LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades, Asunción, Paraguay
ISSN en línea: 2789-3855, 2026

Predicting the use of learning objects: supporting high school education

Predicción del uso de objetos de aprendizaje: apoyo a la educación secundaria

Adriana Pérez López
adriana.pl@teziutlan.tecnm.mx
https://orcid.org/0000-0003-3712-400X
Instituto Tecnológico Superior de Teziutlán /TecNM
Teziutlán – México

Raúl Mora Reyes
raul.mr@teziutlan.tecnm.mx
https://orcid.org/0000-0002-1428-8901
Instituto Tecnológico Superior de Teziutlán/TecNM
Teziutlán – México

Naty Rodríguez Ventura
naty.rv@teziutlan.tecnm.mx
https://orcid.org/0009-0007-5976-4418
Instituto Tecnológico Superior de Teziutlán/TecNM
Teziutlán – México

Sara Aros Alberto
M23TE0050@teziutlan.tecnm.mx
https://orcid.org/0009-0004-4116-499X
Instituto Tecnológico Superior de Teziutlán/TecNM
Teziutlán – México

DOI: https://doi.org/10.56712/latam.v7i4.6409

Redilat
Red de Investigadores Latinoamericanos

LATAM
Revista Latinoamericana de Ciencias Sociales y Humanidades

Artículo recibido: 03 de abril de 2026.
Aceptado para publicación: 03 de septiembre de 2026.
Conflictos de Interés: Ninguno que declarar.

VOLUMEN VII

DOI: https://doi.org/10.56712/latam.v7i4.6409

Predicting the use of learning objects: supporting high school education

Predicción del uso de objetos de aprendizaje: apoyo a la educación secundaria

Adriana Pérez López
adriana.pl@teziutlan.tecnm.mx
https://orcid.org/0000-0003-3712-400X
Instituto Tecnológico Superior de Teziutlán /TecNM
Teziutlán – México

Raúl Mora Reyes
raul.mr@teziutlan.tecnm.mx
https://orcid.org/0000-0002-1428-8901
Instituto Tecnológico Superior de Teziutlán/TecNM
Teziutlán – México

Naty Rodríguez Ventura
naty.rv@teziutlan.tecnm.mx
https://orcid.org/0009-0007-5976-4418
Instituto Tecnológico Superior de Teziutlán/TecNM
Teziutlán – México

Sara Aros Alberto
M23TE0050@teziutlan.tecnm.mx
https://orcid.org/0009-0004-4116-499X
Instituto Tecnológico Superior de Teziutlán/TecNM
Teziutlán – México

Artículo recibido: 03 de abril 2026. Aceptado para publicación: 03 de septiembre de 2026.
Conflictos de Interés: Ninguno que declarar.

Abstract

Education today faces major challenges compared to the situation before 2020, due to changes in the ways of teaching and learning caused by the difficulties of attending classes during the COVID-19 pandemic. Among the main challenges are the lack of didactic resources, limited technological infrastructure, and the geographical location of many schools, particularly those in rural areas. This research analyzes the effectiveness of a predictive model based on the Random Forest algorithm to suggest personalized learning objects for high school students, through the analysis of academic and behavioral variables such as frequency and time of use of digital resources, motivation, and the perceived impact on comprehension. The findings confirm that the predictive methodology not only supports educational personalization but also generated recommendations tailored to each student's profile, optimizing the use of digital resources and significantly improving comprehension and academic performance. Validation, carried out through a diagnostic pre-test and a final post-test, demonstrated a significant average improvement of 45.24% in the academic performance of the experimental group that utilized the personalized learning objects suggested by the model. This study proposes the integration of machine learning models as an adaptive strategy in educational environments, highlighting the capacity of data analytics to promote equity, reduce learning disparities, and strengthen student-centered pedagogy.

Keywords: learning objects, predictive model, academic performance, student motivation

Resumen

La educación se enfrenta hoy en día a importantes retos en comparación con la situación anterior a 2020, debido a los cambios en las formas de enseñar y aprender provocados por las dificultades para asistir a clase durante la pandemia de COVID-19. Entre los principales retos se encuentran la falta de recursos didácticos, la infraestructura tecnológica limitada y la ubicación geográfica de muchos centros educativos, especialmente los situados en zonas rurales. Esta investigación analiza la eficacia de un modelo predictivo basado en el algoritmo Random Forest para sugerir objetos de aprendizaje personalizados a alumnos de secundaria, mediante el análisis de variables académicas y de comportamiento, como la frecuencia y el tiempo de uso de los recursos digitales, la motivación y el impacto percibido en la comprensión. Los resultados confirman que la metodología predictiva no solo favorece la personalización educativa, sino que también generó recomendaciones adaptadas al perfil de cada alumno, optimizando el uso de los recursos digitales y mejorando significativamente la comprensión y el rendimiento académico. La validación, llevada a cabo mediante una prueba de diagnóstico y un examen final, mostró una mejora media del 45,24 % en el rendimiento académico del grupo experimental que utilizó los objetos de aprendizaje sugeridos por el modelo. Este estudio propone la integración de modelos de aprendizaje automático como estrategia adaptativa en entornos educativos, destacando la capacidad del análisis de datos para promover la equidad, reducir las disparidades en el aprendizaje y reforzar la pedagogía centrada en el alumno.

Palabras clave: objetos de aprendizaje, modelo predictivo, rendimiento académico, motivación de los alumnos

Cómo citar: Pérez López, A., Mora Reyes, R., Rodríguez Ventura, N., & Aros Alberto, S. (2026). Predicting the use of learning objects: supporting high school education. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades 7 (4), 2433 – 2452. https://doi.org/10.56712/latam.v7i4.6409*

INTRODUCTION

In 2015, the United Nations General Assembly approved the 2030 Agenda, which is composed of 17 Sustainable Development Goals. In Mexico, the three goals with the least progress are: Goal 9 – Industry, Innovation, and Infrastructure; Goal 15 – Life on Land; and Goal 4 – Quality Education (Global Compact, 2025). The last one is where the use of Learning Objects (LO) is being promoted. According to data from the National Institute of Statistics and Geography (INEGI) and the Secretariat of Public Education (SEP), during the 2022–2023 period, the graduation efficiency rate in upper secondary education in the states of Puebla and Veracruz was 77.2% and 75.6%, respectively (INEGI, 2025).

The predominant factor in these figures is school dropout, which in both states during the same period was 8.7%. The main reasons include: financial difficulties that force students to work, low academic performance and failure, lack of interest in studies, and gaps in prior educational quality (Hernández Osorio, 2023).

Against this backdrop, and based on data from the 2023 National Survey on Availability and Use of Information Technologies in Households (National Institute of Statistics and Geography, 2023), internet usage among youth aged 12 to 17 in rural areas of Mexico reached 92.4%, mostly through mobile devices. This brings additional benefits such as: easier access to digital resources, promotion of knowledge exchange, expansion of educational horizons, and implementation of innovative pedagogical strategies. Therefore, this research focuses on upper secondary schools, including both preparatory and high school institutions (Serna Martínez & Alvites Huamaní, 2021).

This project explores the conceptual foundations that support the use of learning objects, with a focus on their application in rural areas of Teziutlan, a region in the state of Puebla. The study builds upon previous research that evaluated the use of learning objects in educational contexts characterized by limited resources and poor learning conditions. These factors are considered indicators for assessing the practical effectiveness of learning objects, their impact on academic performance, and their value as pedagogical tools for educators within the teaching and learning process. (Orellana-Cordero et al., 2020) The aim of this study is to establish criteria for determining whether learning objects effectively contribute to achieving educational objectives in such environments. (Aros, 2025).

METHODOLOGY

Research approach

The study is based on a quantitative, predictive, and cross-sectional approach, supported by the analysis of historical data on the use of Learning Objects (LO) among higher education students (Palmera Quintero et al., 2023). This methodological design was chosen for its ability to identify behavioral patterns and generate predictive models that help optimize pedagogical strategies.

To carry out this study, the CRISP-DM methodology was implemented—commonly used in data mining projects due to its adaptability to various educational contexts. It consists of six phases (Hotz, 2024), described as follows:

Business Understanding:

This phase focuses on ensuring alignment between the educational objectives of the project and the proposed data analysis, guaranteeing that decisions address real needs within the academic context. Educational environment needs were identified, and key questions were formulated to guide the analysis. The main goal of this phase is to predict which LOs show the highest levels of use and acceptance among high school students in the Teziutlán region, in order to personalize teaching and improve academic performance.

To achieve this, activities such as the following: structured interviews and meetings with teachers, coordinators, and principals to understand their needs and identify pedagogical and technological gaps. Key questions were formulated, such as: What impact would identifying the most-used LOs have? How would the results influence educational planning?, Which LO are considered most necessary to reinforce learning?, Analytical questions were also designed, including: What usage patterns do LO exhibit?; Are certain LO more commonly used in specific subjects?, What factors influence the selection of LO? Study metrics were defined, including: frequency of use for each LO, average interaction time per LO, and preferences by subject, group, or academic level.

This phase also required contextual analysis, evaluating the technological infrastructure of schools and available data sources (surveys, records, logs), while recognizing limitations such as lack of historical data or inconsistencies. A timeline was established based on CRISP-DM phases, with role assignments for data collection, model development, and validation. Necessary tools were identified: Python, MySQL, Jupyter Notebook, Power BI.

Data Understanding

After data collection, dependent and independent variables were identified to address the research problem. The dataset included information such as age, gender, and predominant learning styles. Additional questions were posed to gather insights: How motivated do you feel using this type of learning?, How many hours do you spend studying or reviewing school content?, How often do you access platforms containing Learning Objects (e.g., Kahoot, Duolingo, videos)?, What is the average time you dedicate to each session?, What type of LO is most used?, Which LO do you prefer to use?, How often do you complete the LO?, Do you consider LO to be effective?, Are LO interactive?, Do LO help improve your academic performance?, Table 1 presents the description of the variables used.

Table 1

Description of study variables

NOMBRE_VARIABLE	TIPO_DATO	DESCRIPTION
Edad	Numeric (Integer)	Age of students
Género	Nominal categorical	Gender of students (female, male, other)
Grado	Ordinal categorical	Grade level of students (1st, 2nd, 3rd)
acceso_internet	Binary categorical	Indicate whether the student has access to the internet.
utiliza_OA	Nominal categorical	If you have ever used LO
frecuencia	Ordinal categorical	Frequency of LO use (sometimes, never, frequently, etc.)
tiempo_promedio	Ordinal categorical	Average time spent by students on LO
OA_frecuencia	Ordinal categorical	Frequency with which students use LO
OA_mayorfrecuencia	Nominal categorical	Type of LO used by the student (videos, crossword puzzles, etc.)
OA_mejora	Binary categorical	If you have seen any improvement when using LO
OA_comprensión	Binary categorical	Whether the student understood the content of the learning objectives.
OA_rendimiento	Binary categorical	If LO helped you improve your academic performance
OA_barrera	Binary categorical	The bars presented by the student, in order to use the LO
OA_motivacion	Binary categorical	If the student feels motivated when using LO

Source: own elaboration.

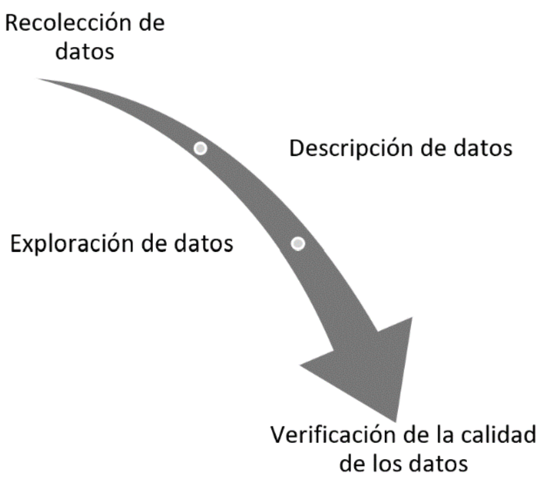
Data Preparation

Once the procedure was completed and based on the data provided in the questions, the analysis was segmented as shown in Figure 1. It begins with data collection, a phase in which records related to OA use are obtained through surveys. Next, the data is described, identifying its characteristics and structure, as well as the most important variables. This is followed by the data exploration phase, where patterns, trends, and relationships between the most important variables are analyzed. Finally, data quality verification is carried out to ensure that the information used is reliable and suitable for the predictive model, which details the data collection activities.

Data preparation process that is part of the CRISP-DM methodology to ensure that the information used is reliable and suitable for the predictive model.

Figure 1

Tasks for data activities



Source: own elaboration.

Data Collection

Data was obtained through surveys, previous studies, and academic grades by group or subject. Key variables were identified, such as age, gender, grade level, internet access, usage frequency, motivation, and comprehension. Null or irrelevant variables were removed, and the target variable was defined: the most frequently used type of learning object.

Data Description

Key variables necessary to build the predictive model were identified and selected, with the aim of improving the accuracy and effectiveness of the analysis.

The variables were obtained through a survey, allowing for the collection of representative data from the analyzed sample. Figure 2 presents the variables considered in the model, along with a detailed description of each one. Extract from the database used in the predictive analysis showing the main variables of the predictive model applied in the research, which allowed us to identify behavior patterns and establish the relationship between OA use and academic performance.

Figure 2

Extracted variables

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Edad	Sexo	Estilo_Aprendi	Motivacion	Horas_Estudic	Frecuencia_A	Tiempo_Sesio	OA_Utilizado	OA_Preferido	Frecuencia_Tı	Valoracion_Ef	Tipo_OA	Mejora_Rendimiento
2	17	1	2	4	3	2	1	6	3	1	2	4	Bueno
3	15	1	3	3	3	2	1	5	4	1	3	3	Bueno
4	16	2	2	1	3	2	1	1	2	2	1	1	Bueno
5	15	2	3	1	3	2	1	1	3	2	1	1	Bueno
6	99	1	4	1	3	2	1	4	3	2	1	1	Bueno
7	17	2	3	5	3	2	1	4	4	2	5	5	Bueno
8	16	1	4	3	1	2	1	5	4	2	3	3	Bueno
9	16	1	1	1	2	1	1	5	1	3	5	5	Bueno
10	16	2	1	5	3	2	1	4	3	3	1	2	Bueno
11	16	2	3	3	1	2	2	4	1	1	3	3	Regular
12	17	1	2	3	1	2	1	4	1	1	2	2	Regular
13	17	2	3	3	3	2	1	2	2	1	3	3	Regular

Source: own elaboration.

Data Exploration

The data was filtered to include only high school students. Columns were renamed to facilitate analysis in Python. Duplicates were removed and formats were standardized. Variables such as age were transformed to numerical format and type_oa_frecuente to binary format using MultiLabelBinarizer. Similar subjects were grouped, metrics such as average usage time were calculated, and multiple responses were encoded in binary columns. The dataset was prepared for predictive analysis.

Table 2

Renaming of data table

Name	Renamed
timestamp	fecha_respuesta
Age	edad
Gender	genero
Grade level	grado
Do you have internet access at home?	acceso_internet
Have you used online learning objects?	uso_objetos_aprendizaje
If yes, how often do you use them?	frecuencia_uso
How long, on average, do you use these resources per session?	tiempo_promedio
What type of learning objects do you use most often?	tipo_oa_frecuente
Where do you access learning objects most often?	lugar_acceso
Do you feel that using learning objects has improved your understanding of the topics?	impacto_comprension
How easy do you find it to understand the content of learning objects?	facilidad_comprension
How do you think these resources influence your academic performance?	impacto_calificacion
What is the main barrier to using learning objects in your case?	barrera_uso
Do you feel more motivated to study when you use learning objects?	motivacion_uso

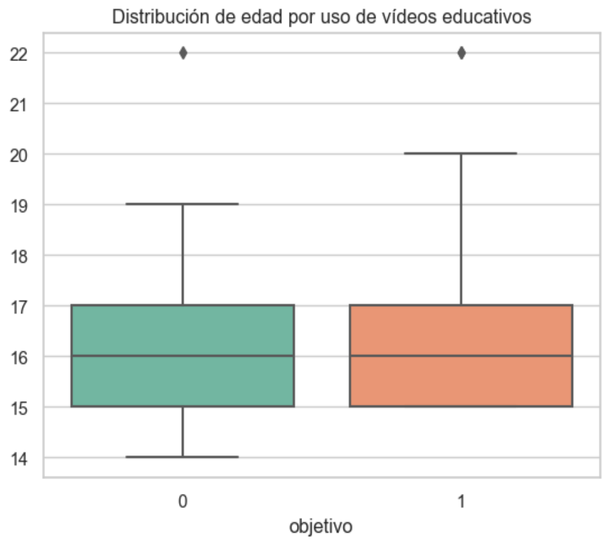
Source: own elaboration.

Data quality verification

Statistical and visual techniques were applied to understand the structure of the data. Boxplots were generated to compare age vs. objective, where it was observed whether there was any difference in the age distribution between those who use and those who do not use educational videos. This information is shown in graphic 1. Box plot identifying that the use of educational videos is not concentrated in a specific age range.

Graphic 1

Boxplot Age vs objective

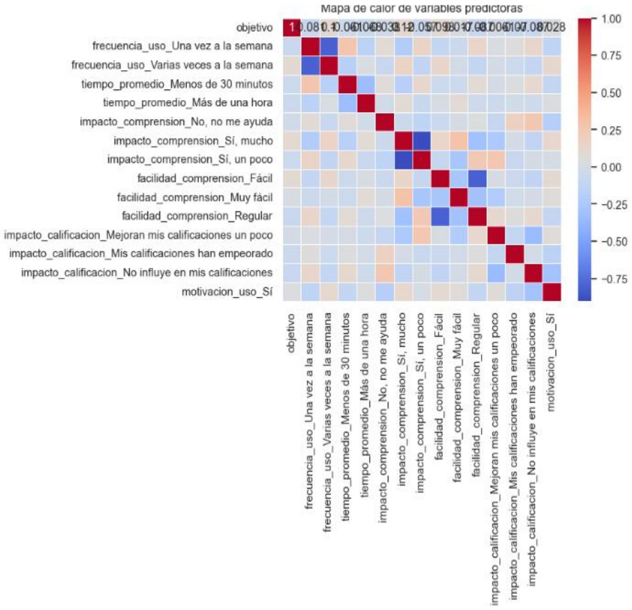


Source: own elaboration.

Through the creation of the heat map, graphic 2, it was possible to observe the correlation between the predictor variables and the target variable, thereby detecting the pattern of relationship between variables.

Graphic 2

Predictive variables and objective



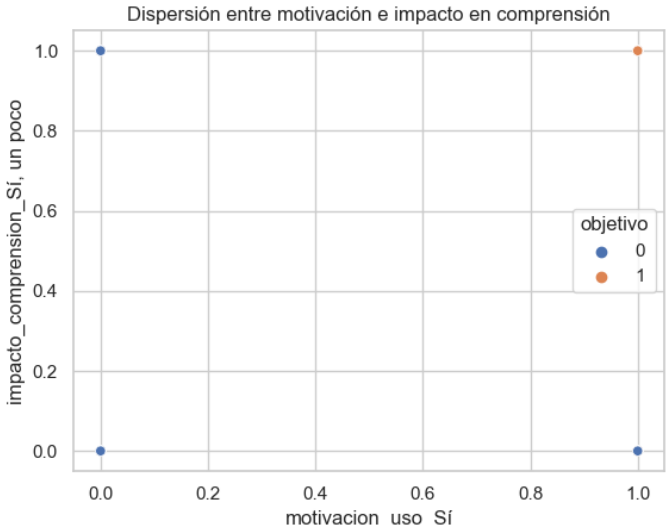
Source: own elaboration.

Heat map of correlations, where shades closer to red show stronger positive correlations, while blues represent negative correlations.

Motivation and impact were evaluated using a scatter plot, which allowed the relationship between both factors to be visualized clearly and accurately, facilitating the identification of patterns, trends, and possible correlations between the participants' level of motivation and the impact generated in the context of the study. These data are shown in graphic 3. The scatter plot showed that students with higher motivation tend to indicate a positive impact on their understanding.

Graphic 3

Motivation and Impacts of LO

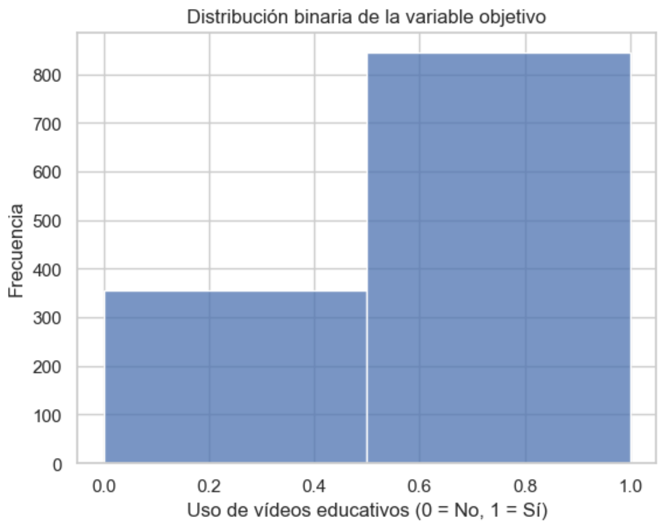


Source: own elaboration.

Lastly, within this phase, a histogram was used to analyze frequency and usage in order to identify who does and does not use videos as a study tool. These results are shown in graphic 4. Binary distribution of the object variable, which is related to the use of educational videos as a learning object. It can be seen that most students (value 1) do use educational videos, while a smaller group (value 0) say they do not use them.

Graphic 4

Use of educational videos



Source: own elaboration.

D. Modeling

In this phase, predictive analysis models were developed to identify patterns in the data and anticipate which learning objects are most used by students. The purpose is to generate useful information for educational decision-making.

To achieve an effective model, key questions were raised regarding algorithm selection, variable transformation, data division, performance evaluation, and model scalability.

Data preparation and division

Historical data collected between December 2024 and April 2025 was used; the target variable was: most used type of learning object, supervised machine learning with multi-class classification was applied, and categorical variables were transformed into numerical variables using one-hot encoding. Missing values were imputed using the most frequent value replacement technique.

Division of the dataset

70% for model training; 10% for validation and hyperparameter tuning; and 20% for final model testing with unseen data.

E. Evaluation

During this phase, the performance of the Random Forest model was evaluated. This model was designed to recommend learning objects to high school students with low performance in mathematics and English.

Cross-validation and test set tools were applied to ensure generalization of results. The metrics applied were: accuracy: overall percentage of correct predictions; precision and recall: quality of positive predictions; F1-score: balance between precision and recall; and confusion matrix: visual identification of hits and misses by class.

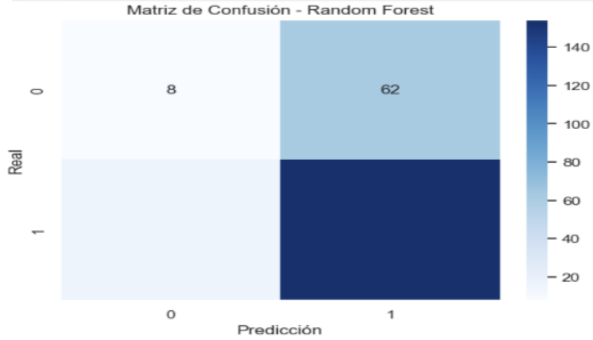
The results of this phase demonstrated the model's high performance in predicting and categorizing student profiles; Random Forest showed resistance to overfitting and good handling of large volumes of variables. With the application of feature importance analysis, it was possible to identify the student attributes most influential in the predictions, graphic 5 shows the results of two of the algorithms mentioned. Comparative results of three machine learning algorithms: Logistic Regression, Random Forest, and KNN, including evaluation metrics and confusion matrices.

Graphic 5

Evaluation of results

	Precisión	Recall	F1-score	Exactitud
Regresión Logística	0.729258	0.982353	0.837093	0.729167
Random Forest	0.712963	0.905882	0.797927	0.675000
KNN	0.709360	0.847059	0.772118	0.645833

```
for nombre, modelo in modelos.items():
    y_pred = modelo.predict(X_test)
    cm = confusion_matrix(y_test, y_pred)
    sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
    plt.title(f'Matriz de Confusión - {nombre}')
    plt.xlabel('Predicción')
    plt.ylabel('Real')
    plt.show()
```



Source: own elaboration.

The model met the objectives set out in the planning phase and is therefore considered viable for practical application, although the possibility of making iterative adjustments remains open if faults are detected in specific subgroups (e.g., improving data or adjusting hyperparameters).

F. Implementation

In this final stage, the models developed are implemented in a real environment with the aim of translating their predictions and findings into concrete actions. The purpose is to ensure that the results of the analysis do not remain theoretical, but are integrated into educational decision-making.

Among the practical applications of the model, teachers receive reports that help them identify the most effective resources; educational platforms are updated with the most relevant learning objects; and automated systems can monitor the use of these resources in real time.

To visualize the results, an interactive tool was developed using Power BI, which allowed for the creation of graphs, maps, and dynamic tables; intuitive filtering and segmentation of data; connection to multiple data sources (SQL, Excel, APIs); configuration of automatic updates; timely detection of trends or problems; and provision of an accessible and cost-effective solution, especially useful in educational environments.

The impact of deploying the predictive model based on historical data made it possible to identify the most frequently used learning objects, revealing patterns of preference among students. This translates into concrete benefits for teachers, administrators, and students, facilitating class planning with more effective resources and supporting data-driven strategic decisions by ensuring access to more useful and personalized materials, respectively.

To sum up, the deployment not only displays statistics, but also transforms data into practical tools that improve the educational experience. Interactive visualizations allow for dynamic exploration of information, promoting informed decisions and continuous improvements in the teaching-learning process.

Population and sample

The study population focused on high schools in the Teziutlán region, with 1,250 students from eight institutions, all selected through non-probabilistic convenience sampling, which included 22 students from the Henry Wallon A.C. Institute, (HW) divided into two groups, one control and one experimental, who were given general OAs and OAs recommended by the model, respectively.

To validate the statistical significance of the improvement observed in the experimental group (Pre-test mean: 10.73, SD: 2.86; Post-test mean: 15.27, SD: 2.10), a paired-samples t-test was conducted(Triola F. Mario, 2004). The results revealed a statistically significant increase in academic performance ($t(10) = 6.42, p < 0.001$), confirming that the personalized recommendations derived from the Random Forest model generated a real pedagogical effect rather than a random variation (Rodríguez, 2023).

DEVELOPMENT

There are cases documented in educational literature where technology has been key to significant improvements in the teaching and learning process. Such is the case of research conducted in Pakistan, which sought to address the lack of studies focused on that country's specific context with regard to global educational trends. While global trends are valuable, it is essential to understand how these innovations and changes manifest themselves in the educational environment of Pakistan, specifically in Punjab. The main objective of the study is to thoroughly analyze the applicability, difficulties, and possible modifications of global educational trends and innovations in the sociocultural and educational environment of Pakistan. The study focuses on the impact of technological advances, pedagogical changes, inclusive education practices, and the globalization of education to provide information that can guide educational policies and practices in Pakistan. A quantitative research design was used. With a cohort size of 160 teachers from Punjab, simple random sampling was used to ensure that each teacher had an equal chance of being selected. Data collection was carried out through a self-developed questionnaire that was pre-tested to ensure clarity and reliability, using face-to-face interviews. The technologies applied include Artificial Intelligence (AI), with adaptive learning platforms to personalize instructional content according to students' needs, and Virtual Reality (VR), providing immersive learning experiences, such as simulations, that exceed the capabilities of traditional classrooms. As a result, the vast majority of participants showed a favorable attitude toward the integration of technology in classrooms (81% agree), the effectiveness of student-centered learning approaches (73% agree), and the relevance of global educational trends in the local context (64% agree). Similarly, there were divergent opinions on the challenges related to technological infrastructure and the importance of cultural competence as a key factor in promoting inclusion. (Hadayat et al., 2024)

Another case is mentioned by (De Medio et al., 2019) in their publication on the use of MoodleRec for course creation via Moodle platforms, the web offers a wide variety of digital resources for learning environments, available to both teachers and students.

In the book Learning Objects as Support in the Teaching-Learning Process (Tovar et al,.2019), Chapter 6 discusses a case at the Colegio de Postgraduates. The issue identified was that despite progress in integrating Information and Communication Technologies (ICTs) into higher and postgraduate education, there remains a need to design LOs that are efficient, reliable, adaptable, and open. The need which is presented in the school before mentioned was to strengthen digital libraries with educational resources aligned to international standards and solid pedagogical frameworks. The ADDIE model (Analysis, Design, Development, Implementation, and Evaluation) was used in the project, along with technological tools such as PHP language, web servers, and metadata managers. The result was the design and development of a specific LO on the entity-relationship model, aimed at students learning database management. It was evaluated using formal instruments and a complete metadata proposal based on IEEE-LOM. This work demonstrated that the developed resources can be used freely by higher education teachers and students under open licenses.

Similarly, (Pulido et al. 2016), in their work on LO as educational resources in algorithm instruction, studied 47 students from Group 301 of the Bachelor's in Administrative Computer Systems (LSCA) at the Faculty of Accounting and Administration, Universidad Veracruzana, Coatzacoalcos campus, during the August 2014 – February 2015 period. The goal was to address challenges in the LSCA program’s Algorithmics course, where students struggled with algorithmic structures such as if-then, for, and while. This led to the design and implementation of LOs as didactic tools to improve understanding of these complex topics. The project used a quantitative, non-experimental, cross-sectional methodology, with an 11-item questionnaire (2 demographic and 9 specific items). Variables evaluated included content, aesthetics, and functionality. The questionnaire was validated through a pilot test (10 students) and Cronbach’s alpha for reliability (>0.7). Results of the three variables were satisfactory: Content: 86% said the LO information was complete; 92% found the content aligned with the curriculum; 78% said it facilitated learning. Aesthetics: 96% found the colors appropriate; 90% said the typography was legible; 70% found the images suitable, though 23% noted some were too large, identifying figure size as an area for improvement. Functionality: 76% found the LO interaction clear, but 82% perceived slow response speed, indicating a need for improvement.

In conclusion, the applied LOs proved effective in terms of content and learning facilitation. However, aspects such as performance speed and visual presentation need enhancement. Overall student perception was positive, suggesting high potential for this type of resource as a didactic tool in higher education.

This study focuses on the use of data science techniques to improve high school Teziutlán, Puebla City. Specifically, it proposes a predictive model based on the Random Forest algorithm to recommend personalized LOs to high school students, aiming to improve academic performance, especially in critical subjects such as mathematics and English. This approach is grounded in the Cross Industry Standard Process for Data Mining (CRISP-DM), ensuring a systematic process for data collection, analysis, and modeling. The study emerges in response to the educational lag exacerbated by the COVID-19 pandemic, which highlighted the need for effective and adaptive digital resources in rural contexts.

RESULTS

The analysis was carried out by carefully cleaning and transforming the data collected in a survey administered to high school students at HW in Teziutlán, Puebla. Categorical variables were transformed into numerical variables using one-hot encoding (pd.get_dummies()), which allowed them to be used in machine learning models. In addition, missing values were completed using the mode strategy with SimpleImputer, ensuring the consistency of the dataset. This is shown in Figure 8 in the code written in Python; code snippet written in Python that is used to prepare data in the predictive model. This process ensures that the data is ready to train the algorithms.

Figure 3

Encoding techniques

```
[7]: df_model = pd.get_dummies(df[variables_exploratorias + ['objetivo']], drop_first=True)
X = df_model.drop(columns='objetivo').astype(int)
y = df_model['objetivo']
imputer = SimpleImputer(strategy='most_frequent')
X_imputed = imputer.fit_transform(X)
X_train, X_test, y_train, y_test = train_test_split(X_imputed, y, test_size=0.2, random_state=42)
```

Source: own elaboration.

The target variable of the predictive classification problem was operationalized as the preference and frequent use of specific Learning Objects (specifically Educational Videos, binarized as 1 = Yes, 0 = No via MultiLabelBinarizer encoding), based on behavioral and motivational predictors. Table 3.

Predictor variables

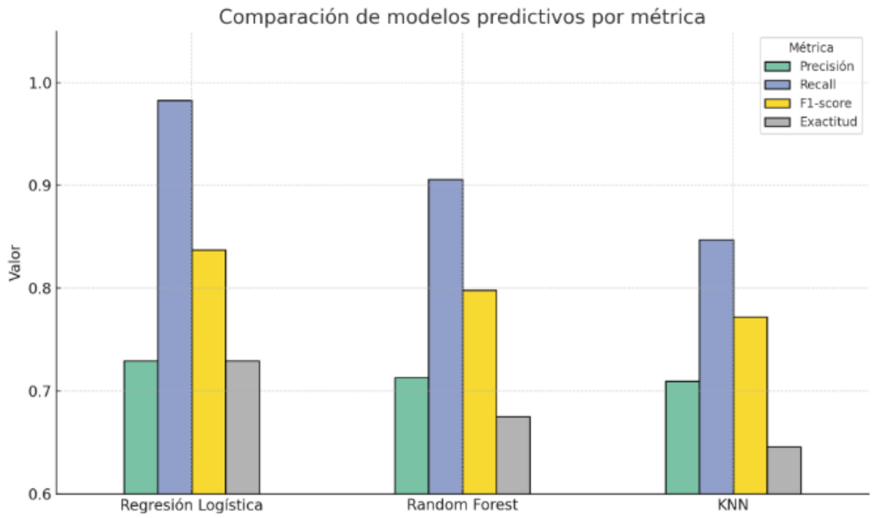
- Frequency of use of learning objects.
- Average time per sesión.
- Impact on comprehension.
- Ease of comprehension.
- Impact on rating.
- Reason for use.

When training and comparing the three supervised classification algorithms, it was observed that the results were very similar in performance, which shows the consistency of the patterns found in the data and the robustness of the target variable.

Details a comparison between each of the models, which provides a broader overview for making the best decision about which algorithm to train

Graphic 6

Comparison of predictive models by metric



Source: own elaboration.

Although Logistic Regression achieved slightly higher global metrics (Accuracy: 0.729, F1: 0.837), Random Forest was selected as the core algorithm for the recommendation engine (Accuracy: 0.675, Recall: 0.906, F1: 0.798). This choice is justified by Random Forest's superior ability to model non-linear interactions among heterogeneous educational variables, its intrinsic resistance to overfitting, and its feature importance capability, which provides actionable pedagogical insights to teachers regarding which factors (e.g., student motivation, session duration) most strongly dictate resource adoption; Table 3.

Table 3

Results of Predictive Models

Algorithm	Precision	Recall	F1-Score	Accuracy
Logistic Regression	0.729	0.982	0.837	0.729
Random Forest (Selected)	0.713	0.906	0.798	0.675
K-Nearest Neighbors (KNN)	0.709	0.847	0.772	0.646

CONCLUSION AND SUGGESTIONS

Figure 9 shows the dashboard, which provided a structured and clear visualization of the impact of the Random Forest model on the academic performance of the experimental group.

A summary card was added showing an average increase of 45.24%, evidencing an overall improvement in academic performance after applying personalized learning objects.

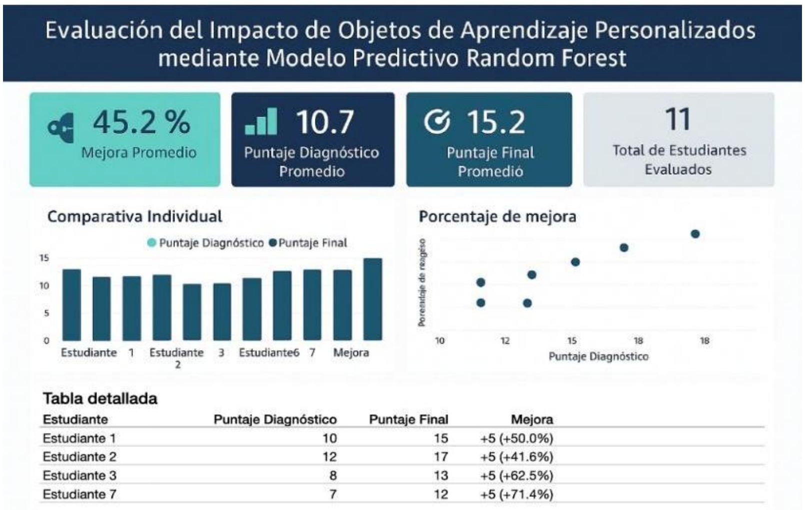
Grouped column charts were used to compare the results of the diagnostic test and the final test per student, allowing individual progress to be observed and confirming that most students showed significant improvements.

In addition, the scatter plot helped analyze the relationship between the initial score and the percentage of improvement, demonstrating that even those who started with low scores made remarkable progress.

Visual dashboard of the conclusions reached in the assessment of the impact of personalized learning objects, using the Random Forest predictive model, including individual student comparisons, the percentage improvement achieved, and a table with specific results.

Graphic 7

Learning Object Impact Assessment



Source: own elaboration.

The application of diagnostic and summative tests, designed to assess basic mathematical skills in high school students, revealed a notable improvement in their academic performance. After using the learning objects suggested by the Random Forest model, students obtained better results in the final test, indicating progress in their understanding of the topics.

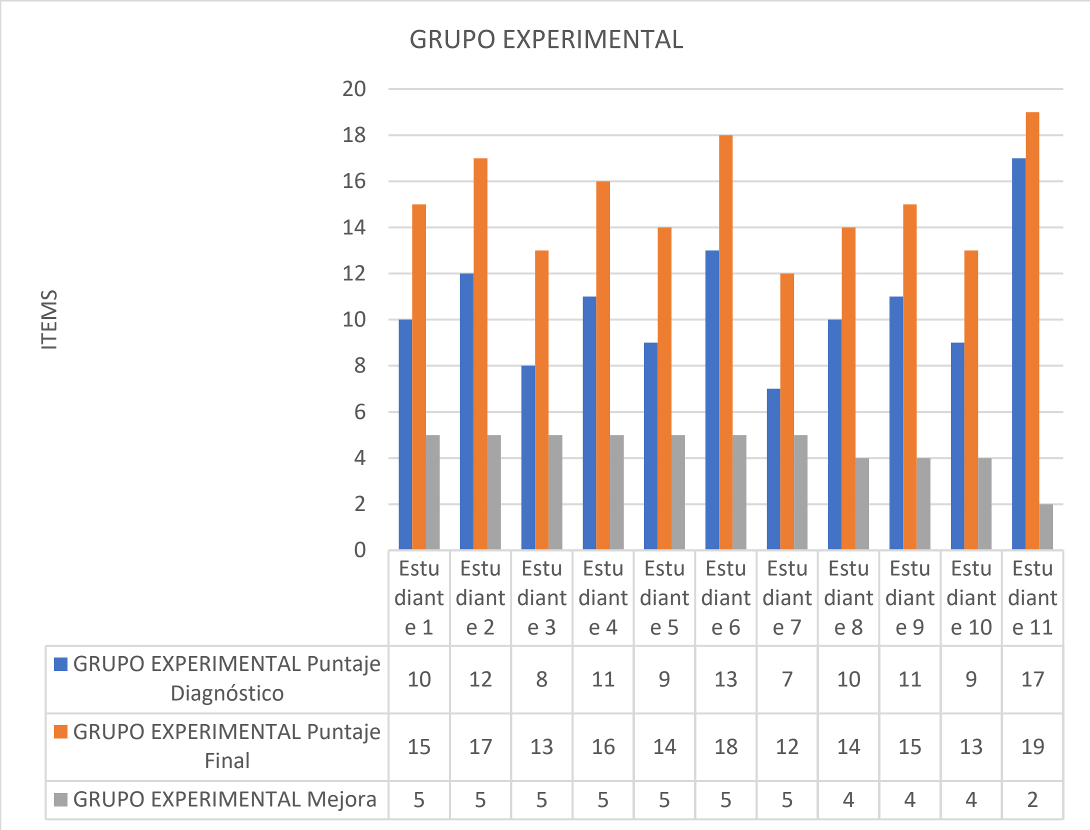
This result supports the use of data science techniques, such as the Random Forest algorithm, to identify behavior patterns and learning profiles that enable the recommendation of personalized digital resources.

The predictive model proved to be an effective educational support tool, classifying learning objects according to variables such as frequency and duration of use, perceived impact on grades, and motivation, which made it possible to accurately predict the most appropriate resource for each student (such as videos, interactive exercises, or simulations).

Thanks to this automated personalization, the individual learning needs were addressed, to achieve an improve of 45.24%in the academic performance. This was validated through the comparison between the results, of both instruments as shown in the graphic 8.

Graphic 8

Experimental comparison



Source: own elaboration.

Furthermore, statistical analysis revealed that progress was uniform across the different groups evaluated, suggesting that the OA recommendation not only benefits a small group of students, but also has a widespread effect. This discovery underscores the importance of incorporating predictive analytics tools into curriculum development and pedagogical decision-making. Likewise, it was demonstrated that student motivation plays a crucial role, given that those who showed greater interest obtained more outstanding results in the final test.

REFERENCES

Aros, A. S. (2025) Predictive Analysis of the Use of Learning Objects for High Schools in the Teziutlán Region [Master's Thesis, Instituto Tecnológico Superior de Teziutlán]

De Medio, C., Limongelli, C., Sciarrone, F., & Temperani, M. (2019). MoodleREC: A recommendation system for creating courses using the Moodle e-learning platform. Computers in Human Behavior. https://doi.org/10.1016/j.chb.2019.106168

Hadayat Ali, H., Muzammal Hussain, M., Fazil, H., Amjad, F., & Nazir, M. (2024). Navigating the future: Trends and innovation in Education. Qualitative Research, 345-366.

Hernández Osorio, J. del C. (2023). La Deserción Escolar en la Educación Media Superior. Ciencia Latina Revista Científica Multidisciplinar, 7(6), 1676–1700. https://doi.org/10.37811/cl_rcm.v7i6.8801

Hotz, N. (December 9, 2024). What is CRISP DM? Retrieved from Data Science PM: https://www.datascience-pm.com/crisp-dm-2/

J. L. García Cué, J. M. Meraz Escobar, & M. A. Jiménez Velázquez, Learning objects as support in the teaching-learning process (p. 87). Mexico: General Directorate of Publications

National Institute of Statistics and Geography. (2025). Interactive tables-Generic. Retrieved from INEGI: https://www.inegi.org.mx/app/tabulados/interactivos/?pxq=Educacion_Educacion_12_987437b7-5398-4067-8b19-a1cd2ec36b7b

National Institute of Statistics and Geography. (2023). National Survey on the Availability and Use of Information Technologies in Households. Retrieved from INEGI: https://www.inegi.org.mx/programas/endutih/2023/

Global Compact. (January 23, 2025). 17 Development Goals to transform the world. Retrieved from Global Compact Network Spain: https://www.pactomundial.org/que-puedes-hacer-tu/ods/

OpenAI. (2024). ChatGPT (Versión GPT-4o) [Modelo de lenguaje grande]. https://chatgpt.com

Orellana-Cordero, M., García-Herrera, D. G., Erazo-Álvarez, J. C., & Narváez-Zurita, C. I. (2020). Objetos virtuales interactivos con Genial.ly: Una experiencia de aprendizaje matemático en bachillerato. CIENCIAMATRIA, 6(3), 309–332. https://doi.org/10.35381/cm.v6i3.403

Palmera Quintero, L. M., Muñoz Morales, L. A., Guerrero Sumalave, D. F., & Holguín Uribe, Y. (2023). Objeto virtual de aprendizaje para la enseñanza: instrumento de apoyo para la inclusión de la informática en niños con discapacidad visual. Mundo Fesc, 13(27), 38–49. https://doi.org/10.61799/2216-0388.135439

Pulido Ocampo, C. A., Martínez Moreno, P., & Vergara Camacho, J. A. (2016). Learning objects as a teaching resource in the Algorithmic educational experience. In J. A. Hernandez Moral, A. Ronquillo Bolaños, A. B. Celis Domínguez, A. Lagunes Domínguez, A. P. Lorandi Medina, & A. D. Otero Escobar, Trends and challenges in educational innovation: an open debate. Mexico: Foundation for International Higher Education.

Rodríguez, N. W. (2023). Objetos virtuales de aprendizaje en el marco de la educación 4.0: una alternativa pedagógica mediada con las TIC para el logro de aprendizajes significativos. Dialéctica, 1(20). https://doi.org/10.56219/dialctica.v1i20.1669

Serna Martínez, R. E., & Alvites Huamaní, C. G. (2021). Plataformas educativas: herramientas digitales de mediación de aprendizajes en educación. HAMUT'AY, 8(3), 66. https://doi.org/10.21503/hamu.v8i3.2347

Tovar Vidal, M., Zepeda Cortés, C., & Castillo Zacatelco, H. (2019). Objetos de Aprendizaje Abiertos basados en ADDIE. Caso de estudio: Colegio de Posgraduados. En J. L. García Cué, R. C. Medina, & M. A. Jiménez Velázquez (Eds.), Los Objetos de Aprendizaje como apoyo en el proceso de enseñanza-aprendizaje (pp. 87-105). Colegio de Posgraduados.

Triola F. Mario. (2004). Estadística (Pearson Educación, Ed.; 9a ed.).

APPENDIX

During the course of this research, artificial intelligence (AI) tools were used to assist with tasks such as preliminary drafting of sections of the document and grammatical correction. The use of AI was limited to auxiliary functions, and all methodological decisions, interpretations of results, and conclusions were made by the authors; AI did not intervene in the methodological design or statistical analysis.

The article is derived from the thesis “Predictive Analysis of the Use of Learning Objects for High Schools in the Teziutlán Region” presented by Sara Aros Alberto to obtain a Master's degree in Computer Systems at the Instituto Tecnológico Superior de Teziutlán on September 3, 2025.