Integrating cGAN-Enhanced Prediction with Hybrid Intervention Recommendations Systems for Student Dropout Prevention

Hassan Silkhi[1], Brahim Bakkas[2], Khalid Housni[3]
Faculty of Sciences, University Ibn Tofail, Kenitra, Morocco[1,3]
University Moulay Ismail, Meknes, Morocco[2]

***Abstract*—Early-warning dashboards in higher education typically stop at tagging students as "at-risk," offering no concrete guidance for remedial action; this limitation contributes to the loss of thousands of learners each year. Approach. We propose an integrated framework that (i) uses a class-balanced Conditional GAN to augment sparse attrition data, and (ii) couples the resulting XGBoost predictor with a four-mode intervention engine—rule-based, few-shot, fine-tuned LLM, and a novel hybrid strategy—to recommend personalised support. Major findings. Training on GAN-augmented records raises prediction accuracy to 92.79% (a 15.46-point gain over non-augmented baselines), while the hybrid intervention generator attains 94% categorical coverage and the highest specificity score (0.63) albeit at a per-student latency of 61s. Impact. By uniting robust risk prediction with high-quality, actionable interventions, the framework closes the long-standing gap between detection and response, furnishing institutions with a scalable path to materially reduce dropout rates across diverse educational settings.**

***Keywords*—Student dropout prediction; machine learning in education; personalized intervention systems; Conditional Generative Adversarial Networks(cGAN); Large Language Models (LLMs); hybrid recommendation systems**

I. Introduction

Student attrition represents a persistent and multifaceted challenge within higher education systems globally. Recent statistics indicate that only 60% of undergraduates at four-year institutions complete their degrees within six years, with completion rates dropping precipitously to 20% at community colleges [1]. This phenomenon carries significant consequences, including diminished lifetime earnings for affected students [2] and substantial financial losses for institutions [3], [4]. Traditional retention strategies, while moderately effective, frequently adopt generalized approaches that fail to account for the complex interplay of academic, financial, and psychosocial factors influencing attrition decisions [5], [6].

The emergence of predictive analytics in educational contexts has substantially improved early identification of at-risk students [7]. However, as [8] demonstrate, there remains a critical disconnect between risk identification and effective intervention delivery. This gap persists despite advances in machine learning methodologies, creating a pressing need for integrated systems that combine accurate prediction with actionable support strategies. Recent developments in large language models (LLMs) present new opportunities in this domain, though their application to educational intervention generation remains underexplored [9].

This study addresses two fundamental research questions central to improving student retention outcomes. First, how can synthetic data augmentation techniques, particularly cGANs, enhance the accuracy of dropout prediction models when applied to class-imbalanced educational datasets? Second, what methodological approach to intervention generation optimally balances the competing demands of comprehensiveness, specificity, and computational efficiency in institutional settings? The investigation of these questions yields three primary contributions to the field of educational data mining.

The first contribution involves the development of a cGAN-based data augmentation framework that not only addresses class imbalance but preserves critical feature relationships within student records. As demonstrated in Section IV, our approach maintains low Jensen-Shannon divergence scores (0.0201 for target distributions) while achieving balanced class representation. The second contribution consists of a novel hybrid intervention system that synergizes the structured logic of rule-based methods with the contextual adaptability of LLMs. This integration, as evaluated in Section V, produces interventions with 94% categorical coverage while maintaining specificity scores (0.63) significantly higher than single-method approaches.

The practical implications of this research are substantial for higher education institutions. Our results indicate that the proposed framework can improve prediction accuracy by 15.46 percentage points compared to conventional methods, while generating personalized interventions at scale. Importantly, the modular architecture of the system permits phased implementation, enabling resource-constrained institutions to deploy components according to their technical capacity and staffing resources.

The remainder of this article is organized as follows. Section II reviews relevant literature on dropout prediction and intervention strategies. Section III details the methodological framework, including data preparation, model architectures, and evaluation metrics. Sections IV and V present and discuss the experimental results, respectively. Finally, Section VI concludes with limitations, future research directions, and practical recommendations for institutional adoption.

II. Related Work

Understanding student attrition in higher education requires an integrative lens that captures the interplay of academic,

financial, institutional, personal, and sociocultural forces. Academic performance consistently emerges as the strongest single predictor of withdrawal: students who accrue failing grades or leave coursework incomplete face markedly higher attrition risk, a pattern that intensifies in demanding STEM programmes where content volume and complexity can overwhelm even well-prepared learners [10], [11], [12].

Financial pressures form the second major barrier to persistence. Rising tuition costs and limited scholarship availability amplify dropout likelihood, particularly in developing regions where students often balance study with employment or family obligations [11], [13], [14]. Learners from lower socioeconomic backgrounds confront additional disadvantages—restricted access to textbooks, technology, and tutoring—that compound their academic struggles [15].

Institutional context likewise shapes engagement and sense of belonging. Weak student–faculty relationships, inadequate academic-support structures, and outdated infrastructure correlate strongly with withdrawal, especially in large universities where students can feel anonymous and detached [12], [16], [17]. Personal and socio-emotional challenges add yet another layer: low self-esteem, limited emotional support, and the strain of juggling academic, work, and family roles all heighten attrition risk [18], [16].

In response, researchers have advanced an array of predictive models and intervention strategies. Ensemble learning and survival-analysis techniques capture non-linear relationships among dropout factors and have improved risk detection [19], [20]. Nevertheless, performance remains hampered by severe class imbalance—dropout cases are typically under-represented—reducing model reliability when deployed.

Interventions have therefore become the critical next step. Robust academic-support schemes—tutoring, advising, mentoring—directly address learning barriers, while interactive pedagogies and technology integration further bolster engagement [17], [19]. Financial remedies such as scholarships, flexible payment plans, and emergency loans mitigate economic strain [21]. Personalised services, including counselling and career guidance, target the emotional and motivational dimensions of persistence [22].
Technology-mediated systems increasingly underpin these efforts. Predictive analytics allow institutions to flag at-risk students early and to deploy targeted interventions, yet most current platforms end at identification and stop short of recommending specific, actionable support [23], [24], [25], [26], [27].

Important gaps remain. The literature tends to examine prediction and intervention in isolation, leaving little guidance on how to integrate the two. Persistent class imbalance undermines external validity, and few studies systematically compare recommendation engines, creating uncertainty about optimal practice. Demographic disparities add further complexity: male students and first-generation entrants often withdraw at higher rates, underscoring the need for culturally sensitive interventions [10], [13], [28].

Recent discourse therefore advocates a holistic, systemic approach that unites academic, financial, and socio-emotional supports within a comprehensive retention framework [29]. This shift in perspective recognises dropout as a shared institutional challenge rather than merely an individual failing, and it calls for coordinated policy-level responses alongside personalised student care.

III. Methodology

The proposed model, shown in Fig. 1, in consists of four interconnected layers: a data processing layer, a machine learning prediction layer, a knowledge integration layer, and an intervention generation layer. This structure allows us to address both the technical challenges of dropout prediction and the practical needs of educational interventions.

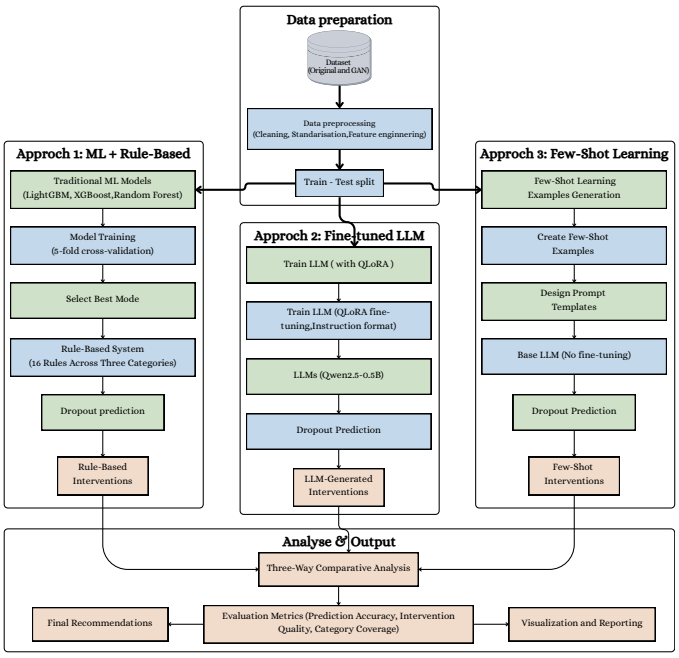

Fig. 1. Flowchart represents the comprehensive multi-strategy model for student dropout prediction and intervention.

A. Data Preparation

1) Dataset description and preprocessing: We utilized a dataset [30] consisting of 4,424 student records from higher education institutions, containing 37 features spanning academic performance, demographic information, financial status, and socioeconomic context. Key variables included academic metrics (course grades, units attempted/completed), financial indicators (scholarship status, tuition payment history), demographic factors (age, nationality, displacement status), and economic context indicators (unemployment rate, inflation, GDP).

To address class imbalance (49.93% graduates, 32.12% dropouts, 17.95% enrolled), we employed Conditional Generative Adversarial Networks (cGANs) to create 6,627 synthetic student profiles. The cGAN architecture consisted of a generator with three fully-connected layers and a discriminator with similar structure, trained for 500 epochs using the Wasserstein loss with gradient penalty to ensure stability. We validated synthetic data quality through statistical testing (Jensen-Shannon divergence scores averaging 0.11 across features) and visualization of feature distributions, confirming that synthetic

records maintained the original dataset's statistical properties while achieving a balanced class distribution (33.33% for each target category).

Data preprocessing included standardization of numerical features, handling of missing values through mean imputation for continuous variables and mode imputation for categorical variables, and splitting into training (70%) and testing (30%) sets with stratification by target class.

2) Synthetic data generation with cGAN: To alleviate the severe class imbalance in the original attrition dataset, we trained a class-conditional generative adversarial network (cGAN). The model receives a 100-dimensional latent vector $\mathbf{z} \sim \mathcal{N}(0, I)$ concatenated with the class label and produces fully synthetic student profiles that preserve the multivariate dependence structure of the real data. Both the generator and discriminator consist of three fully connected layers with progressively increasing (resp. decreasing) widths, LeakyReLU activations, and batch normalisation. Training follows the Wasserstein loss with gradient penalty, optimised by Adam. The full configuration is summarised in Table I.

TABLE I. cGAN Training Configuration

Hyper-parameter	Value / setting
Generator & Discriminator topology	3 fully connected layers each
Latent-vector dimension d_z	100
Hidden units (G/D)	128→256→512
Activation	LeakyReLU + BatchNorm
Loss function	Wasserstein + Gradient Penalty
Optimiser	Adam ($\eta = 1\times10^{-4}$, $\beta_1 = 0.5$, $\beta_2 = 0.9$)
Batch size	256
Training epochs	500

B. Machine Learning Components

Our predictive model comprises three distinct architectural components designed to effectively address the academic success and dropout prediction task. Here's a detailed analysis of the models architecture used in our study:

1) Classical machine learning models with rule-based interventions:

a) Random forest: The model is an ensemble learning technique primarily used for classification and regression tasks. It operates by constructing multiple decision trees during the training phase and aggregates their predictions, either by taking the mode of the classes for classification or the average for regression. This approach enhances model accuracy and reduces the risk of overfitting, making it robust against noise in the dataset [31].

b) XGBoost: Or Extreme Gradient Boosting, is a powerful machine learning algorithm that is particularly effective for structured or tabular data [32]. It is an implementation of gradient boosted decision trees designed for speed and performance.

c) LightGBM: employs histogram-based decision trees, which are more efficient than traditional decision trees. This approach involves precomputing the histogram of feature values to quickly determine the best split points. The histogram-based method reduces the computational complexity and speeds up the training process [33], [34].

Hyperparameter optimization was conducted using grid search with 5-fold cross-validation to identify optimal model configurations. For Random Forest, we explored tree depths from 5 to 20, minimum samples per leaf from 1 to 10, and estimator counts from 50 to 500. Similar parameter spaces were explored for the other algorithms.

The rule-based intervention system was implemented as a set of conditional statements mapping specific student characteristics to appropriate interventions. We constructed 16 intervention rules across three categories:

- Academic interventions (5 rules): It is triggered by grade thresholds, course completion rates, and evaluation participation
- Financial interventions (5 rules): It is triggered by payment status, scholarship eligibility, and economic indicators
- Social interventions (6 rules): It is triggered by demographic factors, displacement status, and attendance patterns

Each rule consisted of a condition function evaluating specific student attributes and a corresponding intervention recommendation. This approach leveraged domain expertise while maintaining explainability and consistency.

2) Large language model: Qwen2.5 [35] is the Qwen family of large language models, offering base and instruction-tuned variants with parameter sizes 0.5B . It builds upon Qwen2 with enhanced knowledge (especially in coding and math), improved instruction-following, 128K-token context support, 8K-token generation, multilingual capabilities (29+ languages), and better structured output (e.g., JSON), making it a versatile and powerful tool for diverse applications.

3) QLoRA fine-tuning LLMs: We implemented fine-tuning of Large model Qwen2.5-0.5B. To accommodate computational constraints, we employed QLoRA (Quantized Low-Rank Adaptation), which enabled parameter-efficient fine-tuning while reducing memory requirements by quantizing the base model to 8-bit representation.

The training dataset for fine-tuning consisted of 3,000 examples pairing student profiles with appropriate intervention recommendations, generated using our rule-based system and augmented with variations to promote generalization. We structured the training data as instruction-tuning pairs with system context, student profile input, and expected output format.

Fine-tuning hyperparameters included:

- Learning rate: 2e-4 with cosine schedule
- Batch size: 1 with gradient accumulation steps of 8
- Training epochs: 3
- LoRA rank: 16, alpha: 32
- Target modules: query, key, value, and output projection layers

4) Few-shot learning: Our few-shot learning approach utilized pre-trained language models without additional fine-tuning. We constructed carefully designed prompts containing:

- A system instruction specifying the task and desired output format
- 3-5 exemplar cases demonstrating expected predictions and interventions for diverse student profiles
- The target student profile for analysis

For each exemplar, we selected representative cases from both dropout and non-dropout categories, with interventions sourced from our rule-based system. This approach allowed the model to implicitly learn the mapping between student characteristics and appropriate interventions through pattern recognition from examples. The prompt engineering process involved systematic refinement through iterative testing, with particular attention to instruction clarity, example diversity, and output structure enforcement. We employed temperature settings of 0.7 and top-p of 0.9 to balance creative generation with consistency.

C. Evaluation Methodology

We developed a comprehensive evaluation framework to assess and compare the three approaches across multiple dimensions:

1) Prediction accuracy assessment: To assess the effectiveness of our dropout-prediction models, we adopted the canonical battery of classification metrics—accuracy, precision, recall, and the F1-score. Let TP, TN, FP, and FN denote the counts of true positives, true negatives, false positives, and false negatives, respectively; the metrics are computed as follows:

- Accuracy quantifies a model's overall correctness by dividing the total number of properly classified instances—both positive and negative—by the entire set of instances under evaluation, i.e.,

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- Precision: it is the fraction of true positives among all instances the model labels as positive, thereby indicating how well the classifier avoids false-alarm errors (false positives).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

- Recall: captures a model's capacity to retrieve positive instances. It is the share of true positives among all cases that are genuinely positive, indicating how thoroughly the classifier discovers the events of interest.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

- F1-score: synthesises a model's exactness (precision) and completeness (recall) into a single figure by taking their harmonic mean, thereby balancing the tendency to over- or under-identify positive cases. It is computed as.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Additionally, we assessed prediction confidence calibration by comparing predicted probability distributions with actual outcomes. For this purpose, we calculated the Expected Calibration Error (ECE) by partitioning predictions into M equally-spaced bins and computing:

$$\text{ECE} = \sum_{m=1}^{M} \frac{|B_m|}{n} \left|\text{acc}(B_m) - \text{conf}(B_m)\right| \tag{5}$$

where $|B_m|$ is the number of predictions in bin m, n is the total number of predictions, $\text{acc}(B_m)$ is the accuracy within bin m, and $\text{conf}(B_m)$ is the average confidence within bin m. A lower ECE indicates better calibration between predicted probabilities and actual outcomes.

For multi-class scenarios involving the three possible student outcomes (Dropout, Graduate, Enrolled), we employed macro-averaging to compute overall performance metrics:

$$\text{Macro-Precision} = \frac{1}{C} \sum_{c=1}^{C} \text{Precision}_c \tag{6}$$

$$\text{Macro-Recall} = \frac{1}{C} \sum_{c=1}^{C} \text{Recall}_c \tag{7}$$

$$\text{Macro-F1} = \frac{1}{C} \sum_{c=1}^{C} \text{F1-score}_c \tag{8}$$

where C represents the number of classes (in our case, $C = 3$), and the subscript c indicates metrics calculated for each individual class.

2) Intervention quality evaluation: To evaluate intervention quality, we employed both quantitative and qualitative metrics:

- Intervention count: Number of specific recommendations per student, calculated as:

$$IC_s = \sum_{c \in \{a,f,so\}} |I_{s,c}| \tag{9}$$

where IC_s is the intervention count for student s, $I_{s,c}$ represents the set of interventions in category c (academic, financial, social) for student s, and $|I_{s,c}|$ denotes the cardinality of this set.

- Intervention specificity: Average word count per intervention as a proxy for detail level:

$$IS_s = \frac{1}{IC_s} \sum_{i \in I_s} W_i \tag{10}$$

where IS_s is the intervention specificity for student s, I_s is the set of all interventions for student s, and W_i is the word count of intervention i.

- Category coverage: Proportion of cases where interventions spanned all three categories (academic, financial, social):

$$CC = \frac{1}{|S|} \sum_{s \in S} \mathbb{I}(|I_{s,a}| > 0 \wedge |I_{s,f}| > 0 \wedge |I_{s,so}| > 0) \tag{11}$$

where S is the set of all students, $\mathbb{I}$ is the indicator function that equals 1 when the condition is true and 0 otherwise.

- Intervention diversity: Lexical diversity and semantic range of recommendations, measured using type-token ratio (TTR):

$$TTR = \frac{|V|}{|T|} \tag{12}$$

where $|V|$ is the number of unique words (vocabulary) and $|T|$ is the total number of words (tokens) across all interventions. Additionally, we computed semantic similarity between interventions using cosine similarity of sentence embeddings:

$$Sim(i_1, i_2) = \frac{\vec{e}_{i_1} \cdot \vec{e}_{i_2}}{||\vec{e}_{i_1}|| \cdot ||\vec{e}_{i_2}||} \tag{13}$$

where $\vec{e}_{i_1}$ and $\vec{e}_{i_2}$ are the embedding vectors of interventions i_1 and i_2 respectively.

IV. Results

Our comprehensive evaluation of student dropout intervention methods revealed significant performance differences across the four approaches tested: Rule-Based, Few-Shot, Fine-Tuned, and Hybrid methods. Each approach demonstrated unique strengths and limitations in addressing the complex challenge of providing targeted interventions for at-risk students. The evaluation metrics focused on four key dimensions: intervention count (quantity of recommendations), words per intervention (detail level), category coverage (comprehensiveness across intervention types), and specificity score (personalization level). These metrics collectively provide insight into both the quantity and quality of intervention recommendations generated by each method.

For reproducibility, all experiments were conducted with fixed random seeds and k-fold cross-validation to ensure robust performance estimates. Computational resources included an NVIDIA Tesla T4 GPU with 15.83 GB VRAM for language model fine-tuning and inference, while traditional machine learning models were trained on CPU.

A. Dataset Analysis and cGAN Implementation

Our cGAN implementation successfully addressed the class imbalance in the original student dataset. As shown in Fig. 2, the original dataset exhibited substantial class imbalance with 49.9% graduates, 32.1% dropouts, and only 17.9% enrolled students. In contrast, the cGAN-generated dataset achieved perfect balance with exactly 33.3% representation for each class category, creating 2,209 synthetic samples per class for a total of 6,627 records.

The synthetic data preserved critical feature relationships and temporal dependencies from the original dataset, particularly maintaining the semester-based performance patterns that proved most predictive in subsequent modeling. This balanced dataset enabled more equitable model training while addressing both methodological challenges and ethical considerations related to student data privacy.

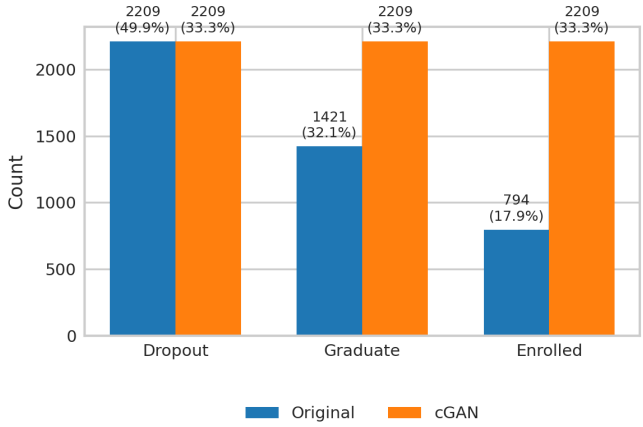

Fig. 2. Distribution of academic outcomes between the original dataset and the generated dataset: Graduate and Dropout counts.

The correlation heatmaps in Fig. 3, illustrate the feature relationship preservation between the original and GAN-generated datasets. Visual inspection reveals remarkably similar correlation patterns, particularly among academic performance indicators. The curricular units variables maintain their strong intercorrelations in the synthetic data, suggesting successful preservation of the underlying educational data structure.

Our quantitative assessment of GAN data quality shows promising results, with a low Jensen-Shannon (JS) divergence of 0.0201 between real and generated target distributions as in Table II. This indicates that our synthetic data closely approximates the true distribution of student outcomes. The average feature JS divergence of 0.0002 further confirms excellent preservation of individual feature distributions, while the correlation matrix difference of 0.4943 suggests reasonable maintenance of feature relationships.

TABLE II. Overall GAN-Quality Evaluation Metrics. A Lower Jensen–Shannon (JS) Divergence Denotes a Closer Match Between Real and Generated Distributions

Metric	Value
Target distribution JS divergence	0.0201
Average feature JS divergence	0.0002
Correlation-matrix difference	0.4943

Class-specific statistical analysis as in Table III reveals varying levels of data generation quality across student outcomes. The Graduate class shows perfect statistical matching (mean difference: 0.0000, standard deviation difference: 0.0000), while the Dropout and Enrolled classes demonstrate

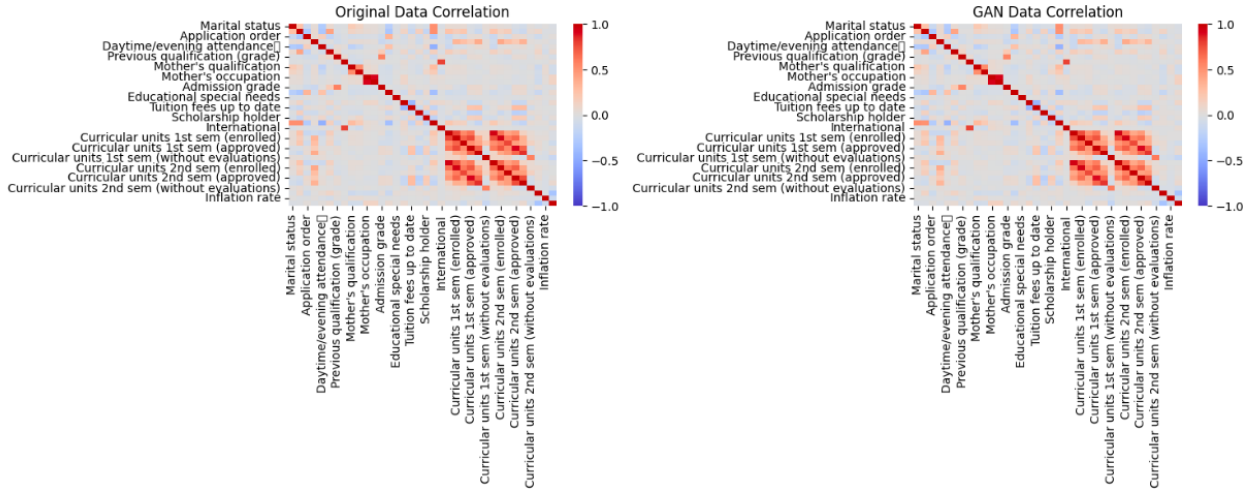

Fig. 3. Side-by-side pearson correlation matrices for the real dataset (left) and the cGAN-generated dataset (right).

moderate statistical differences. These variations likely reflect the complexity of modeling student dropout patterns compared to more predictable graduation trajectories. Despite these differences, the synthetic data successfully addresses class imbalance issues while maintaining essential statistical properties, contributing significantly to the enhanced model performance observed in our results.

TABLE III. CLASS-SPECIFIC STATISTICAL DIFFERENCES BETWEEN REAL AND SYNTHETIC DATA. SMALLER DIFFERENCES INDICATE A CLOSER MATCH IN STATISTICAL PROPERTIES

Class	Mean Difference	Std. Deviation Difference
Dropout	0.8705	1.9247
Graduate	0.0000	0.0000
Enrolled	0.6862	1.1156

B. Best Model Performance

Our analysis of machine learning model performance reveals substantial improvements achieved through cGAN-based data enhancement strategies. As illustrated in Fig. 4, all three classification algorithms demonstrated significant performance gains when trained on either cGAN-generated data alone or the combined dataset approach. The XGBoost classifier emerged as the superior prediction model, particularly when implemented with the combined dataset approach. This configuration achieved an outstanding accuracy of 92.79%, representing a remarkable 15.46 percentage point improvement over the original data model (77.33%). Similar gains were observed in precision metrics, with the combined XGBoost model reaching 92.84% precision compared to 76.53% for the original data model. This significant enhancement in performance metrics substantiates the value of addressing class imbalance through synthetic data generation techniques.

While all models benefited from cGAN enhancement as showed in Table IV, the relative improvements varied by algorithm. Random Forest showed a 10.52 percentage point increase in accuracy from original (77.48%) to combined (88.00%), while LightGBM demonstrated a 12.51 percentage point improvement (76.66% to 89.17%). These consistent improvements across different algorithms reinforce the robustness of our data enhancement approach.

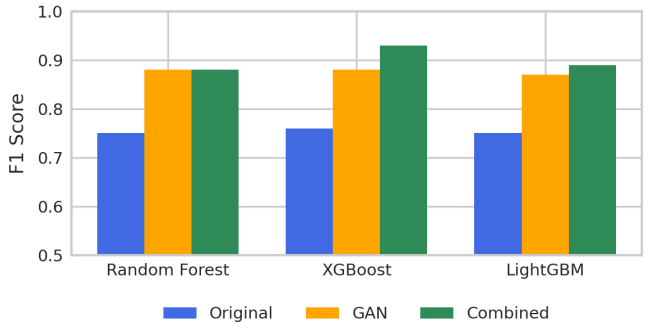

Fig. 4. F1 Score comparison across models demonstrating performance improvements.

The confusion matrices presented in Fig. 5 provide deeper insight into the classification performance across different student outcomes. The combined approach implemented with the XGBoost classifier demonstrated exceptional performance in correctly identifying all three student categories (Dropout, Enrolled, Graduate). Most notably, the combined model achieved remarkable improvement in dropout prediction accuracy, correctly identifying 985 dropout cases compared to just 396 in the original model. This enhanced ability to identify at-risk students is particularly valuable for intervention planning, as it enables institutions to target support resources more effectively. Similarly, the model's improved accuracy in classifying enrolled students (823 correct classifications) and graduates (1269 correct classifications) provides a more comprehensive understanding of student trajectories. The confusion matrices also reveal a reduction in misclassification errors. The combined model showed fewer instances of erroneously classifying dropout students as enrolled or graduated, which is crucial for minimizing false negatives in dropout prediction. This precision is essential for educational institutions seeking to identify all students who might benefit from intervention support.

TABLE IV. PERFORMANCE COMPARISON BETWEEN ORIGINAL AND CGAN-ENHANCED MODELS

Model	Accuracy			Precision			Recall			Training Time (s)		
	O	CG	C	O	CG	C	O	CG	C	O	CG	C
Random Forest	0.7748	0.8788	0.8800	0.7592	0.8828	0.8827	0.7748	0.8788	0.8800	0.5357	0.6903	1.2941
XGBoost	0.7733	0.8819	**0.9279**	0.7653	0.8853	**0.9284**	0.7733	0.8819	**0.9279**	0.2806	0.3223	**0.4661**
LightGBM	0.7666	0.8622	0.8917	0.7593	0.8672	0.8938	0.7666	0.8622	0.8917	0.3306	0.3697	0.5081

Legend: O = Original, CG = cGAN, C = Combined

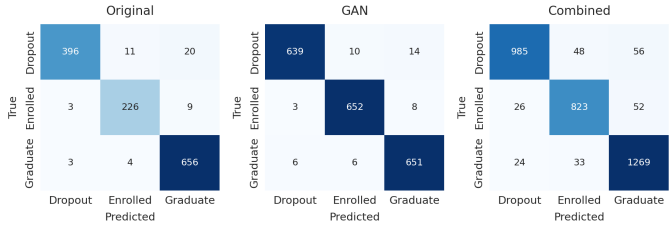

Fig. 5. Confusion matrix for the best model showing balanced classification performance.

C. Comparative Analysis of Intervention Methods

1) Intervention quantity and detail: The quantity and detail of interventions as shown in Fig. 6, varied substantially across the four methods evaluated. The Rule-Based approach generated the fewest interventions per student (2.55) with the shortest average length (8.70 words per intervention). This brevity limited the detail and actionability of the recommendations, resulting in generic guidance that may not adequately address student-specific challenges. In contrast, the Few-Shot method produced a substantially higher number of interventions (13.18) with moderate detail (38.89 words per intervention). The Fine-Tuned approach prioritized depth over breadth, generating a moderate number of interventions (8.64) but with the highest level of detail (61.61 words per intervention). This reflected the Fine-Tuned model's ability to elaborate on specific interventions with contextual information and implementation guidance. The Hybrid method achieved the highest intervention count (15.09) while maintaining a strong level of detail (41.49 words per intervention). This balanced approach suggests that the Hybrid method successfully combines the generative capacity of language models with the structured framework of rule-based systems to produce numerous detailed recommendations.

2) Comprehensiveness and Specificity: The category coverage metric as shown in Fig. 6, revealed important differences in how comprehensively each method addressed the various intervention domains (academic, financial, and social). The Hybrid method demonstrated exceptional coverage (0.94), addressing nearly all relevant intervention categories for each student. The Few-Shot approach achieved moderate coverage (0.79), while the Rule-Based (0.64) and Fine-Tuned (0.52) methods showed more limited scope.

The Fine-Tuned method's lower category coverage was particularly notable, suggesting that while it excelled in generating detailed interventions, it sometimes focused too narrowly on certain intervention types while neglecting others that might be relevant to a student's situation. Regarding specificity, the Hybrid approach again outperformed other methods with a score of 0.63, indicating a high degree of personalization in its recommendations. The Few-Shot (0.49) and Fine-Tuned (0.48) methods demonstrated moderate specificity, while the Rule-Based approach (0.41) generated more generic interventions with limited tailoring to individual student circumstances.

3) Computational efficiency: The computational efficiency analysis, as shown in Fig. 6, illustrates a pronounced disparity in runtime efficiency across the evaluated methods. The rule-based engine produced recommendations almost instantaneously ($\approx$ 0s per student), a speed that makes it well suited to large-scale or real-time deployments. In sharp contrast, the hybrid model required an average of 61.34 s per student, imposing a sizeable computational overhead that could hamper implementation in settings with limited hardware resources or stringent latency demands. This divergence underscores a fundamental trade-off: the hybrid approach delivers more nuanced and context-rich guidance, but its resource intensity may oblige institutions to generate recommendations in scheduled batches rather than on demand.

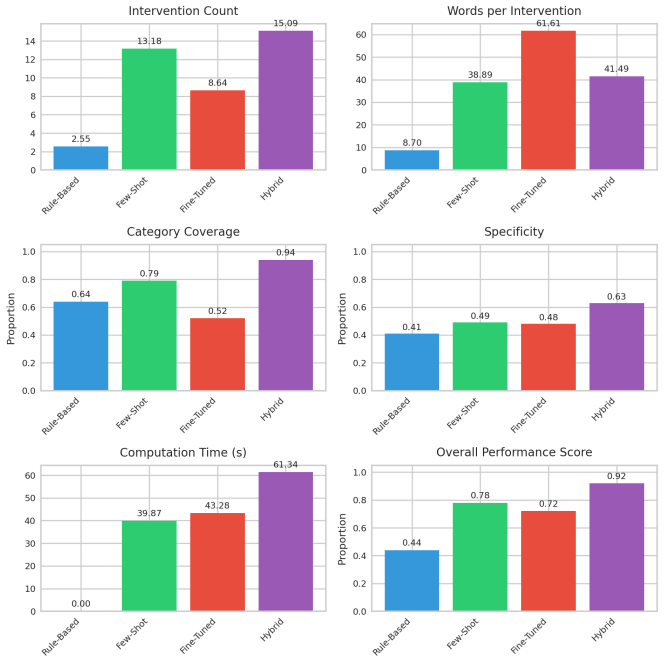

Fig. 6. Quantitative comparison of four student dropout intervention methods across six key performance metrics.

D. Qualitative Analysis of Intervention Content

Beyond quantitative metrics, we conducted qualitative analyses of the intervention content generated by each method, examining the language patterns and recommendation themes through word cloud visualizations.

1) Rule-based method content: The Rule-Based method produced highly structured but limited interventions. The word cloud as shown in Fig. 7, analysis revealed a narrow vocabulary focused primarily on basic academic support terms. Interventions typically followed rigid templates with minimal personalization, such as "attend tutoring sessions" or "meet with academic advisor." While consistently addressing core academic concerns, these interventions lacked depth and contextual relevance to specific student situations.

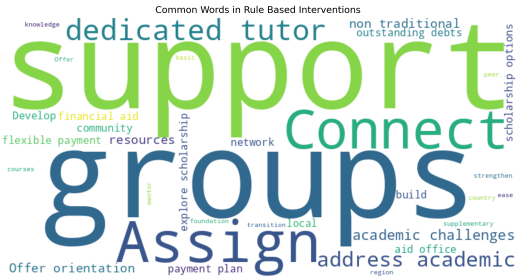

Fig. 7. Word cloud visualization of common terms in rule based method intervention recommendations.

2) Few-shot method content: The Few-Shot method generated interventions with more diverse language and recommendation types. The word cloud as shown in Fig. 8, showed a broader vocabulary spanning academic, financial, and social support domains. Interventions demonstrated increased contextual awareness but sometimes lacked the specificity necessary for highly personalized support. The method excelled in generating a wide range of intervention types but with moderate depth in each.

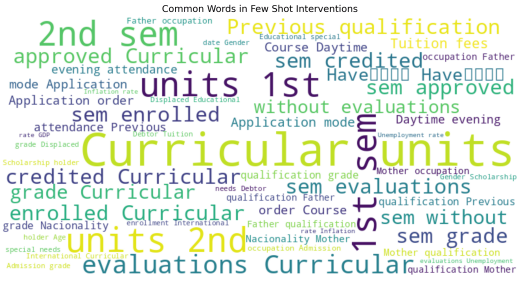

Fig. 8. Word cloud visualization of common terms in few shot method intervention recommendations.

3) Fine-tuned method content: The Fine-Tuned approach produced the most detailed and linguistically sophisticated interventions. The word cloud as shown in Fig. 9, revealed rich, specialized vocabulary related to academic support strategies and student success. These interventions often included specific implementation steps and contextual information, such as detailed tutoring recommendations with scheduling suggestions and expected outcomes. However, the method sometimes overemphasized certain intervention categories while neglecting others, resulting in lower overall category coverage.

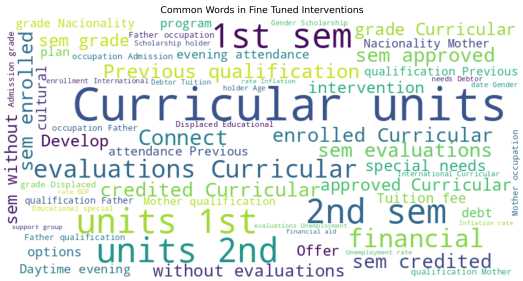

Fig. 9. Word cloud visualization of common terms in fine tuned method intervention recommendations.

4) Hybrid method content: The Hybrid method demonstrated the most balanced intervention content, combining comprehensive coverage with strong specificity. The word cloud as shown in Fig. 10, analysis revealed diverse vocabulary across all intervention domains, with terms related to academic support, financial assistance, and social integration appearing prominently. Interventions were both numerous and substantive, providing specific, actionable recommendations tailored to individual student profiles while addressing the full spectrum of potential support needs.

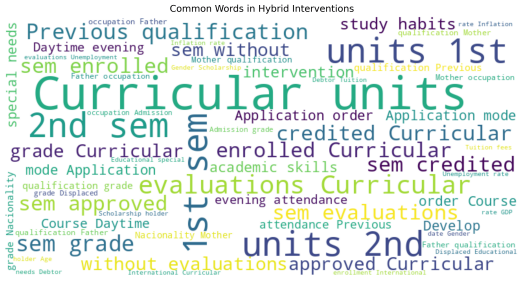

Fig. 10. Word cloud visualization of common terms in hybrid method intervention recommendations.

E. Strengths and Weaknesses Analysis

Our comprehensive evaluation revealed distinct patterns of strengths and limitations as shown in Table V for each intervention method:

1) The Rule-based: method's primary advantage lies in its computational efficiency, making it suitable for resource-constrained environments or large-scale deployment where processing time is a critical concern. However, its limited recommendation quality restricts its effectiveness in providing personalized, actionable support.

TABLE V. COMPARATIVE ANALYSIS OF STUDENT DROPOUT INTERVENTION METHODS: STRENGTHS AND WEAKNESSES

Method	Strengths	Weaknesses
Rule-Based	- Exceptional computational efficiency (0.00s) - Consistent structure across recommendations - Predictable output format	- Limited intervention count (2.55) - Minimal detail (8.70 words) - Low specificity (0.41) - Restricted vocabulary and variety
Few-Shot	- Strong intervention count (13.18) - Good category coverage (0.79) - Moderate specificity (0.49) - Diverse recommendation types	- Moderate detail (38.89 words) - Less precise personalization - Higher computational requirements than Rule-Based
Fine-Tuned	- Exceptional detail (61.61 words) - Rich, sophisticated language - Strong contextual information - Detailed implementation guidance	- Limited category coverage (0.52) - Moderate intervention count (8.64) - Tendency to focus on specific intervention types
Hybrid	- Superior intervention count (15.09) - Excellent category coverage (0.94) - Highest specificity (0.63) - Strong balance between quantity and quality	- Substantial processing time (61.34s) - Higher implementation complexity - Resource-intensive deployment

2) The Few-shot: This method represents a balanced approach that performs reasonably well across all metrics without excelling in any specific dimension. It offers a viable compromise between recommendation quality and computational requirements.

3) The Fine-tuned: This method's strength in generating detailed, context-rich interventions makes it particularly valuable when in-depth support in specific areas is prioritized over comprehensive coverage across all potential intervention categories.

4) The Hybrid method: This consistently outperformed other approaches across most evaluation metrics, demonstrating its effectiveness in generating numerous, comprehensive, and personalized interventions. However, its significant computational requirements represent an important practical limitation that must be considered for implementation.

V. DISCUSSION AND RESULTS

This study set out to examine whether an integrated framework that couples class-balanced dropout prediction with personalised intervention generation can meaningfully reduce student attrition. The empirical evidence supports this premise. By alleviating class imbalance with cGAN-generated samples, the predictive component—most notably the XGBoost classifier—achieved 92.79% accuracy and 92.84% precision, an improvement of 15.46 percentage points over its baseline performance, while maintaining a modest training time of 0.47 seconds. These results confirm that synthetic data augmentation can unlock substantial gains for established machine-learning algorithms.

Equally important, the analysis of intervention strategies showed that methodological integration offers advantages unattainable by any single paradigm. The hybrid engine, which blends rule-based scaffolding with the contextual agility of large language models, produced an average of 15.09 tailored recommendations per student, covered 94% of intervention categories, and attained the highest specificity score (0.63). Although this approach demands greater computational resources—approximately 61 seconds to generate interventions for each student—it consistently delivered the most comprehensive and contextually appropriate guidance. In contrast, the purely rule-based system ran almost instantaneously but lacked depth, while the fine-tuned model generated lengthy, sophisticated advice without achieving comparable breadth across academic, financial and social domains.

Several limitations temper these findings. The evaluation focused on recommendation quality rather than direct measurement of retention outcomes, and the computational requirements of the more sophisticated methods may restrict adoption in resource-constrained settings. Furthermore, certain qualitative dimensions of recommendation relevance are difficult to quantify and therefore fall outside the scope of the current metric suite. Future work should undertake longitudinal field studies to gauge real-world impact, explore optimisation techniques that reduce computational overhead, and incorporate reinforcement-learning feedback loops to refine interventions over time.

VI. CONCLUSION

This research developed an integrated framework combining class-balanced dropout prediction with personalized intervention generation to enhance student retention in higher education. The approach addresses a critical gap in educational technology by moving beyond risk identification to provide actionable remediation strategies.

The cGAN-enhanced XGBoost classifier achieved 92.79% accuracy, a 15.46 percentage point improvement over baseline performance, demonstrating that synthetic data augmentation can significantly improve predictive capabilities when addressing class imbalance in educational datasets.

The hybrid intervention engine, combining rule-based methods with large language models, outperformed individual approaches by generating 15.09 tailored recommendations per student with 94% category coverage and the highest specificity score (0.63). However, this comprehensive approach required 61 seconds per student compared to near-instantaneous rule-based processing, presenting important trade-offs for practical implementation.

The evaluation focused on recommendation quality rather than longitudinal retention outcomes, and computational requirements may limit adoption in resource-constrained environments. Future research should prioritize field studies measuring actual retention rates and explore optimization techniques to reduce processing overhead.

This integrated framework demonstrates that combining cGAN-enhanced prediction with hybrid intervention generation offers institutions a scalable pathway to reduce attrition while maintaining personalized support. The system bridges the critical gap between risk detection and remedial action, representing a significant advancement in educational technology for improving student persistence and academic success outcomes.

REFERENCES

[1] M. Kantrowitz, "Shocking statistics about college graduation rates," *Forbes*, Nov. 18, 2021. [Online]. Available: https://www.forbes.com/sites/markkantrowitz/2021/11/18/shocking-statistics-about-college-graduation-rates/

[2] S. Bayona-Oré, "Dropout in higher education and determinant factors," in *Proc. Int. Conf. Education and New Developments*, pp. 251–258, 2022. doi: 10.1007/978-981-19-2394-4_23

[3] M. Letseka and M. Breier, "Student poverty in higher education: the impact of higher education dropout on poverty," 2008.

[4] "Dropout among students in higher education: a case study," *Universidad, Ciencia y Tecnología*, vol. 27, no. 119, pp. 18–28, 2023. doi: 10.47460/uct.v27i119.703

[5] A. Ahmed et al., "Students' performance prediction employing decision tree," *CTU J. Innovation and Sustainable Development*, vol. 16, Special issue: ISDS, pp. 42–51, 2024. doi: 10.22144/ctujoisd.2024.321

[6] V. Kadu et al., "Student's performance prediction using machine learning algorithms – a comparative study," in *Proc. OTCON*, pp. 1–6, 2024. doi: 10.1109/otcon60325.2024.10687450

[7] J. Chen et al., "When large language models meet personalization: perspectives of challenges and opportunities," *arXiv preprint arXiv:2307.16376*, 2023. doi: 10.48550/arXiv.2307.16376

[8] A. Hawlitschek, V. Köppen, A. Dietrich, and S. Zug, "Drop-out in programming courses – prediction and prevention," *J. Applied Research in Higher Education*, vol. 12, no. 1, pp. 124–136, 2019. doi: 10.1108/JARHE-02-2019-0035

[9] D. Xu et al., "Large language models for generative information extraction: a survey," *arXiv preprint arXiv:2312.17617*, 2023. doi: 10.48550/arXiv.2312.17617

[10] S. I. Doyaoen, "Analysis of education student retention rates: basis for policy formation," *Int. J. Social Science and Education Research Studies*, vol. 4, no. 12, 2024. doi: 10.55677/ijssers/v04i12y2024-07

[11] S. Barragán and L. González Támara, "Complexities of student dropout in higher education: a multidimensional analysis," *Frontiers in Education*, vol. 9, 2024. doi: 10.3389/feduc.2024.1461650

[12] M. N. Ashraf, M. Riaz, and M. Y. Mujahid, "Investigating the causes of dropout in BS programs: insights from public sector universities," *J. Policy Research*, vol. 10, no. 2, pp. 559–567, 2024. doi: 10.61506/02.00269

[13] A. F. Núñez-Naranjo, "Analysis of the determinant factors in university dropout: a case study of Ecuador," *Frontiers in Education*, vol. 9, p. 1444534, 2024. doi: 10.3389/feduc.2024.1444534

[14] Z. Awang Long and M. F. Mohd Noor, "Factors influencing dropout students in higher education," *Education Research International*, vol. 2023, 2023. doi: 10.1155/2023/7704142

[15] F. García, "Risk factors for student dropout in higher education institutions in Latin America," *Salud, Ciencia y Tecnología*, vol. 4, 2024. doi: 10.56294/saludcyt2024.592

[16] O. Lorenzo-Quiles, S. Galdón-López, and A. Lendínez-Turón, "Dropout at university: variables involved on it," *Frontiers in Education*, vol. 8, 2023. doi: 10.3389/feduc.2023.1159864

[17] C. Sin, C. H. Sin, and M. J. Antunes, "Strategies to reduce dropout in higher education," 2024. doi: 10.4995/head24.2024.17082

[18] Y. Garcés-Delgado et al., "The process of adaptation to higher education studies and its relation to academic dropout," *European J. Education*, 2024. doi: 10.1111/ejed.12650

[19] S. Voghoei et al., "Students success modeling: most important factors," *arXiv preprint arXiv:2309.13052*, 2023. doi: 10.48550/arXiv.2309.13052

[20] D. A. Gutierrez-Pachas et al., "Supporting decision-making process on higher education dropout by analyzing academic, socioeconomic, and equity factors through machine learning and survival analysis methods in the Latin American context," *Education Sciences*, vol. 13, no. 2, p. 154, 2023. doi: 10.3390/educsci13020154

[21] A. G. Rincón et al., "Prevention and mitigation of rural higher education dropout in Colombia: a dynamic performance management approach," *F1000Research*, vol. 12, p. 497, 2023. doi: 10.12688/f1000research.132267.2

[22] E. D. Asenjo Muro et al., "Implementing effective sociocultural integration strategies to decrease university student dropout rates," *Sapienza*, vol. 5, no. 4, p. e24079, 2024. doi: 10.51798/sijis.v5i4.832

[23] L. M. Addison and D. Williams, "Predicting student retention in higher education institutions (HEIs)," *Higher Education, Skills and Work-Based Learning*, 2023. doi: 10.1108/heswbl-12-2022-0257

[24] S. Pongpaichet, S. Jankapor, S. Janchai, and T. Tongsanit, "Early detection at-risk students using machine learning," in *Proc. ICTC*, pp. 283–287, 2020. doi: 10.1109/ICTC49870.2020.9289185

[25] M. S. Mosia, "Identifying at-risk students for early intervention – a probabilistic machine learning approach," *Applied Sciences*, vol. 13, no. 6, p. 3869, 2023. doi: 10.3390/app13063869

[26] A. Prasanth and H. Alqahtani, "Predictive modeling of student behavior for early dropout detection in universities using machine learning techniques," in *Proc. ICETAS*, pp. 1–5, 2023. doi: 10.1109/icetas59148.2023.10346531

[27] M. Vaarma and H. Li, "Predicting student dropouts with machine learning: an empirical study in Finnish higher education," *Technology in Society*, vol. 76, p. 102474, 2024. doi: 10.1016/j.techsoc.2024.102474

[28] E. Gök, B. S. Gür, M. Ş. Bellibaş, and M. Öztürk, "Science mapping the knowledge base on student retention in higher education: a bibliometric review of research papers from 1914–2022," *Üniversite Araştırmaları Dergisi*, vol. 7, no. 4, pp. 348–362, 2024. doi: 10.32329/uad.1547067

[29] V. A. Brown and V. A. Lawack, "A re-imagined institutional student retention and success framework to enhance holistic student support," in *Proc. ICFTE*, vol. 1, no. 1, pp. 59–69, 2023. doi: 10.33422/icfte.v1i1.2

[30] V. Realinho, M. V. Martins, J. Machado, and L. Baptista, "Predict students' dropout and academic success," UCI Machine Learning Repository, 2021. doi: 10.24432/C5MC89

[31] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324

[32] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 785–794, 2016. doi: 10.1145/2939672.2939785

[33] A. Kriuchkova, V. Toloknova, and S. Drin, "Predictive model for a product without history using LightGBM. Pricing model for a new product," *Mogilánskij Matematičnij Žurnal*, 2024. doi: 10.18523/2617-7080620236-13

[34] G. Ke et al., "LightGBM: a highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems 30*, pp. 3149–3157, 2017.

[35] Qwen Team, "Qwen2.5-0.5B," Hugging Face Model Repository, 2024. [Online]. Available: https://huggingface.co/Qwen/Qwen2.5-0.5B