Expert Systems

WILEY

Expert Systems

ORIGINAL ARTICLE

Exploring Machine Learning Algorithms for Analysing Students' Attitudes Towards Distance Mathematics Learning

Marina Svičević[1] | Aleksandar Milenković[1] | Lazar Krstić[1] | Miloš Pavković[2]

[1]Faculty of Science, University of Kragujevac, Kragujevac, Serbia | [2]Singidunum University, Belgrade, Serbia

Correspondence: Marina Svičević (marina.svicevic@pmf.kg.ac.rs)

Received: 16 August 2025 | **Revised:** 20 May 2026 | **Accepted:** 22 June 2026

Keywords: data preparation techniques | distance mathematics learning | machine learning algorithms | students' attitudes

ABSTRACT

This study investigates the application of machine learning algorithms to analyse students' attitudes towards distance mathematics education, focusing on perceived effectiveness and students' ability to successfully learn and adopt mathematical content in an online setting. Data were collected from 1154 students at various educational levels using a 28-item Likert scale questionnaire on distance mathematics learning. An ML pipeline incorporating multiple data preparation techniques and machine learning algorithms was applied to two key prediction questions. Unlike predominantly descriptive or single model studies in this area, this approach evaluates both predictive performance and the stability of selected survey items across many model configurations, providing more robust and interpretable pedagogical insights. The results show that Recursive Feature Elimination was the most effective feature selection method, while Random Forest, Ridge Classifier and Categorical Naive Bayes achieved the strongest overall predictive performance across the two questions. These findings confirm the value of combining feature selection techniques and machine learning algorithms to derive robust and interpretable insights from educational survey data, while also highlighting the importance of well-structured distance learning strategies in mathematics education. The methodology is readily adaptable to other academic disciplines, providing educators with a data-driven framework for improving the design and effectiveness of online learning.

1 | Introduction

Distance learning is a concept that has existed for more than a century, with its roots in the correspondence schools of the 19th century. The development of various media, from radio and television to modern information and communication technologies, has significantly shaped this form of education (Matić and Stančić 2021; Sherry 1995). Its implementation is intended for individuals who cannot attend traditional classes and is carried out with the help of various media (Lučev et al. 2022). The main advantages of this approach include flexibility, personalized instruction and broader accessibility for more students (Descamps et al. 2006), leading to the conclusion that it is a complex system that depends on multiple stakeholders and their capabilities (Cassibba et al. 2021).

Distance learning has become an important component of contemporary education, particularly following the COVID-19 pandemic, which accelerated the transition from traditional classroom instruction to online environments across all educational levels (Traxler 2018). In Serbia, this transition represented a significant challenge for mathematics teachers who were primarily accustomed to traditional teaching methods, while students had to adapt to learning in virtual environments (Vučetić et al. 2020). Research conducted during this period showed that students recognized certain advantages of online mathematics instruction, including time savings, better organization and easier access to learning materials, although teachers frequently emphasized the lack of direct communication and concerns regarding students' independent work (Vučetić et al. 2020; Cvijetić et al. 2022).

Similar findings have been reported in international studies examining distance learning during and after the COVID-19 pandemic. Existing research highlights challenges related to technological infrastructure, unequal access to digital resources, reduced interaction and limited preparedness of institutions and teachers for fully online instruction (de la Peña Esteban et al. 2019; Shevchenko et al. 2024; Albanyan 2024; Gurajena et al. 2021; Steyn et al. 2024). At the same time, flexibility, continuity of education and support for geographically or socially disadvantaged students are recognized as important advantages of distance learning. A recent systematic review confirms that technological limitations, organizational barriers and reduced interaction remain among the key factors affecting student engagement and perceived learning effectiveness in online environments (Wang and Huang 2023).

Although there were certain shortcomings in the implementation of online teaching during the pandemic, it also demonstrated positive aspects, which have led to ongoing improvements since the pandemic ended. Research on the integration of artificial intelligence (AI) in distance learning at the higher education level shows that students regularly use AI technology and report a high level of satisfaction with their use. Significant benefits have been observed in areas such as learning through virtual and augmented reality. However, the study also emphasizes the need to address various challenges in order to fully realize the advantages of AI technology in teaching (Mahmudi et al. 2023). This research remains largely focused on AI-enabled distance learning in general and does not directly examine students' attitudes towards distance mathematics learning through a comparative machine-learning evaluation. Additionally, among the proposed models to address the challenges of distance learning, the pedagogical approach known as COMP-PLETE is particularly notable. This approach incorporates eight key elements: community, openness, multimodality, participation, personalization, learning, experience and technological improvement. The model is designed to improve the online learning experience, reduce students' sense of isolation and encourage active participation in the learning community. The authors provide a theoretical rationale and recommendations for implementing this model, with an emphasis on creating dynamic and supportive communities that can effectively address challenges related to technology access, digital literacy and time management (Goria and Konstantinidis 2023). Although COMP-PLETE provides a comprehensive pedagogical framework, it does not offer a subject-specific, data-driven assessment of students' attitudes in distance mathematics education.

Despite the extensive body of research on distance learning, much of the literature addresses online education at a general level and focuses on broad constructs such as students' motivation, engagement and acceptance rather than subject-specific contexts (Mohtar and Md Yunus 2022). Mathematics education represents a particularly demanding context due to the abstract nature of mathematical concepts, the cumulative structure of knowledge and the strong reliance on continuous feedback and interaction. Consequently, findings derived from general distance learning research cannot be directly transferred to distance mathematics education. Moreover, although students' experiences and attitudes in remote mathematics learning have been explored, existing studies remain predominantly descriptive or mixed-method and do not offer a comprehensive, model-based analysis (Mukuka et al. 2021; Mihajlović et al. 2021). Most prior studies in this domain rely on descriptive statistics or qualitative approaches to examine students' perceptions, identifying broad advantages and challenges without offering predictive insights (Vučetić et al. 2020; Cvijetić et al. 2022; Mukuka et al. 2021). Where machine learning has been applied, analyses are typically limited to a single classifier or a narrow feature set, without systematic comparison of feature selection strategies or assessment of model stability across different configurations (Milenković, Krstić, and Svičević 2024). Consequently, it remains unclear which survey items carry consistent predictive value across varying modelling choices, and whether a multi-model pipeline can yield robust and interpretable insights into students' attitudes towards distance mathematics education specifically. This lack of comprehensive analyses that systematically combine multiple feature selection techniques with a wide range of machine learning models provides the primary motivation for the present study.

1.1 | Distance Mathematics Education

Distance mathematics education presents a distinct challenge compared to other fields of study largely due to its interactive nature, the need for continuous feedback and the complexity of mathematical concepts, which often require visual and hands-on support. Mathematics demands a high level of student engagement, which can be further limited in an online environment due to the lack of direct interaction and technical limitations, such as the absence of interactive tools for displaying formulas, graphs or diagrams. At the same time, learning mathematics remotely offers opportunities to utilize digital technologies, such as simulations, interactive applications and visualization tools, that can enrich the teaching process (Engelbrecht and Harding 2005; Lazović 2021; Lowe et al. 2016). Empirical studies indicate that the proper use of multimedia resources and educational software can improve learning outcomes (Engelbrecht and Harding 2005; Murtafiah et al. 2020). During the pandemic, teachers adopted a variety of tools to make mathematical content more accessible to students (Djordjić et al. 2021). Individual factors, such as access to technology, personal learning space and digital literacy, significantly affect students' self-perception in mathematics within the online environment. Studies show that both technological and personal challenges can negatively influence students' perception of mathematical capabilities, whereas pedagogical interventions and curriculum adjustments can help mitigate these challenges (Bringula et al. 2021). Their level of education increases, pointing to a relationship between students' opinion towards online learning and their age (Adalar and Oflaz 2023). During the pandemic in Serbia, most teachers used pen tablets to present mathematical content. Students found this tool highly useful for both lessons that introduced new material and practice sessions. Girls, more than boys, reported that this type of instruction helped them better understand mathematical content, particularly in algebra lessons. When considering students' age, high school students, more than elementary school students, perceive the use of a pen tablet as more effective and consider it more beneficial in mathematics classes (Milenković, Milikić, and Jovović 2024). In addition to technical issues, the lack of direct communication and challenges in monitoring and evaluating student work remain major concerns (Almarashdi and Jarrah 2021; Mihajlović et al. 2021). Teachers also face difficulties in keeping students engaged and providing adequate support in an online setting. Research also points to differences in attitudes between teachers and students towards online mathematics teaching, which can impact its effectiveness (Milenković et al. 2022). Applying the unified theory of acceptance and use of technology (UTAUT) model, research has found that factors such as perceived effectiveness, expected effort, facilitating conditions and social influence all contribute to the acceptance of online teaching. Of these, ensuring that students have access to the necessary technology is the most important factor for successful adoption (Kansiime and Batiibwe 2023). Given all these factors, it is essential to develop new approaches and methods that will enable effective distance learning in mathematics, taking into account the specific characteristics of the domain and the needs of students.

1.2 | Machine Learning as a Tool for Understanding Students' Attitudes

Predicting student performance in programming and mathematics often involves algorithms such as Random Forest, Decision Trees and Naive Bayes, which have demonstrated high accuracy in classifying and predicting student outcomes (Durak and Bulut 2023; Milićević et al. 2024). Additionally, research in interactive learning environments has shown that neural networks perform exceptionally well in predicting learning outcomes, providing the best results in forecasting the timely completion of tasks (Su et al. 2022). At the same time, Naive Bayes outperformed Decision Trees in predicting programming performance, achieving an accuracy of 0.9102, indicating its usability in improving learning strategies and identifying student weaknesses (Sivasakthi and Padmanabhan 2022). Studies in mathematics emphasize the importance of algorithms such as Logistic Regression, k-Nearest Neighbours and Support Vector Machines in analysing student performance (Elouafi et al. 2024). Similarly, time series analysis offers the potential to predict student difficulties with a high level of accuracy, allowing timely interventions and helping to prevent dropout (Shou et al. 2024). Deep learning models, including convolutional neural networks (CNNs), long short-term memory (LSTM) and deep neural networks (DNNs), have also shown significant potential for predicting performance in virtual environments. These algorithms achieve high accuracy when analysing data from online platforms, emphasizing student engagement as a key success factor (Alnasyan et al. 2024). Similarly, the application of the k-Nearest Neighbours algorithm in studies of students' attitudes towards distance mathematics education demonstrates that this model can effectively classify students' perceptions of teaching effectiveness and their ability to understand mathematical concepts in an online environment (Milenković, Krstić, and Svičević 2024). Beyond technological aspects, psychological factors, such as teacher anxiety, can significantly impact student success. Anxiety related to mathematics teaching has been identified as a relevant predictor of student achievement, highlighting the need to integrate psychological support into the teaching process (Awofala et al. 2024).

The application of machine learning to analyse student attitudes and performance provides opportunities for more personalized instruction and the development of more effective educational strategies. This is particularly important in mathematics education and also extends to broader educational contexts.

1.3 | Aim of the Study

This study addresses two research questions:

- Is distance mathematics education effective?
- Can students successfully acquire mathematical content through online learning?

To answer these questions, the study focuses on the application of machine learning to analyse and predict students' attitudes towards distance mathematics education. These research questions are implemented in the questionnaire as a part of dedicated survey items (the full wording of all 28 statements is provided in Appendix A). The central idea of this study is to examine whether machine learning models can predict each student's response to these two target items based on the pattern of responses to the remaining survey questions, which capture how students evaluate different aspects of distance mathematics learning. Specifically, the aim is to develop models capable of accurately predicting (i) students' perceptions of the effectiveness of mathematics teaching in an online environment and (ii) their perceived ability to acquire mathematical content during distance education.

The goal is to achieve this through a comprehensive approach that includes:

- 9 feature selection techniques,
- 14 different machine learning algorithms (23 including variations),
- 5 evaluation metrics applied to both training and test datasets

To the best of our knowledge, there are currently no studies that provide a detailed machine learning-based analysis of students' attitudes towards the quality of distance mathematics instruction in schools across Southeast Europe. By addressing this gap, the present research contributes to a deeper understanding of students' attitudes and provides evidence-based insights to support decision makers in designing more effective online mathematics instruction.

Beyond model comparison, the study also aims to identify survey items that remain consistently informative across different feature selection principles and learning algorithms. This

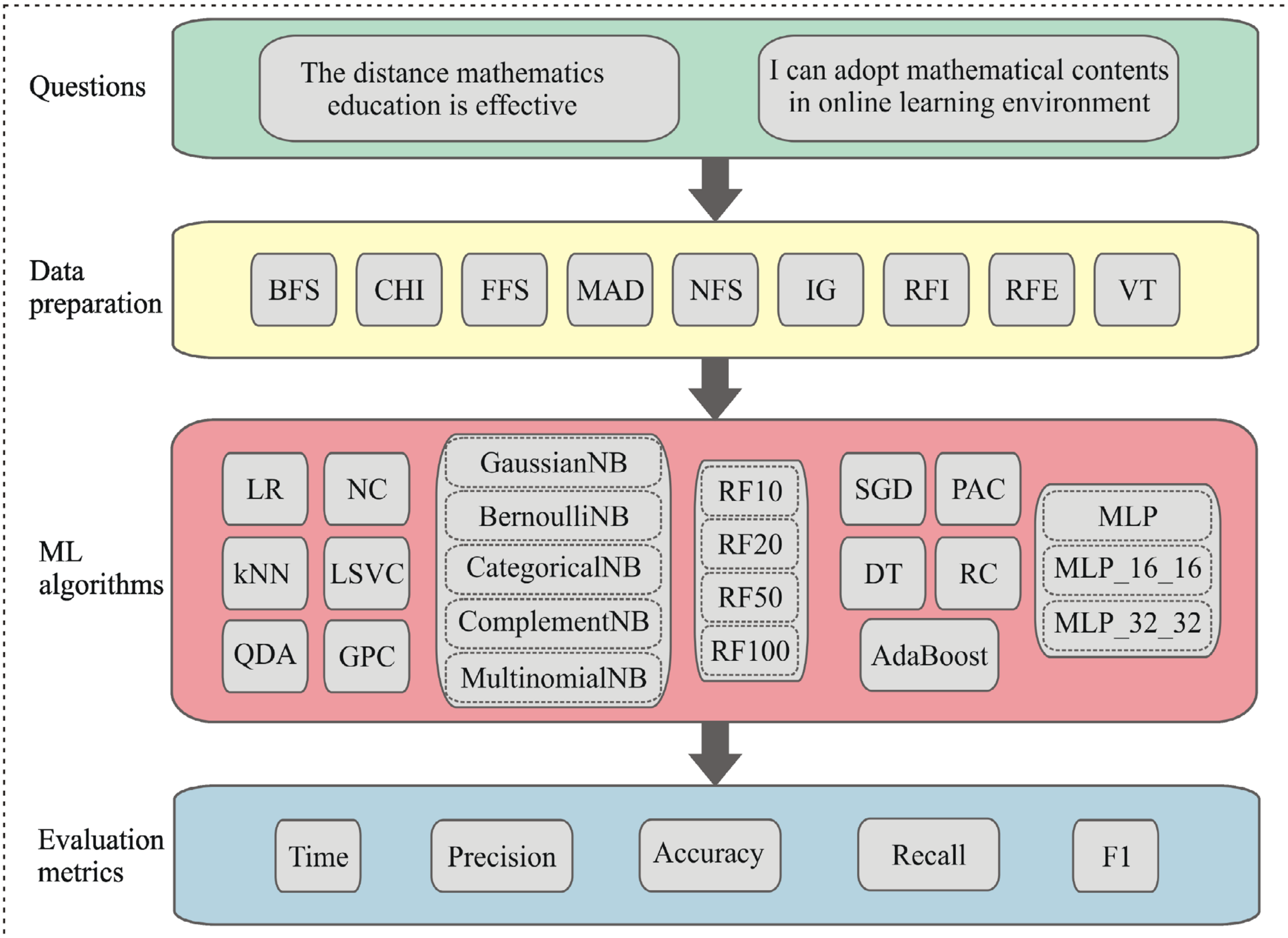

FIGURE 1 | An overview of the model development pipeline, illustrating the four levels of the process: Question formulation, data preparation techniques, machine learning model selection and evaluation metrics.

improves robustness and reduces dependence on any single modelling choice.

The main contributions of this paper are threefold. First, the study examines students' attitudes towards distance mathematics education through a comprehensive machine learning-based approach. Second, it systematically compares multiple feature selection techniques and machine learning algorithms across two key prediction questions derived from questionnaire data. Third, it identifies stable and interpretable survey items across different model configurations, offering data-driven pedagogical insights relevant to the design and improvement of distance mathematics learning.

This paper is structured as follows. Section 2 presents the research methodology, including the dataset description, data preparation techniques, feature selection strategies and the machine learning models applied in the analysis. Section 3 reports the experimental results and comparative performance of the evaluated models across both research questions. Section 4 discusses the findings, interprets the identified pedagogical patterns and outlines their educational implications. Finally, Section 5 summarizes the main conclusions, highlights the study's contributions and suggests directions for future research.

2 | Methodology

This section presents the methodology used to analyse students' attitudes towards distance mathematics learning. The methodological approach consists of three key components: data preparation and selection, application of machine learning algorithms and evaluation of models using a variety of metrics. This research dataset is based on the answers to the survey statements that allow the implementation and comparison of multiple methods and techniques.

The given diagram (Figure 1) visually represents the process of building and evaluating machine-learning models for predicting students' attitudes towards distance mathematics learning. The process is organized into four main levels. The first level includes two research questions that the models are designed to address and predict. The second level covers data preparation using nine different feature selection techniques. The third level involves selecting one of the 23 machine learning algorithms to build the model (Mohammed et al. 2016; Sarker 2021). Finally, the fourth level involves evaluating the model using five metrics, providing a comprehensive assessment of its performance. By combining different data preparation techniques and algorithms, a total of $9 \cdot 23 = 207$ different models were built for each research question. Since this procedure is performed for two distinct questions, a total of 414 models were considered in the study.

Although the study addresses two distinct research questions, both are formulated as binary classification problems derived from the same questionnaire and rely on an identical set of predictor variables. For this reason, the same pool of feature selection techniques and machine learning algorithms was applied to each research question. Importantly, feature selection was performed independently for each target variable, ensuring that the relevance of features was evaluated separately for each research question. This design choice enables a systematic and fair comparison of model behaviour and stability across different predictive tasks, rather than optimizing algorithmic configurations for a single question. Such an approach allows the analysis to focus on the robustness and general applicability of feature selection strategies and machine learning models in the context of students' attitudes towards distance mathematics education.

2.1 | Dataset

A total of 1154 students from various educational levels participated in creating the dataset. The sample included 758 upper primary school students (65.68%) and 396 high school students (34.32%). The participants were drawn from 11 upper primary schools and 5 high schools. Among the students surveyed, 664 (57.54%) were female, while 490 (42.46%) were male. The student sample was collected in Serbia at the end of the 2021/2022 school year using a convenience sampling approach: the survey link was distributed via participating schools through official school and teacher (professor) email channels, and students who received access were invited to complete the questionnaire voluntarily. All questionnaire fields were mandatory, therefore, no incomplete survey forms were submitted and no responses had to be excluded due to missing data. Although the sample was non-probabilistic, its size and structure support meaningful exploratory analysis and the application of machine learning methods, while limitations regarding representativeness are acknowledged. Alongside three demographic questions on gender, school type and grade level, two questionnaire items (12 and 25) corresponded directly to the study's research questions and were used as target variables in the subsequent machine learning analysis.

Responses to all questionnaire statements were collected using a five point Likert scale encoded from 1 to 5. Since Items 12 and 25 operationalize the two research questions as binary classification tasks, their response categories were binarized by grouping 1–2 as disagreement and 4–5 as agreement. Accordingly, prior to the analysis, respondents who selected the neutral midpoint (value 3, interpreted as undecided) on items 12 or 25 were excluded, so that the final sample of 1154 students (76% of the initially collected responses) consisted of only those expressing either disagreement (1–2) or agreement (4–5) with the corresponding statement.

This survey format allowed for a detailed analysis of student perceptions and served as the basis for operationalizing the two research questions introduced in Section 1, which examine the perceived effectiveness of distance mathematics education and students' ability to acquire mathematical content through online learning.

Since the 12th and 25th survey statements were used as target variables corresponding to the two research questions, they were excluded from the input feature set and treated as output variables in the model training process. The remaining questionnaire items were developed in line with prior research on distance mathematics education that highlights the role of digital tools and learning materials, teacher feedback and assessment practices and broader organizational access constraints (Mihajlović et al. 2024; Chin et al. 2022), while also reflecting the Serbian emergency remote education context that combined online platforms with televised instruction via RTS (Ranđelović et al. 2020).

To assess the internal consistency and quality of the questionnaire, reliability analysis was performed using Cronbach's alpha coefficient on all 28 Likert type items. The obtained value of $\alpha = 0.91$ indicates excellent internal consistency. Content validity was ensured through comprehensive coverage of the key dimensions related to distance mathematics learning.

2.2 | Data Preparation

Data preparation is a crucial step in machine learning, with feature selection being one of the most significant phases. Various feature selection techniques are used to extract the most relevant statements from the dataset, which can enhance model performance and reduce computational complexity.

The set of nine feature selection techniques was deliberately chosen to cover complementary categories of feature selection approaches commonly used in machine learning. Specifically, the selected methods include filter-based techniques (Variance Threshold, Median Absolute Deviation, Chi-square, Information Gain), wrapper-based techniques (Recursive Feature Elimination [RFE], Backward Feature Selection, Forward Feature Selection), embedded methods (Random Forest Importance, RFI) as well as a baseline scenario without feature selection. This combination enables a systematic comparison of fundamentally different selection principles in the context of attitude prediction.

Although the questionnaire responses are collected using a Likert scale, the items are treated as ordinal numerical features, which is a standard practice in predictive modelling and educational data mining. In practice, each response option was encoded as an ordered integer from 1 to 5 (preserving response order). This representation is directly compatible with the applied selection methods, including χ^2-based scoring which requires non negative inputs. The applied feature selection techniques do not rely on strict distributional assumptions and are therefore suitable for handling Likert type data in a machine learning setting, where the primary objective is prediction rather than statistical inference (Koklu 2025; Woelk et al. 2025). In this study, the term *data preparation* refers to feature selection (plus the baseline without feature selection) applied to the Likert items prior to model training.

2.2.1 | No Feature Selection (NFS)

In this approach, all data are used without applying any feature selection. This approach is suitable for models that are robust to a large number of input features and can handle potential noise.

2.2.2 | Variance Threshold (VT)

This technique removes features with low variance, eliminating those that have nearly identical values across all input samples. Features with higher variance are more likely to contribute to distinguishing between target classes.

2.2.3 | Median Absolute Deviation (MAD)

This technique is a statistical method used to remove features with low variability. Features with low MAD values often do not contribute to the predictive power of the model and can be eliminated.

2.2.4 | Chi-Square (CHI)

The chi-square test-based technique is used to measure the independence between features and the target variable. The most useful features are those that exhibit the strongest statistical association with the target variable.

2.2.5 | Information Gain (IG)

This method measures how much information each feature provides about the target variable. It is widely used with Decision Trees (DT) and similar algorithms, where features with the highest information gain are considered the most relevant.

2.2.6 | Random Forest Importance (RFI)

This technique uses the Random Forest algorithm to calculate the importance of each feature. Features with the highest importance score are considered most relevant for the model. In this study, features were ranked by their RFI scores, and selection was performed by retaining the most informative items according to the retention criterion described in Section 3.1.

2.2.7 | Recursive Feature Elimination (RFE)

RFE is an iterative method that trains a model and removes the least relevant features based on the weights assigned by the model. The process continues until an optimal set of features remains. In this study, RFE was implemented using a linear estimator, where feature importance was determined based on model coefficients. Prior to applying RFE, all features were normalized to ensure comparability and to prevent scale-related bias in the elimination process. The number of retained features was determined through iterative evaluation of model performance, allowing the selection of a reduced feature subset that preserved predictive accuracy.

2.2.8 | Backward Feature Selection (BFS)

This technique begins with all available features and gradually removes them one by one based on their impact on the model. The final set of features includes those that yield the best results in terms of accuracy or other evaluation metrics.

2.2.9 | Forward Feature Selection (FFS)

The opposite of Backward Feature Selection, this method starts with an empty set of features and adds them one by one, evaluating the model's performance after each addition. The process continues until optimal results are achieved. Both BFS and FFS were implemented as wrapper procedures that iteratively evaluate candidate feature subsets using the same cross-validation protocol as the downstream classification experiments, selecting the subset that yielded the best performance.

FFS and BFS are often slower compared to other methods because they involve multiple steps in adding or removing features. Techniques such as CHI and IG rely on statistical measures to determine feature importance, while RFI and RFE directly use machine learning algorithms for assessment. VT and MAD are simpler methods that do not require complex calculations or modelling. All feature selection techniques were applied using standard implementations with default parameter settings unless otherwise stated. For score-based filter methods (e.g., CHI and IG), a common relevance threshold described in Section 3.1 was used to retain features. For methods with inherent selection rules (e.g., VT/MAD removing only constant items, and RFI using cumulative importance), the corresponding default or method-specific procedure was applied. RFE inherently relies on an iterative model-based elimination procedure.

2.3 | Machine Learning Methods

All machine learning algorithms were applied within a unified experimental framework to ensure comparability across models. A diverse set of well established methods was employed to predict students' perceptions of the effectiveness of distance mathematics learning and their ability to acquire mathematical content in an online environment, selected for their complementary characteristics (e.g., linear and nonlinear decision boundaries, probabilistic modelling and ensemble-based learning). For each combination of feature selection technique and learning algorithm, models were trained separately for each research question using the same cross-validation strategy and evaluation metrics, with the primary aim of performing supervised classification based on students' responses to the selected survey items rather than investigating the theoretical properties of the models themselves.

After excluding the two target statements from the predictors, the remaining survey items were used as inputs. For each research question, each feature selection technique produced a reduced feature set, which was then provided to each classifier to learn a predictive mapping from students' responses to the corresponding binary outcome. Model performance was estimated under a stratified fivefold cross-validation protocol to ensure consistent class proportions across folds and enable a fair comparison across all feature selection × algorithm combinations.

2.3.1 | Logistic Regression (LR)

Logistic Regression was used as a baseline linear classifier due to its interpretability and robustness when handling binary outcome variables derived from Likert-scale data.

2.3.2 | Nearest Centroid (NC)

The NC classifier was applied as a simple distance-based method, assigning instances to classes based on proximity to class centroids in the selected feature space.

2.3.3 | k-Nearest Neighbours (kNN)

The kNN algorithm was employed to model local similarity structures in the data, enabling classification based on the dominant class among neighbouring instances.

2.3.4 | Linear Support Vector Classification (LSVC)

Linear SVC was used to construct linear decision boundaries that maximize class separation, particularly suitable for high-dimensional feature spaces.

2.3.5 | Gaussian Process Classifier (GPC)

The Gaussian Process Classifier was applied to capture nonlinear relationships while providing probabilistic predictions, although its use was limited by computational complexity.

2.3.6 | Decision Trees (DT)

Decision Trees were used to model hierarchical decision rules, allowing for transparent classification based on selected survey items.

2.3.7 | Random Forest (RF)

Random Forest models were implemented as ensemble classifiers combining multiple decision trees to improve generalization and reduce overfitting. Variants with different numbers of trees were considered.

2.3.8 | Adaptive Boosting (AdaBoost)

AdaBoost was applied as an ensemble method that incrementally emphasizes misclassified instances, thereby improving predictive performance in challenging classification scenarios.

2.3.9 | Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron models were used to capture complex nonlinear relationships between input features and target variables. Several network configurations were evaluated to assess model stability.

2.3.10 | Naive Bayes (NB)

Naive Bayes classifiers were employed as probabilistic models based on conditional independence assumptions, offering computational efficiency and robustness in high-dimensional settings.

2.3.11 | Quadratic Discriminant Analysis (QDA)

QDA was used to model class-specific covariance structures, allowing for nonlinear decision boundaries under Gaussian assumptions.

2.3.12 | Stochastic Gradient Descent (SGD)

SGD based classifiers were applied as efficient linear models optimized through iterative updates, suitable for large scale experimental settings.

2.3.13 | Passive Aggressive Classifier (PAC)

The Passive Aggressive classifier was used as an online learning algorithm that updates model parameters only in response to misclassified instances.

2.3.14 | Ridge Classifier (RC)

The Ridge Classifier was applied as a regularized linear model to mitigate multicollinearity and enhance numerical stability.

Hyperparameter tuning was carried out using a comprehensive grid search within cross-validation. For each algorithm, we varied a small set of key capacity and regularization hyperparameters (e.g., neighbourhood size for kNN, regularization strength for linear models, kernel related parameters, network configuration for MLP and the number of estimators for ensemble methods) and selected the configuration that maximized the macro-F1 score. To ensure a fair comparison and prevent information leakage, model specific preprocessing and feature selection were performed using training data only within each cross-validation split, and the selected configuration was then evaluated consistently under the same stratified k-fold protocol used throughout the study.

2.4 | Evaluation Metrics for Machine Learning Models

The evaluation of machine learning models is an important step in the process of developing and implementing algorithms, as it provides a way to measure their accuracy, efficiency and reliability. In this study, various performance evaluation metrics are applied, including time-based metrics, accuracy, precision, recall and the F1 score. Each of these metrics provides valuable insights into different aspects of model performance, ranging from processing speed to the model's ability to correctly classify data.

Time-based metrics assess the efficiency of the model in terms of the time required for training and testing. The average training time, *fit_time_mean*, represents the mean time required to train the model, while the average testing time, *score_time_mean*, indicates the time needed to make predictions. Their standard deviations, *fit_time_std* and *score_time_std*, provide insights into the stability of the execution time:

$$fit\_time\_mean = \frac{1}{n} \sum_{i=1}^{n} t_{fit,i},$$
$$score\_time\_mean = \frac{1}{n} \sum_{i=1}^{n} t_{score,i}.$$

Accuracy measures the overall success of the model and is defined as the proportion of correctly classified samples to the total number of samples:

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

where TP (true positive), TN (true negative), FP (false positive) and FN (false negative) represent different classification outcomes. In this study, both the mean and standard deviation of accuracy for the training and testing datasets were analysed.

Precision evaluates the accuracy of positive predictions and is defined as follows:

$$Precision = \frac{TP}{TP + FP}.$$

Recall (also known as Sensitivity or True Positive Rate) measures the model's ability to identify all actual positive instances:

$$Recall = \frac{TP}{TP + FN}.$$

F1 score represents the harmonic mean of Precision and Recall, serving as a balance between them:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}.$$

In this study, accuracy, precision, recall and *F1* score were analysed through their mean values and standard deviations across both the training and test datasets. By analysing these metrics in their micro, macro and weighted variants, detailed information was obtained on the model's stability and its ability to provide consistent predictions across datasets. This allowed the identification of models with the best performance in line with the defined research objectives.

3 | Results

This section provides a detailed analysis of the performance of machine learning models developed to address two key questions related to the effectiveness of distance mathematics education and the adoption of content in a digital environment. By combining nine different feature selection techniques and 23 machine learning algorithms with two key questions, a total of 414 models were developed in the Python library `Scikit-learn` (Kramer 2016), after which they are evaluated according to multiple criteria. The metrics used in the analysis include accuracy, precision, recall, F1 score and training and testing times, providing a comprehensive assessment of their performance. The aim of this analysis is not only to identify feature selection techniques and models that achieve the best results, but also to understand the impact of different data preparation techniques and evaluation metrics on the overall success of the models in practical applications. To evaluate the models, a fivefold cross-validation strategy was applied, with each fold using 80% of the data for training and 20% for testing. The use of the `StratifiedKFold` method from `Scikit-learn` library, ensured that the class distributions were preserved across all training and testing subsets. This approach guarantees that each class present in the test set is also represented in the training data, ensuring consistency and reliability in assessing model performance.

3.1 | Comparison of Data Preparation Techniques

To better understand the impact of different feature selection strategies, we first compare multiple data preparation techniques applied to the survey dataset. Table 1 summarizes the results of the selection of statements using nine different data preparation techniques for both research questions. To determine relevance, each statement was assigned a score that indicated its importance, as calculated by the respective selection method. A threshold value of 0.7 was applied to score based feature selection outputs to determine whether a feature should be retained. If a statement exceeded the 0.7 threshold within a particular feature selection technique, it was retained for that technique's result set and considered for further analysis. The choice of 0.7 in this case reflects a preference for features with strong relevance while maintaining enough diversity to avoid over-pruning the feature set. It also reduces the reliance on any single method and increases the robustness.

Several features, such as statements 13, 14, 26, 27 and 28, consistently exceeded the threshold across techniques for both questions, indicating their strong predictive relevance. These features are associated with students' self-assessed motivation, which proved to be a key predictive factor. Additionally, this suggests that in other domains, statements that target a similar context may also be effective predictors of learning success. On

TABLE 1 | Overview of selected features by data preparation techniques.

Data preparation	Gender	Grade level	Type of school	Statement No. 1	2	3	4	5	6	7	8	9	10	11	13	14	15	16	17	18	19	20	21	22	23	24	26	27	28
BFS—Q1	+	−	−	+	−	+	−	−	+	+	−	+	−	−	+	−	−	+	−	+	+	+	+	−	+	+	+	−	−
BFS—Q2	+	−	+	−	−	−	−	−	+	+	−	+	+	+	−	+	−	−	+	+	+	+	+	−	−	+	+	−	−
CHI—Q1	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	+	+	−	−	−	−	+	−	−	−	+	+	+	−
CHI—Q2	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	−	−	−	−	+	−	−	−	+	+	+	−
FFS—Q1	+	−	−	+	−	−	+	−	−	−	+	−	−	+	+	−	+	+	−	−	+	+	+	−	+	+	−	+	+
FFS—Q2	−	−	−	+	+	+	+	−	+	−	+	+	−	+	−	−	−	−	−	−	+	+	+	−	+	+	−	+	+
IG—Q1	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	+	+	−	−	−	−	+	−	−	−	+	+	+	−
IG—Q2	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	+	+	−	−	−	−	+	−	−	−	+	+	+	−
MAD—Q1	−	+	−	−	+	−	+	+	−	−	−	+	−	+	+	−	+	−	−	−	−	−	+	+	−	−	+	−	−
MAD—Q2	−	+	−	−	+	−	+	+	−	−	−	+	−	+	+	−	−	−	−	−	−	−	+	+	−	−	+	−	+
NFS—Q1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
NFS—Q2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
RFE—Q1	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+
RFE—Q2	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	−	+	−	−	−	−	+	−	−	−	+	+	+	+
RFI—Q1	+	−	−	+	−	−	−	−	−	−	+	−	−	−	+	+	+	−	−	−	−	−	+	−	−	+	+	−	−
RFI—Q2	−	+	−	−	−	−	−	−	+	+	−	−	+	−	−	−	−	−	−	−	−	+	+	−	−	+	+	+	−
VT—Q1	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
VT—Q2	−	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

the other hand, demographic attributes like Gender and Grade Level were selected sporadically, highlighting their limited generalization across models. Methods like NFS and VT tended to retain nearly all features, whereas CHI, IG and RFE applied more aggressive filtering, underscoring the variability in selection granularity between statistical, information-theoretic and model-based approaches.

To begin the detailed analysis of model performance, the F1 macro mean was selected as the primary metric. This metric was chosen because of several key advantages that contribute to a more comprehensive evaluation of model performance. First, the F1 score combines precision and recall into a single measure, allowing simultaneous consideration of the accuracy of positive predictions and the coverage of all relevant cases. Additionally, its macro variant calculates the average performance for each class, regardless of their representation in the data. This is particularly important in scenarios where there is a class imbalance. Finally, since this metric is measured on the test dataset, it directly reflects the model's ability to generalize predictions on new, unseen examples, which is crucial for assessing the model's practical applicability. Analysing the average values of the F1 metrics in the test dataset for nine different data preparation techniques in the context of both research questions (Figures 2, 3), significant differences in performance were identified. Certain techniques consistently achieve high values for both questions, indicating their stability and effectiveness in various scenarios. However, some techniques show considerable variation between questions, suggesting that their effectiveness may be problem-specific.

For the first question (Figure 2), the RFE technique achieves the highest average F1 macro score of 0.813, making it the best performing technique for this task. Its strength lies in its ability to gradually introduce the most relevant features, thereby optimizing the model's performance. It is followed by the NFS technique, with a score of 0.807, which highlights its efficiency in scenarios where the relationship between features and output classes is well defined. RFE also achieves a high maximum F1 macro score of 0.856, confirming its effectiveness in tasks that require systematic and rapid exploration of the feature space. Techniques such as IG and MAD yield solid but slightly lower results, with scores of 0.849 and 0.844, respectively. Although IG and MAD still offer solid performance, their effectiveness falls behind compared to more technically advanced methods, with FFS achieving the highest score (0.866), followed by CHI (0.859).

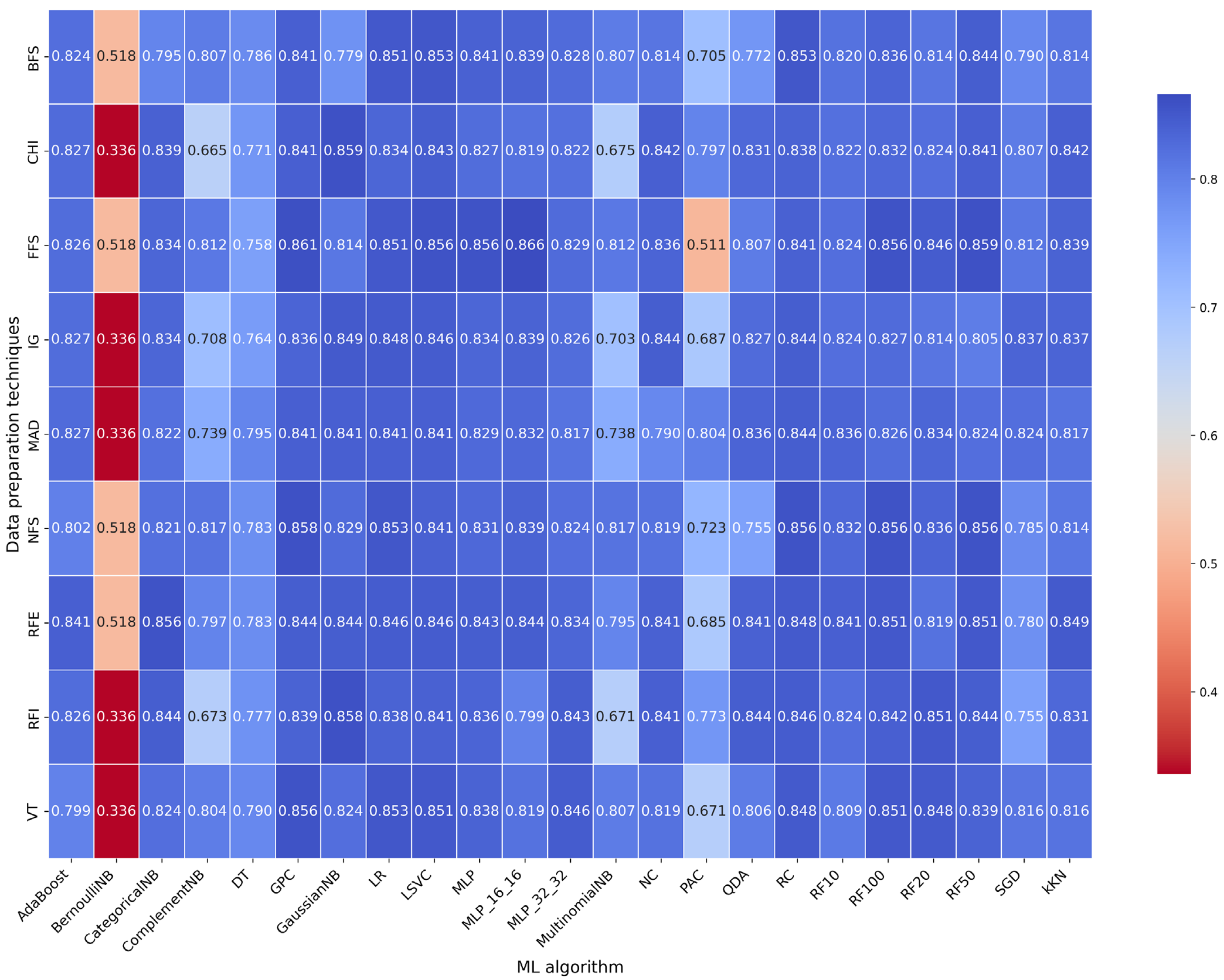

FIGURE 2 | F1 score heatmap of ML algorithms across data preparation techniques for first research question.

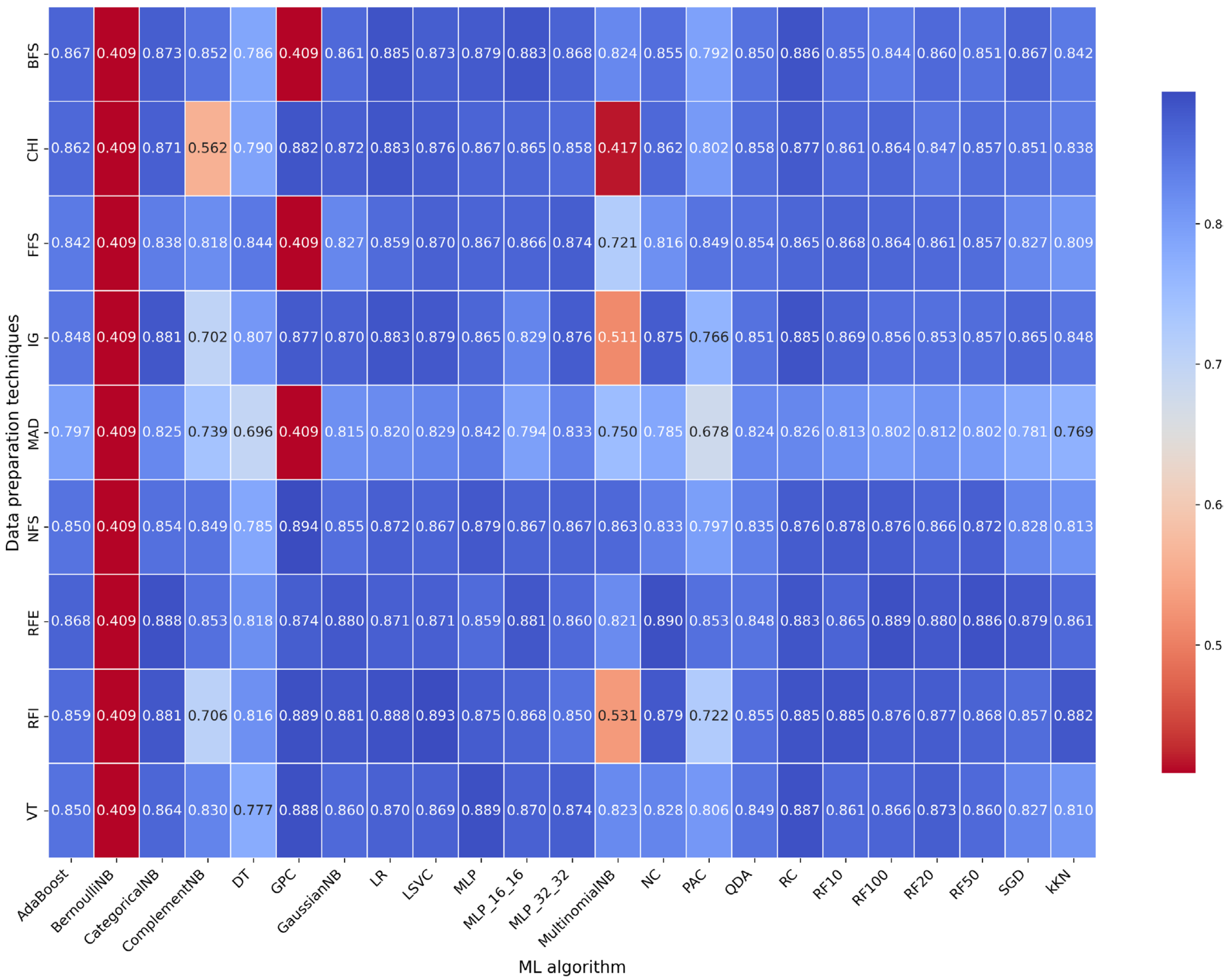

FIGURE 3 | F1 score heatmap of ML algorithms across data preparation techniques for second research question.

For the second question (Figure 3), the RFE technique achieves the highest average F1 macro score of 0.847, making it the best performing technique for this task. Its effectiveness lies in its ability to systematically prioritize the most relevant features, which enhances model performance. It is followed by the NFS technique, with a score of 0.834 and the VT technique with 0.832, both of which demonstrate solid performance in feature selection and model optimization. In terms of maximum F1 macro scores, NFS achieves the highest score of 0.894, confirming its strength in identifying the most predictive features. Close behind are RFI with a score of 0.893 and RFE with 0.890. Techniques such as VT and BFS also deliver high results, with scores of 0.889 and 0.886, respectively. In contrast, methods like MAD perform significantly worse, with a maximum F1 macro score of 0.842, indicating their limited suitability for tasks requiring complex feature interactions.

Based on the analysis of the maximum and average F1 macro mean scores for both questions, we can provide clear recommendations on which techniques should be further used. There are two key points to consider (Table 2): the performance of the technique in terms of its best score (MAX) and the consistency of its performance across different models, represented by the average score (Average). Techniques that achieve exceptionally high maximum values for both questions are NFS, RFI and RFE. NFS, with a maximum score of 0.894 for the second question and an average score of 0.834, demonstrates outstanding applicability and consistent performance. RFI achieves a maximum score of 0.858 for the first question and 0.893 for the second question. Although its average score for the first question (0.793) is not among the highest, it remains one of the top performers for the second question (0.823). RFE shows stability, with maximum scores of 0.856 for the first question and 0.890 for the second question. Its average score for both questions is the highest in the table, indicating its reliability across different tasks. On the other hand, techniques like FFS achieve maximum high values for the first question (0.866), but its average score for the second question (0.805) is lower compared to other techniques. This suggests that, although capable of achieving exceptional results with specific models, its stability is not on the same level as techniques such as NFS and RFE. Techniques such as CHI and VT also demonstrate notable results, but their maximum and average values do not exceed those achieved by NFS, RFI or RFE. They can be useful as secondary choices in specific scenarios, but they are not leading options in general. Based on the given F1 macro mean score analysis, the RFE feature selection technique was identified as one of the most reliable for both research questions.

TABLE 2 | Summary of maximum and average F1 scores for data preparation techniques for both questions.

The distance mathematics education is effective			I can adopt mathematical contents in online learning environment		
Technique	**Average**	**MAX**	**Technique**	**Average**	**MAX**
RFE	0.813	0.856	RFE	0.847	0.890
NFS	0.807	0.858	NFS	0.834	0.894
FFS	0.805	0.866	VT	0.832	0.889
BFS	0.801	0.853	RFI	0.823	0.893
VT	0.799	0.856	BFS	0.816	0.886
MAD	0.797	0.844	IG	0.816	0.885
CHI	0.793	0.859	CHI	0.806	0.883
RFI	0.793	0.858	FFS	0.805	0.874
IG	0.791	0.849	MAD	0.759	0.842

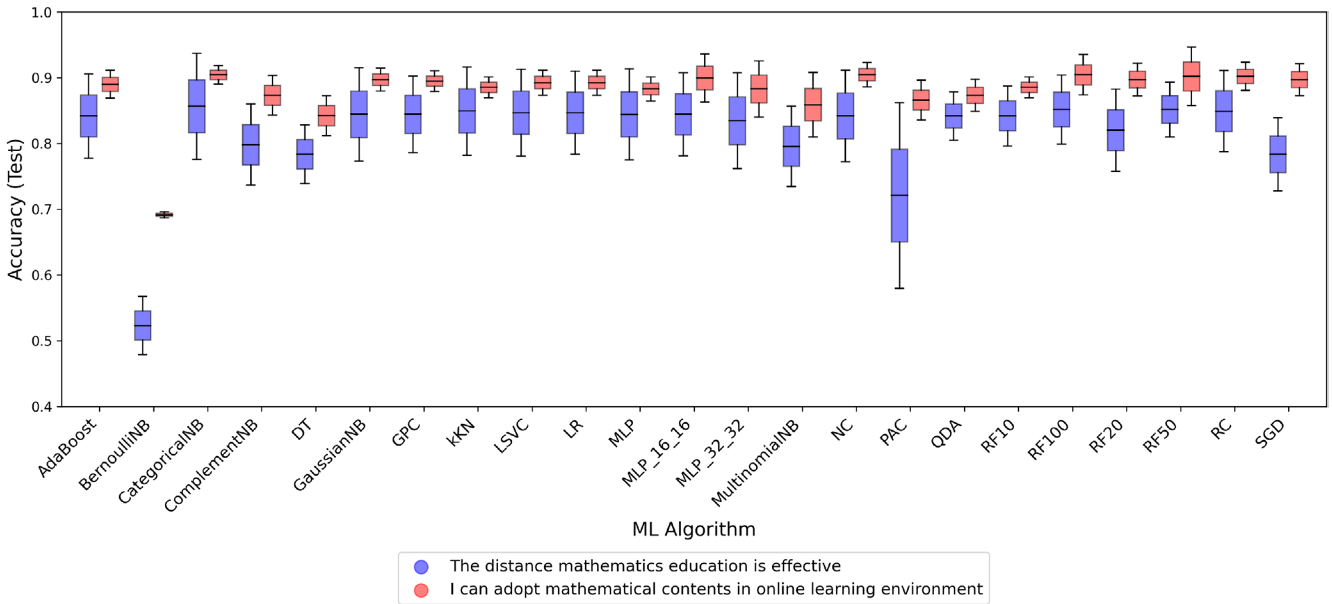

FIGURE 4 | Comparison of mean accuracy for ML algorithms using RFE across two questions.

3.2 | Comparison of Different ML Models

In this section, we present a detailed analysis of the performance of ML models after applying the RFE feature selection technique, which was identified as the most reliable for both research questions. Although the F1 macro mean score was previously used to guide the selection of data preparation techniques, this section extends the evaluation by examining additional metrics, including accuracy, precision, recall and execution time. By comparing the performance of different ML algorithms across these metrics, we aim to identify the most robust models and gain insights into how their performance changes depending on the nature of the research question. The mean accuracy analysis (Figure 4) highlights notable differences among algorithms shown through the box plot, especially when comparing their performance between the two research questions. For the first question (blue), CategoricalNB achieved the best results (0.856). High accuracy was also demonstrated by RF100 (0.852) and RF50 (0.851), while BernoulliNB (0.523) and PAC (0.721) were among the least effective models in this case. Large standard deviations, such as those for PAC (0.141), indicate unstable performance during different training attempts. For the second question (red), most algorithms showed improved mean accuracy. The best performing models were CategoricalNB (0.904), RF100 (0.904), NC (0.904) and RC (0.902). There is a notable improvement in accuracy compared to the first question for algorithms such as BernoulliNB (0.691) and PAC (0.866), indicating that the second question aligns better with their predictive strengths. Compared to the first question, standard deviations are generally smaller, indicating greater stability for algorithms such as RF100 and CategoricalNB. The analysis of macro mean precision (Figure 5) for both research questions reveals differences in algorithm performance and the stability of their results through standard deviations. For the first question (blue), the algorithms with the highest precision values are RF50 (0.857), RC (0.856) and CategoricalNB (0.860). These models demonstrated stable performance with moderate deviations. However, BernoulliNB (0.523) and PAC (0.789) again proved to be the least effective compared to the other algorithms, with significant instability in PAC, having a deviation of 0.079. When considering the second question (red), precision values improved significantly for most algorithms. The best results were achieved by RC (0.892), CategoricalNB (0.889), MLP_16_16 (0.887), RF100 (0.885) and NC (0.885), indicating that these algorithms better respond to the data of this question. BernoulliNB, however, performed poorly, with a precision of only 0.346, suggesting its insufficient ability to correctly classify the data for this question. Regarding the variability of the results, most models demonstrated stable results, with standard deviations below 0.04 for the second question. Exceptions include MLP_32_32 (0.052) and RF50 (0.051), showing slightly higher variability, indicating the potential sensitivity of these models to different training instances and data sets. The analysis of macro mean recall results reveals stable performance for most algorithms, with varying levels of variability between the two questions (Figure 6). For the first question (blue), the best recall results were achieved by CategoricalNB (0.857), RF100 (0.852), RF50 (0.852) and RC (0.849). This shows their ability to cover all relevant features in the data, regardless of class imbalance. Lower-performing results in this context were shown by BernoulliNB (0.522) and PAC (0.720), with the PAC model also affected by high result variability (standard deviation of 0.139). For the second question (red), significant improvements are visible for most algorithms. The best results were recorded by NC (0.897), RF100 (0.895) and RF50 (0.891), indicating their effectiveness in handling more complex and potentially

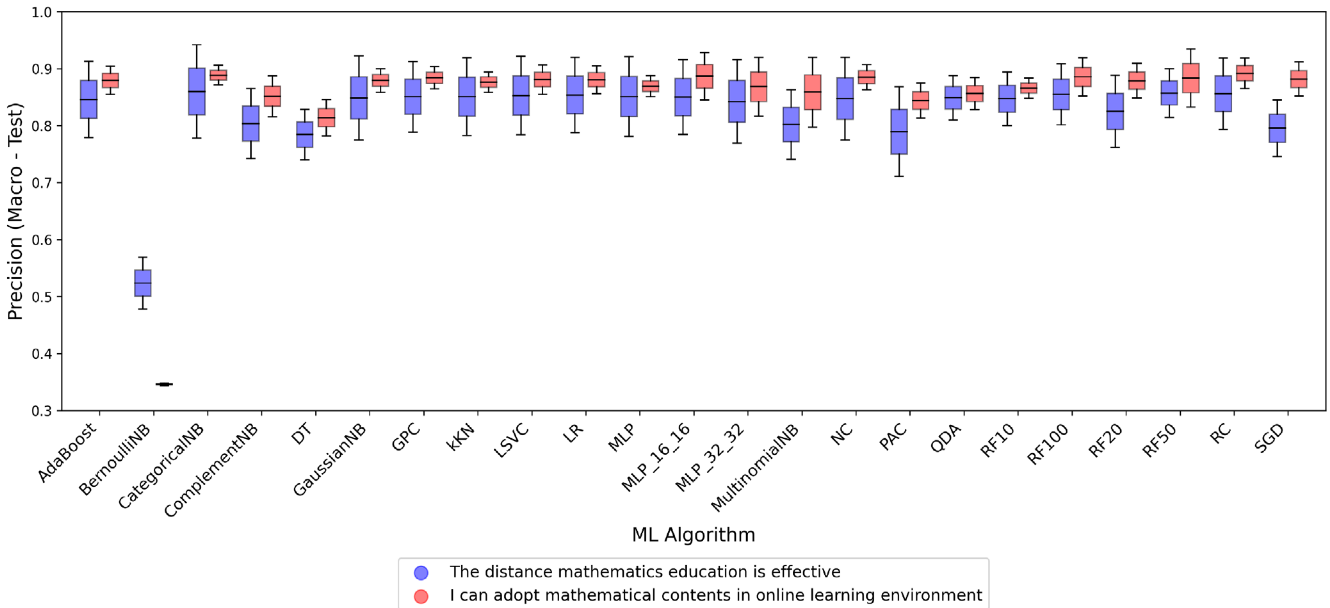

FIGURE 5 | Comparison of macro mean precision for ML algorithms using RFE across two questions.

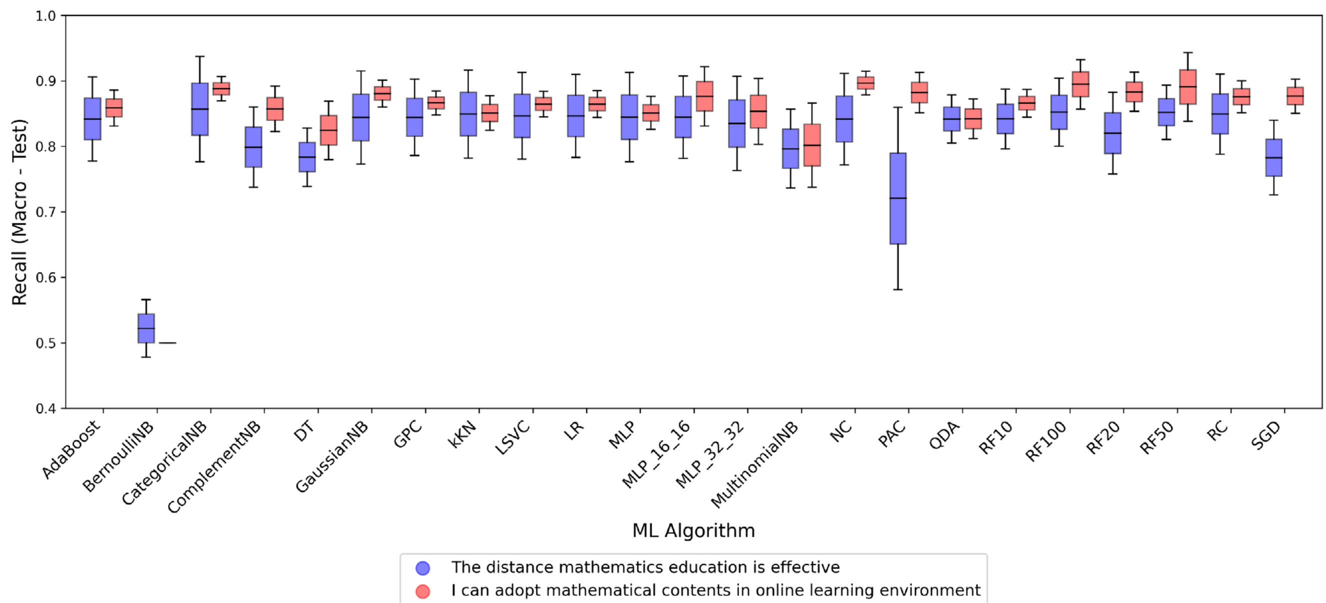

FIGURE 6 | Comparison of macro mean recall for ML algorithms using RFE across two questions.

heterogeneous data. On the other hand, BernoulliNB again exhibits poor performance, with a recall of only 0.500, making it unreliable in this context. The time metrics chart reveals significant differences in training and testing times between the algorithms (Figure 7). Algorithms such as GPC and MLP models stand out with long training times, indicating their high complexity and the need for greater computational resources during training. However, simpler models like CategoricalNB (0.00001) and BernoulliNB (0.00002) have very short training times. On the other hand, algorithms like NC and BernoulliNB draw particular attention because of the testing time, which is relatively high compared to the training time. This suggests that while these models are quick to train, their prediction processes may involve more complexity when applied to new data.

4 | Discussion

This study evaluated various data preparation techniques and machine learning algorithms to assess students' attitudes towards online mathematics education, based on responses to a structured survey found in Appendix A. The application of machine learning in education, as described in the literature review, has shown that certain algorithms such as Decision Trees, Logistic Regression and kNN achieve significant results in predicting student performance and analysing their attitudes (Ersozlu et al. 2024). Algorithms such as Random Forests and Naive Bayes have also proven effective in tasks requiring high accuracy and precision, as confirmed in this study (Durak and Bulut 2023; Milićević et al. 2024). Our results are consistent with these findings, as the RF100 and RC models demonstrate consistently high performance on various metrics. The analysis of the F1 macro mean metric (Figures 2 and 3) indicated that techniques such as RFE, NFS and RFI deliver the best results, confirming the importance of systematic feature selection in model optimization. These results align with previous research, where RFE was used to eliminate irrelevant features, resulting in higher accuracy and better model generalization (Marjuni et al. 2024; Chai et al. 2019).

Using the RFE method, we further identified the statements (Milenković et al. 2022) that have the greatest impact on the predictions of how students perceive the effectiveness of online mathematics instruction. The key factors contributing to perceived effectiveness include:

- the use of digital textbooks and problem collections during distance learning;
- clear instructions from teachers regarding assessment;
- students' agreement that distance mathematics learning has more advantages than in-class learning;
- students' agreement that distance mathematics learning has more advantages than disadvantages;
- whether students' academic performance declined during distance learning;
- students' interest in working in an online environment;
- students' agreement that lessons introducing new definitions and theorems were conducted as successfully as in-class lessons;
- students' age and gender.

In contrast, the results obtained using the RFE technique to assess students' perceptions of their ability to acquire mathematical content during distance learning suggest that the optimal set of statements consists of the following:

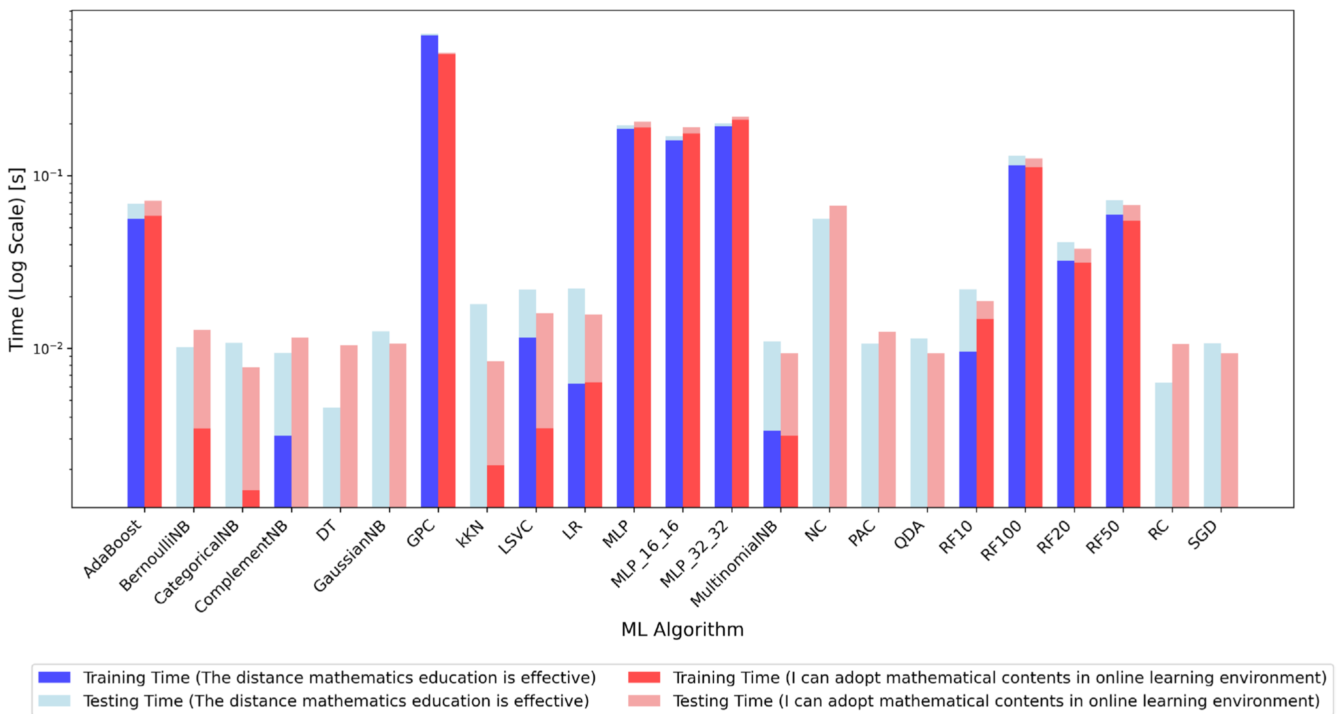

FIGURE 7 | Comparison of training and testing time for ML algorithms (Log Scale).

- students' belief that their mathematics teacher has sufficient knowledge and skills for the successful implementation of distance learning;
- that the mathematics teacher regularly provides feedback on students' work;
- students' level of knowledge and skills necessary to successfully follow mathematics distance learning;
- whether students' work habits and awareness of school obligations are maintained:
- whether students' performance in mathematics has worsened;
- the belief that mathematics content can be successfully taught during distance learning;
- students' agreement that lessons introducing new definitions and theorems were conducted via distance learning as successfully as in-class lessons;
- students' agreement that lessons in which they practice and connect previously learned concepts and theorems are conducted as successfully via distance learning as in-class lessons;
- students' age.

Taken together, the two sets of statements identified through the RFE based analysis reveal distinct but complementary pedagogical patterns. Students tend to perceive distance mathematics education as effective when instruction is supported by well-structured digital materials, transparent assessment criteria and online lessons that successfully replicate the explanatory and conceptual depth of in-class teaching. Motivation and the absence of perceived learning loss emerge as important indicators of successful adaptation to online instruction.

In contrast, students' ability to adopt mathematical content in an online environment is more strongly associated with teachers' pedagogical and technological competence, regular and meaningful feedback and the maintenance of students' work habits and learning routines. These findings suggest that perceived effectiveness and actual content adoption rely on different but interconnected conditions, emphasizing that successful distance mathematics education requires both high quality instructional design and sustained pedagogical support.

It is particularly interesting that techniques such as FFS and CHI showed exceptional performance on the first question, but their average values on the second question were slightly lower. This suggests that, although these techniques can achieve outstanding results in certain cases, their stability is not on the same level as techniques such as RFE and NFS, which consistently demonstrate good performance (Su et al. 2022). When it comes to machine learning algorithms, the performance results for the accuracy metric (Figure 4) reveal significant differences between the models. Algorithms such as CategoricalNB, RF100 and RC demonstrated stable performance with high average values and low standard deviations. In Sivasakthi and Padmanabhan (2022), Naive Bayes was compared with Decision Trees, showing superiority, which was also confirmed in this study. Although kNN showed satisfactory results to predict students' attitudes towards online mathematics education (Milenković, Krstić, and Svičević 2024), this research found that better results can be achieved with other algorithms. It is also important to note that BernoulliNB and PAC showed significantly lower performance, indicating their sensitivity to the specific characteristics of the data. The high performance of CategoricalNB, combined with its simplicity and speed, makes it a practical choice for tasks requiring fast data processing (Sivasakthi and Padmanabhan 2022). The results for precision and recall further confirmed the superiority of the CategoricalNB, RF50 and RC models, which stand out consistently in both metrics. These algorithms demonstrated the ability to classify the data accurately and cover all relevant features, which is crucial in an educational context where identifying all potentially problematic student cases is important. The high recall scores of the RF50 and RC models highlight their ability to recognize various scenarios, which is especially relevant in detecting learning difficulties in mathematics (Milićević et al. 2024; Elouafi et al. 2024). Regarding time metrics (Figure 7), significant differences were observed between the models. Algorithms such as GPC and MLP models exhibited longer training times, indicating their complexity. However, these algorithms also demonstrated a certain level of stability in their results, suggesting that their application is justified for more complex tasks (Alnasyan et al. 2024). However, simpler models such as CategoricalNB and ComplementNB showed fast training times with satisfactory performance, making them practical for scenarios where speed is a key factor.

From a practical teaching perspective, the items identified by RFE as the most relevant predictors for both research questions can be translated into concrete recommendations for distance mathematics instruction. First, digital instructional materials, particularly digital textbooks and problem collections, should be structured to be clear, segmented and consistent and where feasible, include elements that support active engagement during independent work. Second, online instruction should be designed to preserve conceptual depth through step-by-step teacher explanations, representative examples and tasks that emphasize understanding rather than mere reproduction. Third, assessment criteria and expectations should be made fully transparent (through rubrics, clearly defined learning outcomes, deadlines and scoring procedures), so that students have a stable framework for their work and for monitoring their own progress. Fourth, the introduction of a regular feedback cycle is recommended through brief and timely comments, since teacher guidance and feedback emerged as a key component of successful online learning. Fifth, teachers should explicitly support the development and maintenance of students' work habits and learning routines through lesson structure, work plans and reminders, given the relevance of routine maintenance in the selected predictors. Sixth, professional development for teachers in the context of distance education should include not only digital competencies but also strategies for sustaining student engagement, motivation and feedback in online environments. These recommendations translate the survey-based findings by linking the strongest predictors to actionable aspects of instructional design and pedagogical support, and they should be interpreted as associations with students' target responses rather than as causal determinants.

These findings reinforce the view that distance mathematics education is most effective when technological tools are combined with strong pedagogical guidance, rather than treated as a purely technical substitute for face-to-face instruction.

The innovative contribution of this study lies in the systematic integration of multiple feature selection techniques and machine learning models to analyse students' attitudes towards distance mathematics education. Unlike traditional studies that rely primarily on descriptive statistics or single model analyses, this approach enables the identification of stable and robust patterns across a large number of model configurations. By examining both performance metrics and feature relevance, the proposed methodology offers a data driven framework for extracting pedagogically meaningful insights from educational survey data.

This methodological design is particularly valuable in educational research contexts where complex interactions between pedagogical, technological and motivational factors are present. The generality of the pipeline allows it to be applied to other subject domains and educational settings, making it a flexible tool for evidence-based instructional planning and decision making.

5 | Conclusion

This study demonstrated that the proper selection of data preparation techniques and machine learning algorithms can significantly impact the accuracy and reliability of models for analysing students' attitudes in the context of two research questions: 'Is distance mathematics education effective?' and 'Can students successfully adopt mathematical content through online learning?'

The results showed that techniques such as RFE and NFS achieved the best performance for both research questions, with algorithms such as RF100, RC and CategoricalNB consistently demonstrating the highest values for F1 score, accuracy, precision and recall. These models produced stable and accurate results, particularly in identifying the most relevant statements within the student responses.

The set of statements identified through RFE as a stable predictor of students' agreement on the effectiveness of distance learning and their perceived ability to learn mathematical content online should serve as a foundation for instructional planning. These statements can guide the design and implementation of distance mathematics instruction to improve learning outcomes. These findings underscore specific factors educators should consider when designing online learning environments.

Moreover, techniques such as RFE and NFS, along with models such as RF100, can be applied to different types of educational domains, contexts and environments. This means that this approach is not limited to mathematics education. By applying this feature selection and model evaluation pipeline to domain-specific surveys in other fields, educators can gain actionable insights and better prepare instruction to the learning needs of students.

The study also highlights that smaller models, like CategoricalNB, are suitable when execution speed and large dataset handling are priorities. In contrast, more complex algorithms such as MLP and GPC, while requiring longer training times, are beneficial in scenarios requiring deeper analysis.

In conclusion, combining feature selection methods such as RFE with high-performing models like RF100 and RC is recommended to optimize the prediction and analysis of students' attitudes towards online mathematics education.

Author Contributions

Marina Svičević: conceptualization, methodology, visualization, formal analysis, writing original draft, writing – review and editing. **Aleksandar Milenković:** conceptualization, investigation, data curation, formal analysis, writing – review and editing. **Lazar Krstić:** software, data curation, visualization. **Miloš Pavković:** methodology, software, formal analysis, writing original draft, writing – review and editing.

Acknowledgements

The authors are supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia, Contract No. 451-03-34/2026-03/200122.

Funding

The authors have nothing report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

Adalar, F., and G. Oflaz. 2023. "A Scale Development Study Determining the Attitudes of Secondary School Students Towards Distance Education of Mathematics." *International e-Journal of Educational Studies* 7, no. 13: 153–169. https://doi.org/10.31458/iejes.1232169.

Albanyan, A. M. 2024. "The Quality of Distance Learning During Covid-19: Perspectives of Saudi University Students." *Heliyon* 10, no. 13: e33731. https://doi.org/10.1016/j.heliyon.2023.e00000.

Almarashdi, H., and A. M. Jarrah. 2021. "Mathematics Distance Learning Amid the Covid-19 Pandemic in the UAE: High School Students' Perspectives." *International Journal of Learning, Teaching and Educational Research* 20, no. 1: 292–307. https://doi.org/10.26803/ijlter.20.1.16.

Alnasyan, B., M. Basheri, and M. Alassafi. 2024. "The Power of Deep Learning Techniques for Predicting Student Performance in Virtual Learning Environments: A Systematic Literature Review." *Computers and Education: Artificial Intelligence* 6: 100231.

Awofala, A. O. A., S. O. Akinoso, C. O. Adeniyi, S. H. Jega, A. O. Fatade, and A. A. Arigbabu. 2024. "Primary Teachers' Mathematics Anxiety and Mathematics Teaching Anxiety as Predictors of Students' Performance in Mathematics." *ASEAN Journal of Science and Engineering Education* 4, no. 1: 9–24.

Bringula, R., J. J. Reguyal, D. D. Tan, and S. Ulfa. 2021. "Mathematics Self-Concept and Challenges of Learners in an Online Learning Environment During Covid-19 Pandemic." *Smart Learning Environments* 8, no. 1: 22. https://doi.org/10.1186/s40561-021-00168-5.

Cassibba, R., D. Ferrarello, M. F. Mammana, P. Musso, M. Pennisi, and E. Taranto. 2021. "Teaching Mathematics at Distance: A Challenge for Universities." *Educational Sciences* 11, no. 1: 1. https://doi.org/10.3390/educsci11010001.

Chai, Y., C. Lei, and C. Yin. 2019. "Study on the Influencing Factors of Online Learning Effect Based on Decision Tree and Recursive Feature Elimination." In *Proceedings of the 10th International Conference on E-Education, E-Business, E-Management and E-Learning*, 52–57. Association for Computing Machinery.

Chin, K. E., F. F. Jiew, and A. Jupri. 2022. "Mathematics Teachers Online Teaching Experience in Times of School Closures: The Case of Malaysia." *Malaysian Journal of Learning and Instruction* 19, no. 1: 59–84. https://doi.org/10.32890/mjli2022.19.3.

Cvijetić, M., M. Marić, and M. Beljanski. 2022. "Prednosti i nedostaci nastave na daljinu iz ugla nastavnika i roditelja." *Inovacije u Nastavi* XXXV, no. 2: 1–15. https://doi.org/10.5937/inovacije2202001C.

de la Peña Esteban, F. D., J. A. Lara Torralbo, D. Lizcano Casas, and M. A. Martínez Rey. 2019. "Expert System for Problem Solving in Distance University Education: The Successful Case of the Subject Operations Management." *Expert Systems* 36, no. 5: e12444.

Descamps, S. X., H. Bass, G. B. Evia, R. Seiler, and M. Seppälä. 2006. "E-Learning Mathematics." In *Proceedings of the International Congress of Mathematicians (ICM)*, 1743–1768. EMS Press.

Djordić, D. M., M. M. Cvijetić, and R. D. Damjanović. 2021. "Iskustva učitelja i nastavnika tokom realizacije nastave na daljinu usled pandemije virusa korona (covid-19)." *Inovacije u Nastavi—Časopis Za Savremenu Nastavu* 34, no. 2: 86–103. https://doi.org/10.5937/inovacije2102086D.

Durak, A., and V. Bulut. 2023. "Classification and Prediction-Based Machine Learning Algorithms to Predict Students' Low and High Programming Performance." *Computer Applications in Engineering Education* 32, no. 1: e22679. https://doi.org/10.1002/cae.22679.

Elouafi, A., I. Tammouch, S. Eddarouich, and R. Touahni. 2024. "Evaluating Various Machine Learning Methods for Predicting Students Math Performance in the 2019 Timss." *Indonesian Journal of Electrical Engineering and Computer Science* 34, no. 1: 565–574.

Engelbrecht, J., and A. Harding. 2005. "Teaching Undergraduate Mathematics on the Internet." *Educational Studies in Mathematics* 58, no. 2: 253–276.

Ersozlu, Z., S. Taheri, and I. Koch. 2024. "A Review of Machine Learning Methods Used for Educational Data." *Education and Information Technologies* 29: 22125–22145.

Goria, C., and A. Konstantinidis. 2023. "A Participatory Pedagogical Model for Online Distance Learning: Ideation and Implementation." *Turkish Online Journal of Distance Education* 24, no. 1: 145–161.

Gurajena, C., E. Mbunge, and S. Fashoto. 2021. "Teaching and Learning in the New Normal: Opportunities and Challenges of Distance Learning Amid COVID-19 Pandemic." Available at SSRN 3765509.

Kansiime, G., and M. S. K. Batiibwe. 2023. "Adoption of Online Mathematics Learning in Ugandan Government Universities During the COVID-19 Pandemic: Pre-Service Teachers' Behavioural Intention and Challenges." *Discover Education* 2, no. 1: 12.

Koklu, N. 2025. "A Multidimensional Analysis of the 21st Century Competencies Scale Through AI-Driven Data Mining Techniques." *Scientific Reports* 15, no. 1: 43618.

Kramer, O. 2016. *Machine Learning for Evolution Strategies*. Vol. 20. Springer.

Lazović, T. S. 2021. "Učenici sa smetnjama u razvoju na nastavi matematike tokom pandemije virusa covid-19." *Pedagoška Stvarnost* 67, no. 2: 174–185.

Lowe, T., B. Mestel, and G. Williams. 2016. "Perceptions of Online Tutorials for Distance Learning in Mathematics and Computing." *Research in Learning Technology* 24: 1–14. https://doi.org/10.3402/rlt.v24.30630.

Lučev, I., L. Lukić Jakopčević, and V. Špiljak. 2022. "Experiences of Foreign-Language Teachers at the Tertiary Level in the Online Teaching Environment in Croatia in 2020." *Strani Jezici: Časopis Za Primijenjenu Lingvistiku* 51, no. 1: 23–54.

Mahmudi, A. A., R. Fionasari, B. Mardikawati, and L. Judijanto. 2023. "Integration of Artificial Intelligence Technology in Distance Learning in Higher Education." *Journal of Social Science Utilizing Technology* 1, no. 4: 190–201.

Marjuni, A., A. Z. Fanani, and R. S. Basuki. 2024. "Enhancing Machine Learning Algorithms Using Recursive Feature Elimination For Student Performance Prediction." In *In 2024 International Seminar on Application for Technology of Information and Communication (Isemantic)*, 409–414. IEEE.

Matić, I., and D. Stančić. 2021. "Što smo naučili u/o nastavi na daljinu?" *Sociologija i Prostor* 59, no. 3: 413–435.

Mihajlović, A., V. Ristić, and E. Kopas-Vukašinović. 2024. "University Students Attitudes About Online Mathematics Teaching During the Pandemic." *Zbornik Radova Pedagoskog Fakulteta, Uzice* 27, no. 26: 119–138. https://doi.org/10.5937/ZRPFU2426119M.

Mihajlović, A., N. Vulović, and S. Maričić. 2021. "Teaching Mathematics During the Covid-19 Pandemic—Examining the Perceptions of Class Teachers and Mathematics Teachers." In *Nauka, nastava, učenje u izmenjenom društvenom kontekstu*, edited by S. Marinković and J. Stamatović, 501–518. Pedagoški fakultet u Užicu. https://doi.org/10.46793/NNU21.501M.

Milenković, A., S. Aleksić, and A. Saković. 2022. "Differences in the Attitudes of Mathematics Teachers and of Students Towards Online Mathematics Instruction." *Nastava i Vaspitanje* 71, no. 3: 363–383. https://doi.org/10.5937/nasvas2203363M.

Milenković, A., L. Krstić, and M. Svičević. 2024. "Assessment of Students' Attitudes Towards Distance Mathematics Learning Using k-Nearest Neighbour Classifier." In *28th scientific conference Elementary Mathematics Education Banská Bystrica - Osrblie,* April 24-26, 2024. Univerzita Mateja Bela, Fakulta prírodných vied v Banskej Bystrici.

Milenković, A., M. Milikić, and T. Jovović. 2024. "Embracing Technology in Distance Mathematics Education: Investigating Students' Perceptions of the Teachers' Usage of the Pen Tablet." *Research in Pedagogy* 14, no. 1: 28–45.

Milićević, M., B. Marinović, and L. Jeftić. 2024. "Machine Learning Methods as Auxiliary Tool for Effective Mathematics Teaching." *Computer Applications in Engineering Education* 32, no. 6: e22787.

Mohammed, M., M. B. Khan, and E. B. M. Bashier. 2016. *Machine Learning: Algorithms and Applications*. CRC Press.

Mohtar, M., and M. Md Yunus. 2022. "A Systematic Review of Online Learning During Covid 19: Students Motivation, Task Engagement and Acceptance." *Arab World English Journal* 19: 202–215.

Mukuka, A., O. Shumba, and H. M. Mulenga. 2021. "Students Experiences With Remote Learning During the Covid-19 School Closure: Implications for Mathematics Education." *Heliyon* 7, no. 7: e07523.

Murtafiah, W., S. Suwarno, and N. D. S. Lestari. 2020. "Exploring the Types of a Material Presentation by Teachers in Mathematics Learning During the Covid-19 Pandemic." *Journal of Physics: Conference Series* 1663, no. 1: 012043. https://doi.org/10.1088/1742-6596/1663/1/012043.

Ranđelović, B. M., E. Karalić, and D. Đukić. 2020. "The Digitalization of the Learning Process in Serbia During the COVID-19 Crisis." In *Proceedings of the Scientific Conference "—"*, 203–216. Pedagoški fakultet u Užicu. https://doi.org/10.46793/STEC20.203R.

Sarker, I. H. 2021. "Machine Learning: Algorithms, Real-World Applications and Research Directions." *SN Computer Science* 2, no. 3: 160.

Sherry, L. 1995. "Issues in Distance Learning." *International Journal of Educational Telecommunications* 1, no. 4: 337–365.

Shevchenko, V., N. Malysh, and O. Tkachuk-Miroshnychenko. 2024. "Distance Learning in Ukraine in COVID-19 Emergency." *Open Learning: The Journal of Open, Distance and e-Learning* 36, no. 1: 4–19. https://doi.org/10.1080/02680513.2021.1967115.

Shou, Z., M. Xie, J. Mo, and H. Zhang. 2024. "Predicting Student Performance in Online Learning: A Multidimensional Time-Series Data Analysis Approach." *Applied Sciences* 14, no. 6: 2522. https://doi.org/10.3390/app14062522.

Sivasakthi, M., and K. R. A. Padmanabhan. 2022. "Prediction of Students Programming Performance Using naïve Bayesian and Decision Tree." In *Soft Computing for Security Applications*, vol. 1428, 97–106. Springer.

Steyn, A. A., C. van Slyke, G. Dick, H. Twinomurinzi, and L. B. Amusa. 2024. "Student Intentions to Continue With Distance Learning Post-Covid: An Empirical Analysis." *PLoS One* 19, no. 1: e0293065.

Su, Y.-S., Y.-D. Lin, and T.-Q. Liu. 2022. "Applying Machine Learning Technologies to Explore Students Learning Features and Performance Prediction." *Frontiers in Neuroscience* 16: 1018005.

Traxler, J. 2018. "Distance Learning—Predictions and Possibilities." *Educational Sciences* 8, no. 1: 35. https://doi.org/10.3390/educsci8010035.

Vučetić, I., N. Vasojević, and S. Kirin. 2020. "Mišljenje učenika srednjih škola u srbiji o predvidjanjima onlajn učenja tokom pandemije covid-19." *Nastava i Vaspitanje* 69, no. 3: 345–359. https://doi.org/10.5937/nasvas2003345V.

Wang, Q., and Q. Huang. 2023. "Engaging Online Learners in Blended Synchronous Learning: A Systematic Literature Review." *IEEE Transactions on Learning Technologies* 17: 594–607.

Woelk, M., M. Nam, B. Häckel, and M. Spörrle. 2025. "From Questionnaires to Heatmaps: Visual Classification and Interpretation of Quantitative Response Data Using Convolutional Neural Networks." *Applied Sciences* 15, no. 19: 10642.

Appendix A

A Survey

Gender

a. Female

b. Male

Choose your current grade level: ________.

Type of school

a. Primary school

b. High school

For the next statements mark a number that aligns with your opinion (1—I strongly disagree; 2—I somewhat disagree; 3—I am undecided; 4—I somewhat agree; 5—I strongly agree).

1. During distance mathematics learning, we use digital textbooks and problem collections.
2. During distance mathematics learning, we use dynamic software (GeoGebra, MathLab, Wolfram Mathematica…) with instructions and support from the teacher.
3. During distance mathematics learning, we use presentations, images, scanned materials and various other documents.
4. During distance mathematics learning, we use applications and platforms for solving tests and quizzes (quizziz, kahoot…).
5. During distance mathematics learning, we largely follow mathematics classes broadcasted on RTS (public TV service), as instructed by the teacher.
6. I believe that the mathematics teacher who teaches me has enough knowledge and skills for successful implementation of distance learning.
7. During distance learning, the mathematics teacher regularly provides feedback on my work.
8. During distance learning, I received clear instructions regarding assessment from the mathematics teacher.
9. During distance learning, the mathematics teacher regularly informs my parents about my work.
10. I believe I have enough knowledge and skills necessary to successfully follow mathematics distance learning.
11. I have experience with distance mathematics learning, even before the SARS-CoV-2 pandemic.
12. **I believe that distance mathematics learning is effective**.
13. I believe that distance mathematics learning has more advantages than disadvantages.
14. I believe that distance mathematics learning has more advantages compared to mathematics learning conducted in a traditional school environment.
15. I am interested in this type of work and collaboration with the mathematics teacher.
16. I regularly follow mathematics lessons, to the extent that corresponds to following lessons in a traditional school environment.
17. I experience technical difficulties while following distance learning.
18. I complete math homework independently during distance learning.
19. I regularly complete math homework during distance learning.
20. During distance learning, I retained my work habits and awareness of school obligations.
21. During distance learning, my performance in mathematics worsened.
22. During distance learning, my overall school performance worsened.
23. I believe that communication between the mathematics teacher, my peers and me is at a satisfactory level.
24. Mathematics content is such that, for the most part, it can be successfully taught in mathematics classes during distance learning.
25. **I believe that mathematics content is such that I can successfully adopt and understand it through distance learning**.
26. Mathematics lessons where we are introduced to new concepts, theorems and content are successfully conducted via distance learning, without significant differences compared to lessons held in a traditional school environment.
27. Mathematics lessons where we review, practice and connect concepts and theorems we previously learned are successfully conducted via distance learning, without significant differences compared to lessons held in a traditional school environment.
28. Mathematics tests (written assignments, control exercises) are successfully conducted via distance learning.