for testing was reserved for evaluating the effectiveness of binary classification performance on data that the model had never encountered before. In the data splitting process, the random_state=42 parameter was used to fairly compare the models with the same training and testing data.

D. Model Development

In this study, the performance efficiency of binary classification using ensemble and traditional machine learning models was investigated across three different academic time periods to predict whether students enrolled in associate degree programs at a public university will graduate on time. The three academic time periods are as follows:

1) Ensemble Learning Methods

- AdaBoost (Adaptive Boosting): It creates a strong model that iteratively improves performance by combining weak learners (usually decision trees) and assigning higher weights to misclassified examples. It is widely used for classification and regression problems [21].
- CatBoost: Optimized for categorical data and can automatically process categorical features without user intervention. It is preferred for classification, regression and time series analysis due to its speed and accuracy [22].
- LightGBM (Light Gradient Boosting Machine): A fast and lightweight model that uses a histogram-based approach optimized for large datasets and high-dimensional features. Therefore, it is frequently used in classification and regression tasks [23].
- Gradient Boosting: Used in classification and regression problems that minimizes error rates by optimizing the loss function and sequentially combining weak learners [24].
- XGBoost (Extreme Gradient Boosting): Developed with optimizations for speed and efficiency, especially used in classification, regression and ranking tasks on large datasets [25].
- Random Forest: Used for both classification and regression tasks, consisting of multiple decision trees, capable of generalization and resistant to overfitting [26].

2) Traditional Machine Learning Models:

- Logistic Regression (LR): A linear model widely used for binary classification problems [27].
- Artificial Neural Networks (ANN): A multi-layer neural network inspired by biological neural networks used for both classification and regression tasks for complex and large data sets [28].
- Support Vector Machines (SVM): Uses hyperplanes to separate data. It can achieve high accuracy in non-linear problems using kernel functions. Applicable for both classification and regression tasks [29].
- Decision Tree (DT): A model that branches data for classification or regression. It is easy to interpret and provides quick results for small datasets [30].

The hyperparameters of the model were determined based on values commonly recommended in the relevant literature and insights gained from previous studies. In this study, cross-validation or automated hyperparameter optimization techniques (such as Grid Search, Random Search, and Bayesian Optimization) were not used, and the parameters of the models employed are presented in detail in Table 3.

Table 3. The parameters of the models.

Model	Important Parameters
AdaBoost	n_estimators=100, learning_rate=0.1, algorithm='SAMME.R', random_state=42
CatBoost	iterations=100, learning_rate=0.1, depth=3, random_seed=42, verbose=0
LightGBM	n_estimators=100, learning_rate=0.3, random_state=42
Gradient Boost	n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42
XGBoost	n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42
Random Forest	n_estimators=100, criterion='gini', random_state=42
Logistic Regression	penalty='l2', C=1.0, solver='lbfgs', random_state=42
ANN	hidden_layer_sizes=(100,), activation='relu', solver='adam',batch_size='auto', learning_rate_init=0.001, random_state=42
SVM	C=1.0, kernel='rbf', degree=3, gamma='scale', random_state=42
Decision Tree	criterion='gini', splitter='best', random_state=42

Ensemble and classical machine learning models were trained and tested using a computer equipped with an 13th Gen Intel® Core™ i5-13500H processor (2.60 GHz), Windows 11 IoT Enterprise LTSC operating system, 16 GB RAM, NVIDIA GeForce RTX 4050 graphics card, and a 64-bit architecture, utilizing only CPU resources. The classification models were developed in Python using the Pandas, Keras, and Numpy libraries within the Anaconda Navigator environment. The training dataset was used to train the models, followed by evaluation on a separate test dataset.

E. Model Evaluation

The performance of the models was evaluated using several classification metrics:

- Accuracy: This metric represents the proportion of correctly predicted instances (both true positives and true negatives) relative to the total number of instances in the dataset. It is computed using the formula:

  Accuracy = (TP + TN) / (TP + TN + FP + FN)

- Precision: Precision indicates the ratio of correctly predicted positive instances to all instances classified as positive. It is calculated as:

  Precision = TP / (TP + FP)

- Recall (Sensitivity or True Positive Rate): Recall measures the model's ability to correctly identify actual positive instances. It is determined by:

  Recall = TP / (TP + FN)

- F1-Score: The F1-Score is the harmonic mean of precision and recall, offering a balanced evaluation of the model's performance. It is derived using:

  F1-Score = (2 * Precision * Recall) / (Precision + Recall)

- Total Time: Total time refers to the entire duration from the initialization and training of a machine learning model to making predictions on test data, measured in seconds.

IV.EXPERIMENTAL RESULTS AND DISCUSSION

In this study, the performance of models used for predicting on-time graduation was evaluated at three different prediction times: at the beginning of the first term, at the end of the first term, and at the end of the second term. The comparison between ensemble learning methods and classical machine learning methods was conducted using classification metrics, namely accuracy, recall, precision, and F1-score.

In the binary classification dataset, 38% of the 3171 students (1220 students) graduated on time, while 62% (1951 students) did not graduate on time. The dataset was split into 70% training (2219 students) and 30% testing (952 students). In the training set, 866 students (39%) graduated on time, whereas 1353 students (61%) did not. In the test set, 354 students (37%) graduated on time, while 598 students (63%) did not graduate on time.

The experimental results and overall evaluation for these three time periods are presented as subsections.

A. Prediction Results at the Beginning of the First Term

Prediction results obtained at the beginning of the first term using demographic information, high school performance, and entrance exam results are presented in the Table 4.

At the beginning of the first term, ensemble methods such as XGBoost, GradientBoost, CatBoost, and AdaBoost achieved the highest accuracy rates, with 67.6%, 67.3%, 67.2%, and 67.0%, respectively. However, when evaluated based on the F1-score, Logistic Regression (LR) demonstrated the most balanced performance, achieving the highest F1-score of 51.7%. Despite their high accuracy, ensemble methods exhibited lower F1-scores due to their low recall values. This indicates that these models struggled to identify positive cases, specifically students who graduated on time. LightGBM (51.6%) and Artificial Neural Networks (ANN) (51.4%) achieved the second and third highest F1-scores, demonstrating a balanced performance. Decision Trees (DT) showed the weakest performance, with an accuracy of 58.6% and an F1-score of 44.7%. Overall, while ensemble methods yielded higher accuracy, traditional models such as LR and ANN provided a more balanced performance in terms of the F1-score. Therefore, they are better suited for tasks requiring a balance between false positives and false negatives. The confusion matrices of LR, which achieved the highest F1-score among traditional models, and LightGBM, which had the second-best F1-score among ensemble models, are presented in Figure 2.

Table 4. Prediction results at the beginning of the first term.

Models	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)	Total Time (second)
AdaBoost	67.0	47.3	58.3	39.8	0.467
ANN	63.9	51.4	51.4	**51.4**	**1.850**
CatBoost	67.2	48.5	58.3	41.5	0.317
DT	58.6	44.7	44.4	44.9	0.041
GradientBoost	67.3	49.1	58.4	42.4	0.628
LightGBM	64.4	51.6	52.2	51.1	0.394
LR	66.8	**51.7**	56.3	47.7	**0.026**
RF	65.5	47.9	54.7	42.7	0.675
SVM	66.4	47.2	56.7	40.4	0.418
XGBoost	**67.6**	49.5	**59.0**	42.7	0.110

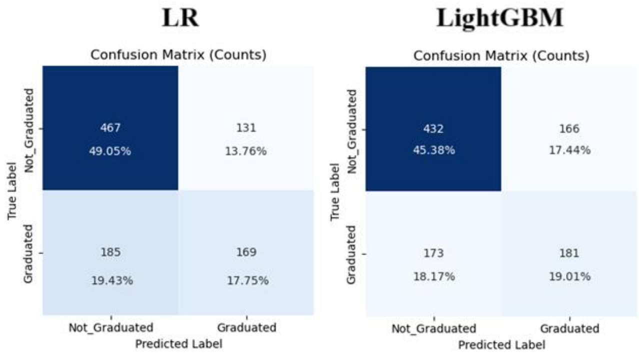

Figure 2. The confusion matrices of LR and LightGBM.

The total time results demonstrate that while the Artificial Neural Network (ANN) model is the slowest at 1.850 seconds, simpler models like Logistic Regression (LR, 0.026s) and Decision Tree (DT, 0.041s) operate significantly faster. Notably, XGBoost emerges as the most balanced performer, achieving the highest accuracy (67.6%) within a relatively short duration (0.110s).

B. Prediction Results at the End of the First Term

The results obtained when demographic information, high school performance, entrance exam data, and first-term academic performance data were included are presented in the Table 5.

Table 5. Prediction results at the end of the first term.

Models	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)	Total Time (second)
AdaBoost	77.0	69.9	68.1	71.8	0.464
ANN	76.2	68.1	67.8	68.4	**1.862**
CatBoost	78.3	**71.7**	69.5	**74.0**	0.334
DT	71.5	63.4	60.7	66.4	0.047
GradientBoost	77.8	70.7	69.6	71.8	0.675
LightGBM	76.2	67.7	68.2	67.2	0.421
LR	**78.4**	70.5	**71.5**	69.5	**0.037**
RF	77.3	68.8	70.4	67.2	0.600
SVM	77.4	69.1	70.4	67.8	0.369
XGBoost	78.2	71.3	69.7	72.9	0.104

At the end of the first term, Logistic Regression (LR) achieved the highest accuracy rate of 78.4%. Ensemble methods such as CatBoost, XGBoost, and GradientBoost followed, with accuracy rates of 78.3%, 78.2%, and 77.8%, respectively. However, when evaluated based on the F1-score, CatBoost (71.7%) demonstrated the best performance, followed by XGBoost (71.3%) and GradientBoost (70.7%). These ensemble methods exhibited high precision and recall values, indicating their effectiveness in identifying students who graduated on time. On the other hand, Decision Trees (DT) showed the weakest performance, with an accuracy of 71.5% and an F1-score of 63.4%, maintaining its low performance from the beginning of the term. Overall, although ensemble methods performed well in terms of accuracy and F1-score, traditional models such as LR and ANN provided more stable results. The confusion matrices of CatBoost, the ensemble model with the highest F1-score, and LR, the best-performing traditional model, are presented in Figure 3.

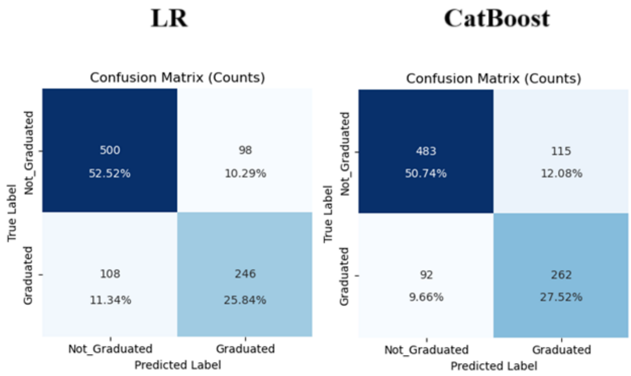

Figure 3. The confusion matrices of LR and CatBoost.

The total time analysis reveals ANN as the slowest model (1.862s), while Logistic Regression (LR, 0.037s) and Decision Tree (DT, 0.047s) maintain their speed advantage. CatBoost demonstrates the best accuracy-precision-recall balance (78.3%, 69.5%, 74.0%) with competitive speed (0.334s), whereas XGBoost (78.2%, 0.104s) again shows remarkable efficiency.

C. Results at the End of the Second Term

The results obtained by adding second-term academic performance data to the first-term data are presented in the Table 6.

Table 6. Prediction results at the end of the second term.

Models	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)	Total Time (second)
AdaBoost	79.5	74.0	70.0	78.5	0.498
ANN	77.4	70.4	68.6	72.3	**1.884**
CatBoost	79.9	74.6	70.5	79.1	0.332
DT	76.7	69.0	68.2	69.8	0.035
GradientBoost	79.4	74.3	69.4	79.9	0.763
LightGBM	78.0	71.6	69.0	74.3	0.419
LR	79.7	73.0	72.3	73.7	**0.041**
RF	**80.9**	75.1	**72.8**	77.7	0.576
SVM	80.1	74.0	72.1	76.0	0.315
XGBoost	80.4	**75.6**	70.3	**81.6**	0.140

At the end of the second term, the highest accuracy, as in the first term, was achieved by Random Forest (RF) with 80.9%, followed closely by XGBoost (80.4%), Support Vector Machines (SVM) (80.1%), and CatBoost (79.9%). When evaluated based on the F1 score, XGBoost (75.6%) achieved the highest value, followed by RF (75.1%), CatBoost (74.6%), and GradientBoost (74.3%). These results demonstrate that ensemble models exhibited both high precision and recall values, showing their effectiveness in identifying graduating students. Among traditional models, Logistic Regression (LR) continued to show balanced performance with 79.7% accuracy and a 73.0% F1 score. Decision Trees (DT) continued to exhibit the weakest performance. Overall, ensemble methods such as XGBoost, RF, and CatBoost continued to demonstrate strong predictive power in terms of both accuracy and F1 score. However, the LR traditional model provided more stable results, making it valuable for tasks requiring a balance between false positives and false negatives. The confusion matrices of XGBoost, the ensemble model with the highest F1-score, and LR, the best-performing traditional model, are presented in Figure 4.

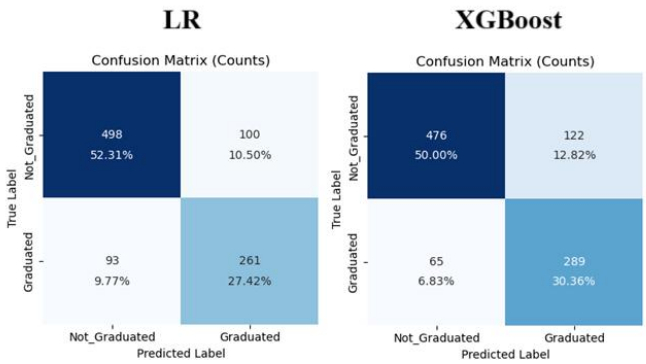

Figure 4. The confusion matrices of LR and XGBoost.

The total time analysis at the end of the Second Period reveals Artificial Neural Network (ANN) remains the slowest model (1.884s), while Decision Tree (DT: 0.035s) and Logistic Regression (LR: 0.041s) maintain their consistent speed advantages across all evaluation periods.

D. Overall Evaluation

In the general evaluation, models were assessed based on F1 scores due to the class imbalance in the dataset. The F1 score (the harmonic mean of precision and recall) was preferred because correctly classifying the minority class, the students who did not graduate on time (38%), is crucial in our imbalanced dataset. The F1 scores of the models are presented in Figure 5.

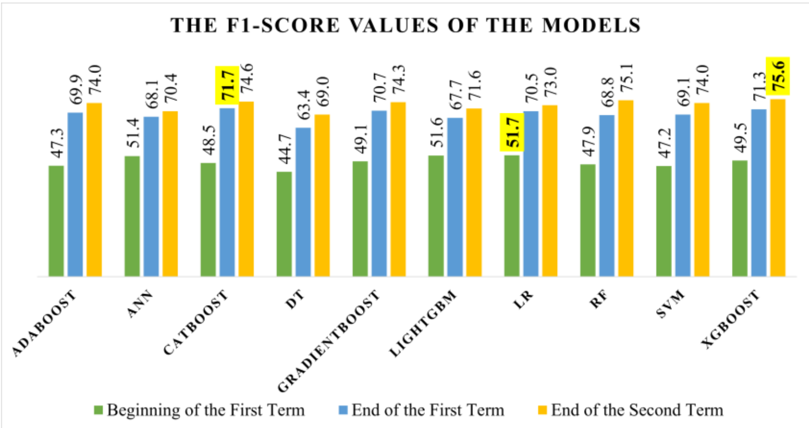

Figure 5. The accuracy values of the models.

According to Figure 5, ensemble models exhibit higher performance compared to traditional models in terms of F1 score. At the beginning of the first term, traditional models (ANN, DT, LR, SVM) had higher initial values, while ensemble models (AdaBoost, CatBoost, GradientBoost, LightGBM, RF, XGBoost) showed a better performance increase in subsequent time periods. At the end of the first term, the traditional LR model (51.7) achieved the highest performance. However, by the end of the first and second terms, the ensemble models CatBoost (71.7) and XGBoost (75.6) reached the highest F1 scores. The performance of traditional models remained relatively limited. As a result, ensemble models—particularly with the addition of new academic performance features—produced more successful results. Moreover, the inclusion of new features at each prediction stage was a significant factor in improving the F1 scores of all models.

Additionally, the distribution of total times of all models according to the three academic time periods is given in Figure 6. The three-period performance data clearly reveal variations in the computational times of the models. While LR (Logistic Regression) and DT (Decision Tree) consistently remain the fastest models (0.03-0.05s), ANN (Artificial Neural Network) demonstrates the slowest performance (1.85-1.88s). CatBoost and LightGBM emerged as the most stable models (±3% fluctuation), while XGBoost attracted attention with its balanced performance (0.10-0.14s).

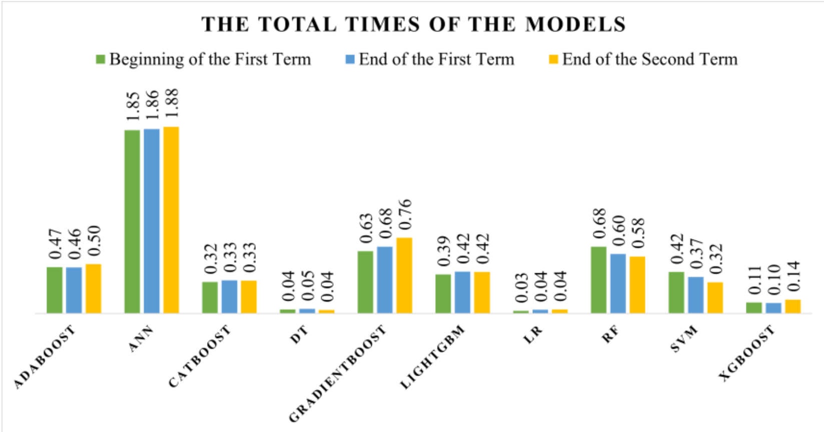

Figure 6. The total times of the models.

V.CONCLUSION

The experimental findings of this study demonstrate that the classification performance of both traditional and ensemble-based machine learning models improves with the inclusion of within-term academic features. In the initial prediction phase, traditional models such as Logistic Regression and Artificial Neural Networks produced more balanced results, whereas in subsequent stages, ensemble learning methods like XGBoost, CatBoost, and GradientBoost achieved higher success after the integration of within-term performance data.

Given that the minority class constituted 38% of the dataset, the F1-score was prioritized as the evaluation metric. The results revealed that traditional models performed better in early-stage predictions, while ensemble learning methods exhibited superior performance in later stages enriched with additional features. These findings highlight the suitability of traditional methods for early predictions, while ensemble models become more effective as data volume increases.

The analysis of model total times reveals that Logistic Regression (LR) and Decision Trees (DT) remain the optimal choices for low-latency applications, while CatBoost and XGBoost demonstrate greater suitability for large-scale projects due to their balanced performance. In contrast, Artificial Neural Networks (ANN) are computationally prohibitive in resource-constrained environments, given their significantly higher runtime requirements.

This study contributes to the literature by examining the role of machine learning in predicting on-time graduation for associate degree students across different stages. The findings may support the development of early intervention strategies for at-risk students. Future research aims to enhance prediction performance through optimization techniques, hybrid models, or deep learning approaches. Additionally, incorporating supplementary variables such as psychosocial factors and classroom interaction data could further improve early-stage prediction accuracy.

REFERENCES

[1] Bako, H. S., Ambursa, F. U., Galadanci, B. S., Garba, M. (2023). Predicting Timely Graduation of Postgraduate Students Using Random Forests Ensemble Method. *Fudma Journal of Sciences*, 7(3), 177–185.

[2] Rismayati, R., Ismarmiaty, I., Hidayat, S. (2022). Ensemble Implementation for Predicting Student Graduation with Classification Algorithm. *International Journal of Engineering and Computer Science Applications (IJECSA)*, 1(1), 35–42.

[3] Pandey, M., Taruna, S. (2014). A Comparative Study of Ensemble Methods for Students' Performance Modeling. *International Journal of Computer Applications*, 103(8), 26–32.

[4] Kang, Z. (2019). Using Machine Learning Algorithms to Predict First-generation College Students' Six-year Graduation: A Case Study. *International Journal of Information Technology and Computer Science*, 11(9), 1–8.

[5] Desfiandi, A., Soewito, B. (2023). Student Graduation Time Prediction Using Logistic Regression, Decision Tree, Support Vector Machine, and Adaboost Ensemble Learning. *IJISCS (International Journal of Information System and Computer Science)*, 7(3), 195.

[6] Law, T.-J., Ting, C.-Y., Ng, H., Goh, H.-N., Quek, A. (2024). Ensemble-SMOTE: Mitigating Class Imbalance in Graduate on Time Detection. *Journal of Informatics and Web Engineering*, 3(2), 229–250.

[7] Ananto, N. (2024). Leveraging Ensemble Learning for Predicting Student Graduation: A Data Mining Approach. *Prosiding Seminar Nasional Forum Manajemen Indonesia - e-ISSN 3026-4499*, 2(353–365.

[8] Anggrawan, A., Hairani, H., Satria, C. (2023). Improving SVM Classification Performance on Unbalanced Student Graduation Time Data Using SMOTE. *International Journal of Information and Education Technology*, 13(2), 289–295.
[9] Ariska, F., Sihombing, V., Irmayani, I. (2022). Student Graduation Predictions Using Comparison of C5.0 Algorithm With Linear Regression. *SinkrOn*, 7(1), 256–266.
[10] Hanafi, W., Chrisnanto, Y. H., Ningsih, A. K. (2023). Student Graduation Prediction System Based on Academic And Non-Academic (Eq) Data Using C4.5 Algorithm. *JUMANJI (Jurnal Masyarakat Informatika Unjani)*, 7(1), 64.
[11] Kurniawan, D., Anggrawan, A., Hairani, H. (2020). Graduation Prediction System On Students Using C4.5 Algorithm. *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 19(2), 358–365.
[12] Lagman, A. C., Alfonso, L. P., Goh, M. L. I., Lalata, J. P., Magcuyao, J. P. H., Vicente, H. N. (2020). Classification Algorithm Accuracy Improvement for Student Graduation Prediction Using Ensemble Model. *International Journal of Information and Education Technology*, 10(10), 723–727.
[13] Mehta, S. (2023). Playing Smart with Numbers: Predicting Student Graduation Using the Magic of Naive Bayes. *International Transactions on Artificial Intelligence (ITALIC)*, 2(1), 60–75.
[14] Muliani, S. S., Sihombing, V., Munthe, I. R. (2024). Implementation of Exploratory Data Analysis and Artificial Neural Networks to Predict Student Graduation on-Time. *SinkrOn*, 8(2), 1188–1199.
[15] Rachmawati, D. A., Ibadurrahman, N. A., Zeniarja, J., Hendriyanto, N. (2023). Implementation of the Random Forest Algorithm in Classifying the Accuracy of Graduation Time for Computer Engineering Students at Dian Nuswantoro University. *Jurnal Teknik Informatika (Jutif)*, 4(3), 565–572.
[16] Samuel, Y. T., Hutapea, J. J., Jonathan, B. (2019). Predicting the Timeliness of Student Graduation Using Decision Tree C4.5 Algorithm in Universitas Advent Indonesia. In: *2019 12th International Conference on Information & Communication Technology and System (ICTS)*. IEEE, Surabaya, Indonesia, 276–280
[17] Tholib, A., Fadli Hidayat, M. N., Yono, S., Wulanningrum, R., Daniati, E. (2023). Comparison of C4.5 and Naive Bayes for Predicting Student Graduation Using Machine Learning Algorithms. *International Journal of Engineering and Computer Science Applications (IJECSA)*, 2(2), 65–72.
[18] Riadi, I., Umar, R., Anggara, R. (2024). Comparative Analysis of Naive Bayes and K-NN Approaches to Predict Timely Graduation using Academic History. *International Journal of Computing and Digital Systems*, 15(1), 1163–1174.
[19] Allen, R. E., Diaz, C., Gant, K., Taylor, A., Onor, I. (2016). Preadmission Predictors of On-time Graduation in a Doctor of Pharmacy Program. *American Journal of Pharmaceutical Education*, 80(3), 43.
[20] Lovelace, M. D., Reschly, A. L., Appleton, J. J. (2017). Beyond School Records: The Value of Cognitive and Affective Engagement in Predicting Dropout and On-Time Graduation. *Professional School Counseling*, 21(1), 1096-2409–21.1.70.
[21] Freund, Y., Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), 119–139.
[22] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., Gulin, A. (2017). CatBoost: unbiased boosting with categorical features. https://doi.org/10.48550/ARXIV.1706.09516
[23] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: *31st Conference on Neural Information Processing Systems (NIPS 2017)*. Long Beach, CA, USA, 1–9
[24] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.
[25] Chen, T., Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, San Francisco California USA, 785–794.
[26] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
[27] Hosmer, D. W., Lemeshow, S. (2000). Applied Logistic Regression, 1st ed, Wiley, New York.
[28] Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep learning, The MIT press, Cambridge.
[29] Cortes, C., Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
[30] Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. (2017). Classification And Regression Trees, 1st ed, Routledge, New York.