An Adaptive Feature Selection Algorithm for Student Performance Prediction

By

Koushik Roy, 0122230054

Submitted in partial fulfilment of the requirements
of the degree of Master of Science in Computer Science and Engineering

July 14, 2024

Department of Computer Science and Engineering
United International University

Declaration

I, **Koushik Roy**, declare that this thesis titled, **An Adaptive Feature Selection Algorithm for Student Performance Prediction** and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a MSc degree at United International University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at United International University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

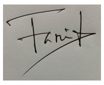
Signed: Koushik

Date: July 14, 2024

(koushik Roy)

Certificate

I do hereby declare that the research works embodied in this thesis titled, **An Adaptive Feature Selection Algorithm for Student Performance Prediction** is the outcome of an original work carried out by **Koushik Roy** under my supervision.
I further certify that the dissertation meets the requirements and the standard for the degree of MSc in Computer Science and Engineering.

Signed: 

Date: July 14, 2024

(Prof. Dr. Dewan Md. Farid)
Department of Computer Science and Engineering,
United International University,
Dhaka-1209, Bangladesh.

Abstract

Educational Data Mining (EDM) is used to ameliorate the teaching and learning process by analyzing and classifying data that can be applied to predict the students' academic performance, and students' dropout rate, as well as instructors' performance. The prediction of student performance is complicated by the vast and diverse range of variables from academic records to behavioral and health metrics. In this thesis book, we have introduced a new Adaptive Feature Selection Algorithm (AFSA) by amalgamating an ensemble approach for initial feature ranking with normalized mean ranking from five distinct methods to enhance robustness. The proposed method iteratively selects the best features by adjusting its threshold based on each feature's rank to ensure significant contributions to model accuracy and also effectively reduces dataset complexity. We have tested the performance of the proposed feature selection algorithm using five machine learning classifiers: Logistic Regression (LR), K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Naïve Bayes (NB) classifier, and Decision Tree (DT) classifier on four student performance datasets. The experimental results highlight the proposed method significantly decreases feature count by an average feature reduction factor of 5.7, significantly streamlining datasets while maintaining competitive cross-validation accuracy, marking it as a valuable tool in the field of educational data analytics.

Acknowledgements

This work would not have been possible without the input and support of many individuals throughout my study period. I would like to express my deepest gratitude to everyone who contributed to it in one way or another.

First and foremost, I extend my sincerest thanks to my thesis supervisor, Prof. Dr. Dewan Md. Farid, whose guidance and insights were invaluable throughout this journey.

I am profoundly grateful to Prof. Dr. Md. Motaharul Islam and Prof. Dr. Mohammad Nurul Huda for their unwavering support and encouragement.

I owe a great deal of gratitude to my thesis examiners, Prof. Dr. Swakkhar Shatabda, Prof. Dr. Raqibul Mostafa, and Prof. Dr. Md. Saddam Hossain Mukta, for their rigorous reviews and valuable feedback.

In addition, I would like to acknowledge my course professors, Prof. Dr. Swakkhar Shatabda, Prof. Dr. Dewan Md. Farid, Prof. Dr. Md. Saddam Hossain Mukta, and Prof. Dr. Khondaker Abdullah -Al-Mamun, who have significantly shaped my understanding and knowledge. Their teachings have been instrumental in my academic development.

Last but not least, I owe everything to my family, including my parents, for their unconditional love and immense emotional support throughout this journey.

Publication List

The main contributions of this research are either published or accepted or in preparation in journals and conferences as mentioned in the following list:

Journal Articles

1. Koushik Roy, Huu-Hoa Nguyen, and Dewan Md Farid. Impact of dimensionality reduction techniques on student performance prediction using machine learning. CTU Journal of Innovation and Sustainable Development, 15:93–101, October 2023

2. Koushik Roy and Dewan Md Farid. An adaptive feature selection algorithm for student performance prediction. IEEE Access, 12:75577 – 75598, May 2024

Conference Papers

1. Koushik Roy, Huu-Hoa Nguyen, and Dewan Md. Farid. Impact of dimensionality reduc- tion techniques on student performance prediction using machine learning. In International Conference on Intelligent Systems and Data Science (ISDS-2023), pages 1–15, Can Tho University (CTU), Vietnam, November 2023

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Educational Data Mining (EDM) has emerged as a pivotal area of research, representing the confluence of data mining, machine learning, and statistical analysis techniques with the aim of enriching and innovating the educational process. This interdisciplinary field harnesses the potential of vast datasets generated within educational contexts to unlock insights into student learning behaviors, engagement levels, and academic performance. The evolution of EDM reflects a broader trend towards data-driven decision-making and personalized education, prompted by the digital transformation of educational spaces and the advent of Big Data technologies.

1.1.1 The Digital Transformation of Education

The last few decades have witnessed a seismic shift in how educational content is delivered, accessed, and assessed, largely due to the digitization of educational environments. Digital platforms such as learning management systems (LMS), student information systems (SIS), and massive open online courses (MOOCs) have become integral to the educational experience, creating a rich tapestry of data on every aspect of the student learning journey. These platforms facilitate a continuous flow of information, capturing myriad data points from student engagement metrics and learning progress to social interactions and resource utilization. This digital transformation has expanded the horizons of educational research, offering new avenues to explore how students learn and how educational outcomes can be optimized.

1.1.2 Big Data's Role in Educational Innovation

The concept of Big Data refers to datasets that are vast, complex, and rapidly evolving, characteristics that aptly describe the data landscape in education. The sheer volume of data generated through educational technologies presents both challenges and opportunities for educators and researchers. Big Data analytics, powered by advanced data mining and machine learning algorithms, allows for the extraction of meaningful patterns and insights from this data deluge. These insights have the potential to revolutionize educational practices by enabling personalized learning paths, predictive interventions, and real-time feedback mechanisms, thereby catering to the unique needs and potentials of each student.

1.1.3 The Promise of Personalized Education

One of the most compelling aspects of EDM is its ability to underpin personalized education. By analyzing detailed data on individual learning styles, preferences, and performance, educational practitioners can tailor teaching approaches, content, and support to meet each student's specific needs. This bespoke approach to education has the potential to significantly improve learning outcomes by addressing students' strengths and weaknesses more effectively than traditional one-size-fits-all teaching methods.

1.1.4 Challenges and Opportunities

Despite the promising potential of EDM, its application is not without challenges. Issues related to data privacy, ethical considerations in data handling, and the need for robust data governance frameworks are paramount. Moreover, the complexity of educational data, which includes not only quantitative but also qualitative and unstructured data, necessitates sophisticated analytical techniques and tools. However, these challenges also present opportunities for innovation in data analysis methods, ethical standards, and policy frameworks that can guide the responsible use of educational data.

1.1.5 Research Efforts and Future Directions

As noted in the work of [1, 2, 3], research in EDM has predominantly focused on leveraging data insights to predict and enhance student performance. These efforts have laid a solid foundation for the field, highlighting the transformative potential of data-driven educational strategies. Moving forward, the scope of EDM research is set to broaden, encompassing areas such as emotional and cognitive aspects of learning, the dynamics of teacher-student interactions, and the impact of environmental factors on educational outcomes. The future of EDM lies in its ability to offer holistic insights that can guide not only academic but also socio-emotional interventions, contributing to the overall well-being and success of students.

Educational Data Mining stands at the forefront of educational innovation, offering promising avenues to enhance learning experiences and outcomes. The continued evolution of this field, fueled by advancements in technology and analytical methods, promises to usher in a new era of personalized, data-driven education.

1.2 Challenges in Educational Data Analysis

Analyzing educational Big Data involves navigating a labyrinth of complexities, each presenting unique challenges to researchers and educators alike. The data generated within educational settings is inherently multi-dimensional, capturing a wide spectrum of information that ranges from academic performance and learning behaviors to demographic backgrounds and psychosocial metrics. This diversity in data types and sources necessitates a nuanced approach to data analysis, where the following challenges are most pronounced:

1.2.1 Complexity of Multi-Dimensional Datasets

Educational datasets are characterized by their complexity and multi-dimensionality. Academic records, for instance, provide quantitative measures of student achievement, while demographic information offers contextual backgrounds of the learner population. Behavioral patterns, gleaned from interaction data with learning management systems, and health metrics introduce additional dimensions that can influence educational outcomes. The integration and analysis of these disparate data types require sophisticated models that can accommodate the heterogeneity of educational data.

1.2.2 Computational Demands

The volume of data generated in educational settings is staggering, growing exponentially with the adoption of digital learning tools and platforms. This creates significant computational challenges, as traditional data processing techniques may not be scalable or efficient enough to handle such vast datasets. Researchers and practitioners must leverage advanced computational methods and technologies, such as cloud computing and parallel processing, to manage and analyze educational Big Data effectively.

1.2.3 Model Interpretability and Overfitting

The goal of predictive analytics in education is not just to predict outcomes accurately but also to derive insights that can inform teaching practices and learning interventions. This requires models that are not only accurate but also interpretable, meaning that educators and researchers can understand the factors driving the predictions. However, the complexity of educational data and the sophisticated models used to analyze it often lead to challenges in interpretability. Furthermore, the risk of overfitting, where models perform well on training data but poorly on unseen data, is heightened due to the vastness and variety of features available in educational datasets. Careful feature selection and model validation practices are crucial to mitigate these issues.

1.2.4 Ethical Considerations and Data Privacy

As educational data mining involves the analysis of sensitive information, ethical considerations and data privacy concerns are paramount. The handling of student data must adhere to strict ethical guidelines and privacy regulations, such as the General Data Protection Regulation (GDPR) in Europe and the Family Educational Rights and Privacy Act (FERPA) in the United States. Ensuring the anonymity and security of student data while extracting meaningful insights poses a delicate balance that must be maintained throughout the research process.

1.2.5 Addressing these Challenges

Previous studies [4, 5] underscore the critical need for developing advanced analytical tools and methodologies capable of overcoming these challenges. The development of sophisticated tools for feature prioritization and analysis is crucial for managing the complexity of educational datasets. These tools must not only be powerful in terms of computational capability but also sensitive to the ethical and privacy concerns associated with handling educational data.

The path forward involves a collaborative effort among educators, researchers, and technologists to innovate and refine data analysis techniques in education. By embracing interdisciplinary approaches and advancing ethical data governance frameworks, the educational research community can navigate the challenges of Big Data analysis, unlocking the full potential of educational data mining to enhance learning outcomes and educational strategies.

1.3 Objective and Contributions

This study is positioned at the forefront of educational data mining, introducing the Adaptive Feature Selection Algorithm (AFSA) as a novel solution to the multifaceted challenges associated with analyzing educational Big Data. The development and application of AFSA are guided by two primary objectives, which collectively aim to enhance the predictive analytics landscape in education. Additionally, this study makes several significant contributions to the field, advancing our understanding and capabilities in educational data analysis.

1.3.1 Objectives of the Study

Development of AFSA

The primary objective of this research is to develop an advanced feature selection algorithm that leverages the strengths of various existing methodologies to intelligently identify and prioritize the most critical features within educational datasets. This objective is driven by the recognition that the effectiveness of predictive analytics is contingent upon the ability to distill complex datasets into their most informative components. By doing so, AFSA aims to reduce computational complexity, mitigate the risk of overfitting, and enhance the interpretability of predictive models. This development represents an extension and significant advancement of our prior work [6, 7], pushing the boundaries of what is currently achievable in educational data mining.

Evaluation of AFSA's Impact

The second objective focuses on the rigorous evaluation of AFSA's impact on key performance metrics, such as accuracy, precision, recall, f1 score, feature reduction factor, and training time. Special emphasis is placed on the algorithm's efficacy in predicting student performance, a critical area of interest within educational data mining. This evaluation will not only assess the technical merits of AFSA but also its practical implications, shedding light on the algorithm's potential to facilitate more targeted and effective educational interventions.

1.3.2 Contributions to the Field

Advancement in Feature Selection Methodologies

AFSA represents a significant leap forward in feature selection techniques, particularly within the context of educational data mining. By synthesizing the strengths of multiple feature ranking methodologies, AFSA introduces a more adaptive and effective approach to feature selection. This contribution is expected to set a new benchmark for the development of predictive models in education, offering a powerful tool for researchers and practitioners alike.

Enhanced Predictive Analytics in Education

Through the development and application of AFSA, this study contributes to the enhancement of predictive analytics capabilities in the educational sector. By improving the accuracy and interpretability of predictive models, AFSA facilitates a deeper understanding of student performance and learning dynamics. This, in turn, enables educational institutions to make more informed decisions, tailor learning experiences to individual needs, and allocate resources more efficiently.

Foundation for Future Research

Beyond its immediate contributions, this study lays the groundwork for future research in educational data mining and feature selection. By demonstrating the efficacy of AFSA and its impact on predictive analytics, this research opens new avenues for exploration and innovation. It encourages further investigation into adaptive feature selection methodologies and their application across different domains within education.

This study, through its objectives and contributions, underscores the transformative potential of the Adaptive Feature Selection Algorithm in the realm of educational data mining. It not only advances our methodological capabilities but also enriches our understanding of educational processes, ultimately contributing to the development of more effective and personalized educational strategies.

1.4 Structure of the Thesis

The structure of this thesis is designed to provide a comprehensive exploration of the Adaptive Feature Selection Algorithm and its application in the field of educational data mining. Following this introduction, Chapter 2 presents a thorough review of the literature, covering key developments in EDM and the role of feature selection techniques in analyzing educational data. Chapter 3 offers an overview of existing feature selection techniques, setting the stage for the introduction of AFSA. Chapter 4 delves into the methodology and theoretical framework of AFSA, comparing its approach with other feature selection algorithms. Chapter 5 discusses the experimental setup, presents the results, and evaluates the performance of AFSA in the context of student performance prediction. Finally, Chapter 6 summarizes the findings of this research, acknowledges its limitations, and outlines potential directions for future studies in educational data mining and feature selection.

1.5 Significance of the Study

The significance of this study extends far beyond the technical development of a new feature selection algorithm; it represents a paradigm shift in the approach to educational data mining and its application in the educational sector. The introduction of the Adaptive Feature Selection Algorithm (AFSA) marks a crucial step forward in harnessing the power of Big Data to enhance educational outcomes and strategies. Below, we outline the multifaceted implications of this research and its contributions to the field.

1.5.1 Advancements in Predictive Analytics

By enhancing the efficiency and efficacy of feature selection in educational datasets, AFSA significantly improves the predictive analytics capabilities within the educational sector. This advancement allows for the development of more accurate and interpretable predictive models, which are instrumental in understanding and forecasting student performance. The improved accuracy and clarity of these models enable educators and administrators to make informed decisions based on robust data-driven insights.

1.5.2 Personalization of Learning

One of the most transformative aspects of this study is its potential to facilitate personalized learning. The insights derived from more refined predictive models can be used to tailor educational content, methodologies, and support services to meet the unique needs of each student. This personalization of the learning experience is expected to foster a more engaging and effective educational environment, ultimately leading to improved learning outcomes and student satisfaction.

1.5.3 Resource Optimization

The application of AFSA also contributes to the more efficient allocation of educational resources. By identifying the most salient factors influencing student performance, educational institutions can target their interventions and support services more precisely, optimizing the use of limited resources. This targeted approach not only enhances the effectiveness of educational programs but also ensures that resources are directed where they are most needed.

1.5.4 Contribution to Educational Data Mining

This research contributes significantly to the field of educational data mining by introducing an innovative approach to feature selection. The Adaptive Feature Selection Algorithm represents a leap forward in the analysis of complex educational datasets, offering a more nuanced and effective tool for researchers and practitioners. This contribution is expected to inspire further research and innovation in the field, leading to the development of even more sophisticated analytical tools and techniques.

1.5.5 Implications for Policy and Practice

Beyond its technical contributions, the study has important implications for educational policy and practice. The insights gained through the application of AFSA can inform policy decisions, guide the development of educational strategies, and shape the future of educational practices. By providing a clearer understanding of the factors that influence student performance, this research empowers policymakers and educators to enact changes that are grounded in empirical evidence.

In sum, the significance of this study lies in its potential to revolutionize the way we approach educational data mining, from enhancing predictive analytics and personalizing learning experiences to optimizing resource allocation and informing policy. Through the development and application of the Adaptive Feature Selection Algorithm, this work demonstrates the transformative power of Big Data in education, offering valuable insights and tools to improve educational strategies and outcomes.

Table 1.1: Summary of Research Studies of Educational Data Mining.

Year	**Method**	**Dataset**	**Objective**	**Reference**
2020	Naive Bayes, J48	Educational Benchmark Dataset	Data Analysis	Karthikeyan et al. [8]
2020	Association Rule Mining, Principal Component Analysis	Real Academic Data	Data Analysis	Crivei et al. [9]
2020	Genetic Algorithm	Kaggle Repository Data	Prediction	Farissi et al. [10]
2021	Text Mining, k-Nearest Neighbors	Student Teaching Evaluation Data	Data Analysis	Okoye et al. [11]
2021	Naive Bayes, Random Forest	Benchmark Student Data	Data Analysis	Kumar et al. [12]
2021	Harris Hawks Optimization	UCI Dataset	Prediction	Turabieh et al. [13]
2021	Deep Neural Network	Public 4-Year University Data	Prediction	Nabil et al. [14]
2022	Improved Evolutionary Algorithm with Neuro-Fuzzy Classification, Chaotic Whale Optimization	Benchmark Student Performance Data	Data Analysis	Duhayyim et al. [15]
2022	Deep Cognitive Diagnosis Framework, Improved K-means, CNN	Real-World Data	Prediction	Gao et al. [16]
2022	Improved K-means, CNN	University Dataset	Data Analysis and Prediction	Feng et al. [17]

Table 1.2: Summary of Feature Selection Methods in EDM.

Method	**Type**	**Study**
Manual selection		[18], [19], [20], [21], [22], [23]
Filter	Correlation	[24], [25], [26], [27], [28], [29], [30]
	Information Gain	[31], [32]
	Classifier	[33], [34]
Wrapper	Genetic algorithm	[35], [36], [37]
	PCA	[38]

Chapter 2

Literature Review

Educational Data Mining (EDM) has emerged as a transformative approach to analyzing educational data, aiming to improve learning outcomes and enhance educational processes. This chapter provides an extensive review of the literature on EDM, focusing on its foundations, methodologies, the application of feature selection techniques, and the integration of these techniques with predictive modeling.

In recent years, the field of education has experienced a surge in interest regarding the utilization of data mining techniques, particularly within the framework of Educational Data Mining (EDM). EDM represents the convergence of data mining technology and educational data, with the primary goal of gaining deeper insights into students' learning processes and enhancing the overall educational experience [39, 40]. Educational practitioners and researchers have consistently emphasized the pressing need for meaningful EDM tools that guide decision-making and provide valuable insights [39, 40]. One of the notable areas of research within EDM is the application of data mining algorithms for the analysis and prediction of academic performance. This research landscape has been summarised in Table 1.1, highlighting the diverse methods and datasets used by researchers to achieve various objectives.

2.1 The Emergence of Educational Data Mining

The field of education has witnessed a remarkable transformation with the advent of data mining techniques, placing Educational Data Mining (EDM) at the heart of this evolution. The inception of EDM marked a pivotal shift in educational research and practice, leveraging data mining technology to delve into vast amounts of educational data. This synergy aims to extract profound insights into students' learning processes, thereby significantly enhancing educational experiences [39, 40].

The genesis of EDM can be traced back to the late 1990s and early 2000s, a period characterized by the digital revolution in education. As educational institutions increasingly adopted digital tools and platforms, the volume of data generated by students' interactions with these technologies began to grow exponentially. This burgeoning data presented an untapped resource for understanding and improving education, catalyzing the emergence of EDM as a distinct field. Researchers and educators alike recognized the potential of applying data mining techniques to educational data, aiming to uncover patterns and insights that traditional analysis methods could not reveal.

Several factors have contributed to the rapid ascent of EDM within the academic and educational communities. The digitalization of education, including the adoption of Learning Management Systems (LMS), online courses, and digital assessment tools, has provided a rich source of data on student behavior, engagement, and performance. Concurrently, advancements in data mining technologies and methodologies have equipped researchers with the tools necessary to analyze this data effectively. The growing emphasis on personalized education and evidence-based teaching strategies has further fueled interest in EDM, highlighting its potential to tailor educational experiences to individual learners' needs.

Despite its promising applications, the nascent field of EDM faced several challenges. Early efforts were often limited by the available technology, data privacy concerns, and the lack of standardized data formats across different educational platforms. However, breakthroughs in machine learning algorithms, data storage, and processing technologies gradually overcame these obstacles. The development of specialized EDM tools and frameworks, along with increasing collaboration between data scientists, educators, and policymakers, has played a crucial role in advancing the field. These efforts have led to significant achievements in understanding and predicting student performance, identifying at-risk students, and developing adaptive learning systems.

The emergence of Educational Data Mining represents a significant milestone in the intersection of data science and education. By harnessing the power of data mining technology to analyze educational data, EDM has opened new avenues for enhancing teaching and learning processes. The field continues to evolve, driven by technological advancements, a deeper understanding of learning analytics, and the ongoing quest to improve educational outcomes [39, 40].

2.2 The Necessity for EDM Tools

Educational Data Mining (EDM) tools have emerged as essential instruments for educational practitioners and researchers. These tools enable the extraction of meaningful patterns from large educational data sets, aiding in the decision-making processes that shape educational strategies and influence student outcomes. EDM tools facilitate a deeper understanding of learning processes and student behavior, which can lead to more effective educational interventions [39, 40].

2.2.1 Impacts on Educational Decision Making

EDM tools significantly enhance the capability of educational administrators and teachers to make informed decisions. By analyzing patterns and trends within educational data, these tools can highlight successful teaching strategies, predict resource needs, and identify potential areas for improvement.

2.2.2 Enhancing Student Outcomes

The use of EDM tools also plays a crucial role in enhancing student outcomes. By providing insights into student performance and learning habits, educators can tailor their approaches to meet the individual needs of students, potentially reducing dropout rates and improving academic success [39].

2.3 Predictive Modeling in EDM

Predictive modeling is a cornerstone of Educational Data Mining, focusing on the application of data mining algorithms to predict and analyze academic performance. This approach not only helps in forecasting student successes but also in pinpointing challenges students might face, allowing for timely interventions.

2.3.1 Methods and Algorithms

A variety of methods and algorithms are employed in predictive modeling within EDM. Techniques such as decision trees, neural networks, and regression analysis are commonly used to analyze educational data. These methods provide a robust framework for understanding and predicting student performance, thereby enabling educators to offer targeted support.

2.3.2 Application to Student Performance

The application of predictive modeling in EDM extends to predicting student performance across various dimensions such as grades, course outcomes, and standardized testing. Researchers use historical data to train models that predict future performance, which can be crucial for early identification of students who might require additional support [40].

2.4 Feature Selection in Educational Data Mining

Feature selection is a critical component in Educational Data Mining (EDM), significantly impacting the refinement and accuracy of predictive models. This process involves identifying the most relevant features from a dataset that are closely linked to student performance. Esteemed researchers such as Estrera and Ramaswami have highlighted the importance of this process [41, 42].

2.4.1 Common Feature Selection Techniques

Various techniques for feature selection have been widely employed in the EDM community. Techniques like Chi-square Statistics and Information Gain are notable for their effectiveness in identifying impactful features [43]. Table 1.2 offers a concise overview of the feature selection methods used in recent EDM studies, illustrating the diversity of approaches in this field.

Manual and Filter-based Techniques

Manual selection and filter-based techniques such as correlation analysis and information gain are commonly used due to their simplicity and effectiveness in reducing dimensionality while maintaining the integrity of the data [18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30].

Wrapper Methods

Wrapper methods, including genetic algorithms and Principal Component Analysis (PCA), are tailored to specific learning models and can offer more nuanced feature selection by iteratively testing different subsets of features for optimal performance [33, 34, 35, 36, 37, 38].

2.4.2 Integration with Predictive Modeling

The integration of feature selection with predictive modeling represents a crucial advancement in EDM. This synergy is essential for developing more accurate and interpretable models that can predict student performance effectively. Innovative methodologies like the Adaptive Filter-based Selection Algorithm (AFSA) are being introduced to enhance model performance by optimally selecting features in conjunction with predictive algorithms.

2.4.3 Challenges and Opportunities

While the field has seen substantial growth, the existing research often adopts a narrow focus on individual algorithms or specific selection techniques. This approach has led to a fragmented understanding of how best to apply these methods comprehensively [44]. The future of EDM lies in overcoming these challenges by fostering a more integrated and holistic approach to feature selection and model development.

The continuous evolution of feature selection methodologies in EDM opens new avenues for research. Future studies should aim to develop comprehensive frameworks that integrate various feature selection techniques with advanced predictive models. Such integration is crucial for enhancing the accuracy, efficiency, and applicability of EDM tools in real-world educational settings.

Chapter 3

Feature Selection Techniques

In various fields such as machine learning and data analysis, the process of selecting the most relevant features from a high-dimensional dataset plays a critical role in building accurate predictive models. This section explores different feature selection techniques, highlighting their strengths, weaknesses, and applications, drawing insights from several research papers for reference.

3.1 Filter Methods

Filter methods represent a category of feature selection techniques that assess the relevance of each feature individually. Notable examples include statistical tests like chi-squared tests, mutual information, and correlation-based feature selection (CFS) [45, 46]. These methods generate a ranked list of features based on their individual significance, making them particularly suitable for datasets with a high number of features. However, they may not adequately capture complex interdependencies between features and are less effective at uncovering intricate feature relationships in the data. This limitation highlights the need for more advanced techniques, such as wrapper and embedded methods, that consider feature interactions and offer greater predictive power by integrating with the classifier algorithm. The choice of an appropriate significance threshold for filter methods can be crucial, and techniques like controlling for family-wise error rate (FWER) or false discovery rate (FDR) through methods like Bonferroni correction or Benjamini-Hochberg procedure are often employed to address this challenge. Additionally, in machine learning applications, the selection of the optimum threshold can be treated as a hyper-parameter and determined through cross-validation as part of the model selection process.

3.2 Wrapper Methods

Wrapper methods utilize the performance of a chosen classifier as a metric to select the best feature subset. These methods consider feature interactions and redundancies and aim to identify the best-performing set of features for a given classifier [47, 48]. They are computationally heavier than filter methods and might lead to overfitting in some cases due to the choice of classifier [49]. Unlike filter methods that produce ranked feature lists, wrapper methods generate a ”best” feature subset as the output. This means users are spared the task of determining an optimal threshold or the number of selected features, as the output is already a feature subset. However, this approach has its drawbacks, as it does not readily reveal the relative importance of individual features within the selected set. Additionally, wrapper methods are dependent on the chosen classifier, so the selected features may not remain optimal when a different classifier is employed. This lack of generalisability can lead to issues when applying the model to external datasets. Despite these limitations, wrapper methods have been shown to result in superior performance compared to filter methods, making them a valuable tool for feature selection in machine learning tasks.

3.3 Embedded Methods

Embedded methods integrate feature selection within the classifier algorithm itself during training [45]. Examples include decision tree-based algorithms (e.g., decision tree, random forest, gradient boosting) and regularisation models like LASSO or elastic net. These methods strike a balance between filter and wrapper methods, offering computational efficiency compared to wrapper methods while incorporating the classifier's bias into feature selection, which can improve classifier performance [50]. Embedded methods can capture feature interactions, with some decision tree-based algorithms even considering higher-order interactions. However, they typically require careful handling of highly dimensional datasets due to limitations in their ability to detect feature interactions as the number of features increases. Unlike some multivariate filters, decision tree-based algorithms like random forest do not automatically eliminate redundant features, which can impact their performance. To address this, hybrid methods that combine feature selection and random forest or other algorithms have been proposed. Penalized methods like LASSO, on the other hand, can discard redundant features but require explicit inclusion of interaction terms for feature interactions, which can be computationally prohibitive in highly dimensional data settings. Two-stage or hybrid strategies have been suggested to reduce search spaces in such cases.

3.4 Hybrid Methods

Hybrid methods seamlessly blend various feature selection strategies to capitalize on their respective advantages. An illustrative example involves initiating the process with univariate filter methods, which efficiently reduce the feature set size. This initial reduction curtails the computational load for subsequent wrapper or embedded methods, which are more computationally intensive but can capture intricate feature dependencies and interactions [51, 52]. These hybrid approaches strike a balance between computational complexity and performance. They offer a middle-ground solution between simple filter methods and the more computationally demanding wrapper and embedded methods. Notably, they tend to yield superior performance compared to standalone filter methods, while being less computationally burdensome than pure wrapper methods. For instance, Alzubi et al. [53] demonstrated the effectiveness of a hybrid CMIM + RFE-SVM strategy in classifying patients with various conditions, underscoring the superiority of this combined approach. Moreover, hybrid methods, while advantageous, have their limitations. They may inadvertently overlook relevant interacting features with no discernible individual effects, especially if these interactions are exclusively complex. This potential drawback arises because most filter methods are ill-equipped to model feature-feature interactions. However, using filter algorithms that can account for feature interactions can help mitigate this issue.

3.5 Integrative Methods

Integrative methods incorporate external knowledge to narrow the feature search space [54, 55]. By leveraging information from various sources, researchers can strategically filter and prioritize features based on their relevance to the task of interest. Specialized software and databases can assist in selecting "interesting" features by considering factors like prior knowledge, pathway information, and statistical evidence. This approach significantly reduces computational complexity by focusing the analysis on features associated with relevant information. For instance, Ma et al. [54] successfully identified feature interactions related to a specific outcome by narrowing their search to features previously linked to the outcome, features within known pathways, and those participating in relevant interactions. In another example, D'Angelo et al. [55] uncovered significant interactions associated with a particular condition by confining their search to relevant regions of interest. However, it's essential to acknowledge a limitation in these integrative approaches—the reliance on external a priori knowledge. The completeness of external data sources can impact the discovery of novel features lying beyond their scope.

3.6 Ensemble Methods

Ensemble feature selection methods are designed to enhance the feature selection process by amalgamating the outputs of multiple feature selection algorithms. The fundamental premise behind ensemble methods is to leverage the diversity inherent in individual algorithms, thereby improving the overall stability and resilience of the selection process, particularly when confronted with fluctuations in the input data. These approaches have demonstrated their effectiveness in various domains, showing that combining different feature selection techniques can often outperform relying on a single method. The key lies in carefully choosing a set of diverse feature selection algorithms to be included in the ensemble, such as filter and embedded methods, while ensuring that they produce distinct feature subsets. Evaluating the diversity among these algorithms is vital, with various metrics available for this purpose. Moreover, the success of ensemble feature selection methods also hinges on how the partial outputs from each algorithm are consolidated into a final output, a process known as aggregation. Several aggregation methods have been proposed, including union, intersection, mean, median, and weighted sum of feature rankings. Alternatively, majority voting systems can be employed to determine the final outcome by considering the collective predictions of classifiers trained on the feature subsets generated by each algorithm. In practice, ensemble feature selection methods have exhibited their potential in diverse applications. They have uncovered hidden relationships and insights that might have eluded single feature selection methods. The collective wisdom of multiple algorithms can enhance the robustness of feature selection, ultimately leading to more reliable and insightful results [56].

3.7 Exhaustive Searches for Higher-Order Feature Interactions

Exhaustive searches play a crucial role in the identification of significant feature interactions, with a focus on both pair-wise and higher-order interactions. Various algorithms, as mentioned in the literature [57, 58], have been developed for this purpose. However, it's important to note that these exhaustive searches are computationally demanding processes. In practical applications, researchers often employ hybrid or two-stage approaches to effectively manage the computational complexity associated with exhaustive searches [59, 60]. These approaches aim to reduce the feature space before initiating the exhaustive search, thus optimizing the overall efficiency of the process. It is worth mentioning that while exhaustive searches are invaluable for detecting higher-order feature interactions, they do come with computational challenges, especially when applied to large-scale datasets. As such, the research community continually seeks efficient and scalable algorithms to detect these interactions, particularly for higher-order cases. Exhaustive searches are a vital tool for identifying feature interactions in both pair-wise and higher-order scenarios, but their computational demands often necessitate the adoption of hybrid or two-stage approaches to manage the complexity effectively. Ongoing efforts focus on developing more efficient algorithms for detecting feature interactions, especially in cases involving higher-order interactions. In summary, the choice of feature selection method depends on the dataset and research goals. There is no one-size-fits-all approach, and researchers often adopt hybrid or ensemble methods to harness the strengths of multiple techniques.

Chapter 4

Methodology

This research delved into investigating the impact of feature selection techniques on student performance prediction using machine learning, with a focus on incorporating insights from educational data mining (EDM). Figure 4.1 encapsulates the essence of educational data mining, portraying a cyclic process where students actively interact, use, participate, and communicate within educational systems, generating valuable student data. This data, in turn, undergoes mining techniques, providing targeted recommendations to students while simultaneously revealing discovered knowledge to educators. This dual functionality forms a continuous loop. This research investigated the impact of feature selection techniques on student performance prediction using machine learning. The methodology involved data cleaning, exploratory data analysis, preprocessing, feature selection, and evaluation. Multiple machine learning algorithms were employed and evaluated using cross-validation. The experimental results and findings are discussed in detail. A visual representation of the methodology is provided in Figure 4.2, encapsulating the sequential flow from data cleaning to the evaluation of machine learning algorithms. The incorporation of EDM principles not only enhances the predictive modeling aspect but also aligns the study with the broader context of leveraging mined educational data to offer targeted insights for both students and educators, contributing to the continuous improvement of educational systems.

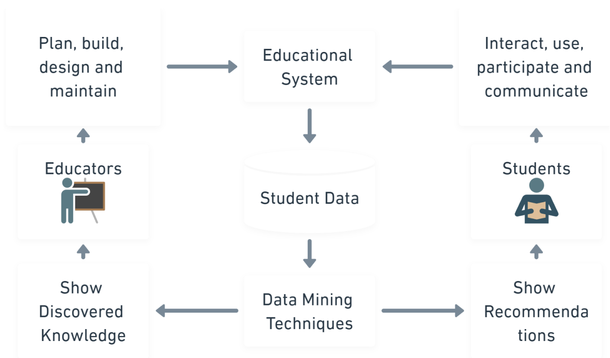


Figure 4.1: Different Interconnected Components of EDM.

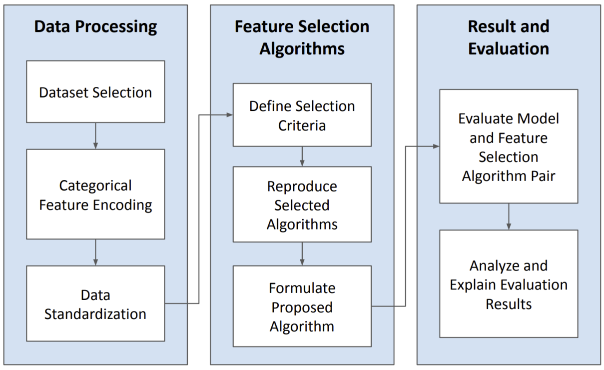


Figure 4.2: Block Diagram of Methodology.

Table 4.1: Details comparison of the dataset used in this study

Dataset	Year	Source	Instance	Attributes	Feature	Associated Tasks
XAPI[31]	2016	Kaggle	480	16	Demographic, Academic, Behavioral	Classification
SSP[61]	2014	UCI ML Archive	1044	33	Demographic, Academic, Behavioral, Health	Classification, Regression
HESP[62]	2022	Kaggle	145	31	Demographic, Academic, Behavioral	Classification, Regression
WOC2[63]	2020	GitHub	486	9	Academic	Classification

4.1 Dataset Description

The research incorporates a comparative analysis of four distinct educational datasets, each offering unique insights into student performance and behavior. The first dataset, xAPI-Educational Mining Dataset (xAPI), was procured from Kaggle in 2016 [31]. Comprising 480 instances and 16 attributes, xAPI includes demographic, academic, and behavioral features primarily tailored for classification tasks. The second dataset, Secondary Student Performance Dataset (SSP), was retrieved from the UCI ML Archive in 2014 [61]. With 1044 instances and 33 attributes, SSP encompasses a diverse range of features, including demographic, academic, behavioral, and health-related attributes. This dataset is versatile, and suitable for both classification and regression tasks. The third dataset, Higher Education Student Performance Dataset (HESP), was sourced from Kaggle in 2022 [62]. Comprising 145 instances and 31 attributes, HESP covers demographic, academic, and behavioral features, making it applicable for classification and regression tasks. The fourth dataset, Student Performance Prediction - Western-OC2-Lab (WOC2), was obtained from GitHub in 2020 [63]. With 486 instances and 9 attributes focusing on academic features, WOC2 is specifically designed for classification tasks. Table 4.1 provides a comprehensive overview of the key characteristics of these datasets, including the year of acquisition, source, number of instances, number of attributes, types of features, and associated tasks. This comparison aids researchers in understanding the diverse strengths and potential applications of each dataset in educational research and predictive modeling.

4.2 Exploratory Data Analysis

The foundational step in this study involved a comprehensive examination of the dataset through a meticulous process of Exploratory Data Analysis (EDA). This phase aimed not only to ensure the integrity and reliability of the dataset but also to extract meaningful insights that would guide subsequent decisions in data preprocessing, feature engineering, and model selection. The EDA process comprised a multifaceted exploration of the dataset's characteristics, shedding light on its intricacies and patterns.

4.2.1 Data Cleaning Procedures

Rigorous data cleaning procedures were implemented to rectify potential discrepancies and enhance the dataset's quality. Missing values were systematically addressed through imputation or removal, ensuring that the dataset was free from gaps that could compromise subsequent analyses. Duplicate entries were identified and removed to maintain data integrity, while columns with constant values were eliminated as they did not contribute meaningful information for model training.

4.2.2 Exploration of Dataset Characteristics

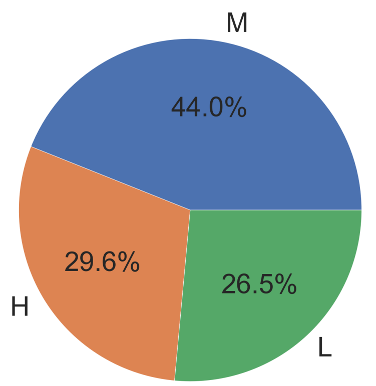
The heart of the EDA involved a deep dive into the dataset's characteristics. Descriptive statistics were calculated to provide an overview of the dataset. From the pie charts in Figure 4.3 that shows target column distribution of all four datasets, it can be observed that XAPI, SSP and HESP data distribution is fairly balanced whereas the WCO2 data distribution is imbalanced. This observation has been addressed in the data preprocessing section where state-of-the-art data balancing methods have been used to counteract the issue.

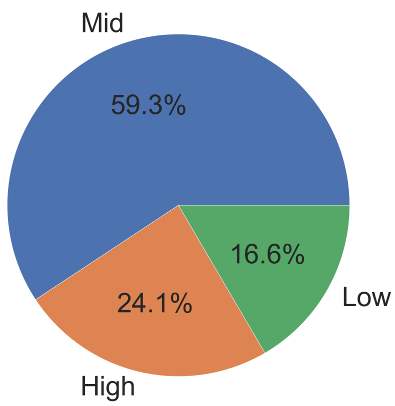
4.2.3 Attribute Types and Semantic Analysis

An in-depth examination of attribute types and semantics was conducted to unravel the nature of the features. Categorical variables were identified and analyzed for their cardinality, while numerical variables underwent statistical scrutiny to comprehend their distributions and potential outliers. Semantic analysis aimed to uncover the meaning and relevance of each attribute in the context of the study, laying the groundwork for informed feature engineering.

4.2.4 Data Visualisation

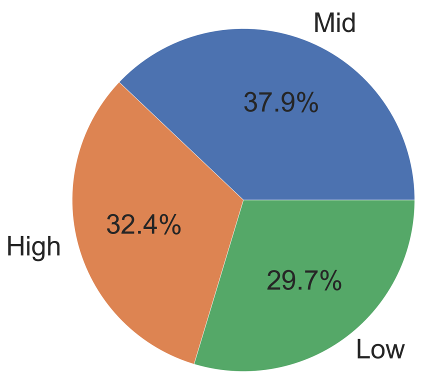
The visualization of data played a crucial role in gaining intuitive insights into complex relationships within the dataset. Correlation heatmaps and class distribution pie charts were employed to visualize the interactions between different features, providing a visual narrative of potential dependencies and correlations. This visual exploration facilitated

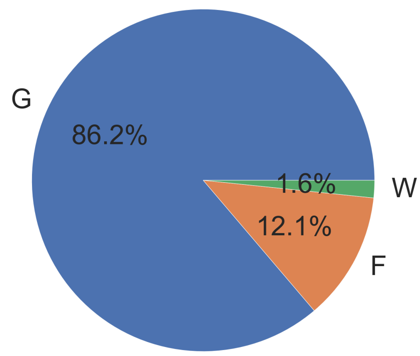




(a) Target column data distribution of XAPI Dataset

(b) Target column data distribution of SSP dataset





(c) Target column data distribution of HESP dataset

(d) Target column data distribution of WOC2 dataset

Figure 4.3: Target column data distribution of four datasets

the identification of potential patterns and outliers, guiding subsequent decisions in the modeling process.

A correlation matrix is a tabular representation that displays the correlation coefficients between multiple variables, where each cell in the table represents the correlation between two specific variables. This matrix typically computes pairwise correlations primarily using the Pearson correlation coefficient. This coefficient quantifies the linear relationship between variables, ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no correlation. The primary utility of calculating a correlation matrix in feature selection processes for machine learning models lies in its ability to identify columns within the DataFrame that exhibit the highest absolute correlations. This identification is crucial as features with high correlation may convey redundant information. Consequently, eliminating one of such highly correlated features can significantly reduce the dimensionality of the model without substantial loss of information.

In data analysis, particularly when dealing with complex datasets, visualizing the full correlation matrix is often impractical due to its extensive size. To effectively manage this challenge and enhance the clarity of data visualization, we focus on selecting the most informative columns by identifying those with the highest absolute correlation values. This approach allows us to distill the correlation matrix to its most significant relationships, simplifying the visualization and making it more accessible for interpretation while preserving the critical insights from the larger dataset.

The correlation matrix depicted in Figure 4.4 illustrates the relationships between various features in the XAPI dataset, reflecting educational data dynamics. Notably, Nationality and Place of Birth exhibit a high positive correlation (0.8), indicating a strong alignment between students' nationalities and their birthplaces. Conversely, StageID and GradeID demonstrate a strong negative correlation (-1.0), suggesting an inverse relationship possibly due to the dataset's educational stage and grade level encoding. Academic engagement indicators such as raisedhands, VisITedResources and AnnouncementsView positively correlate with each other and with the final class feature, highlighting that increased student engagement is associated with better academic performance. Moreover, ParentAnsweringSurvey and ParentschoolSatisfaction share a moderate positive correlation (0.5), revealing that parent satisfaction with the school correlates with their participation in surveys, which also relates positively to the Relation feature, suggesting a link between parent-school relationship quality, satisfaction, and survey participation.

The correlation matrix presented in Figure 4.5 represents the interrelations among various factors in the SSP dataset, hinting at the complexities within student socio-personal dynamics. Notable observations include the positive correlation between age and failures (0.3), indicating that older students tend to have more academic failures, potentially reflecting challenges like grade repetition. The matrix reveals a strong positive correlation between Medu (mother's education) and Fedu (father's education) at 0.6, underscoring the aligned educational backgrounds within families. Interestingly, the job sectors of mothers (Mjob) and fathers (Fjob) are moderately correlated (0.2), suggesting some level of sim-

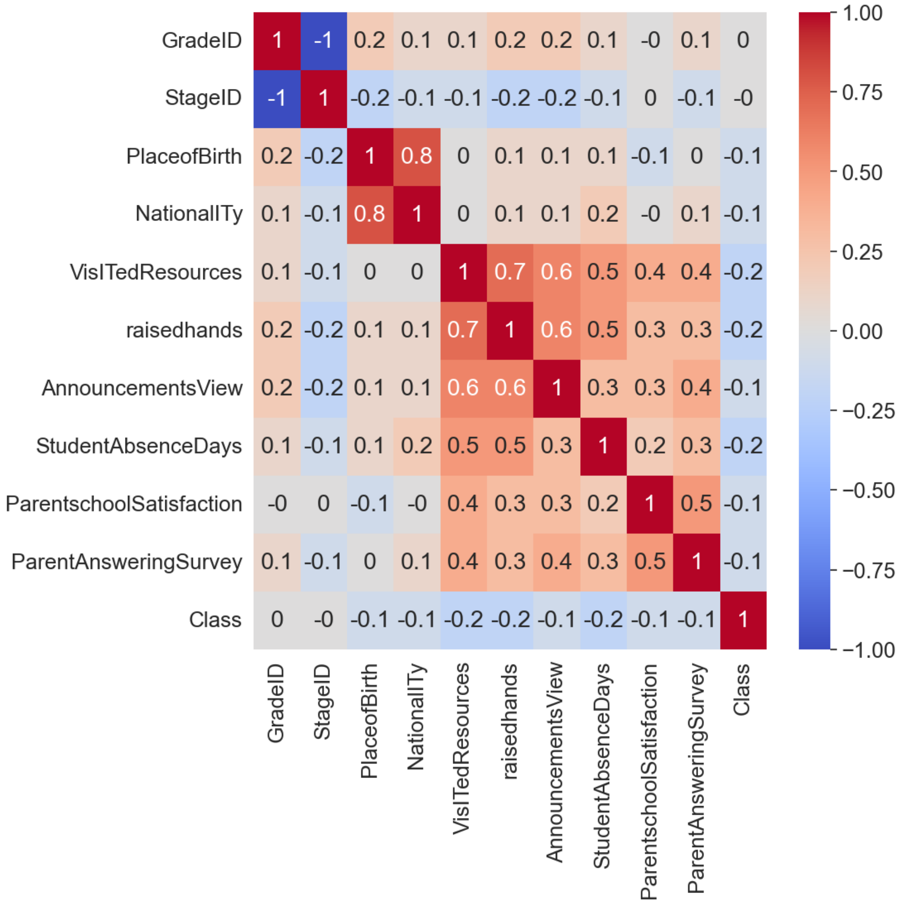


Figure 4.4: Correlation heatmap of dataset XAPI

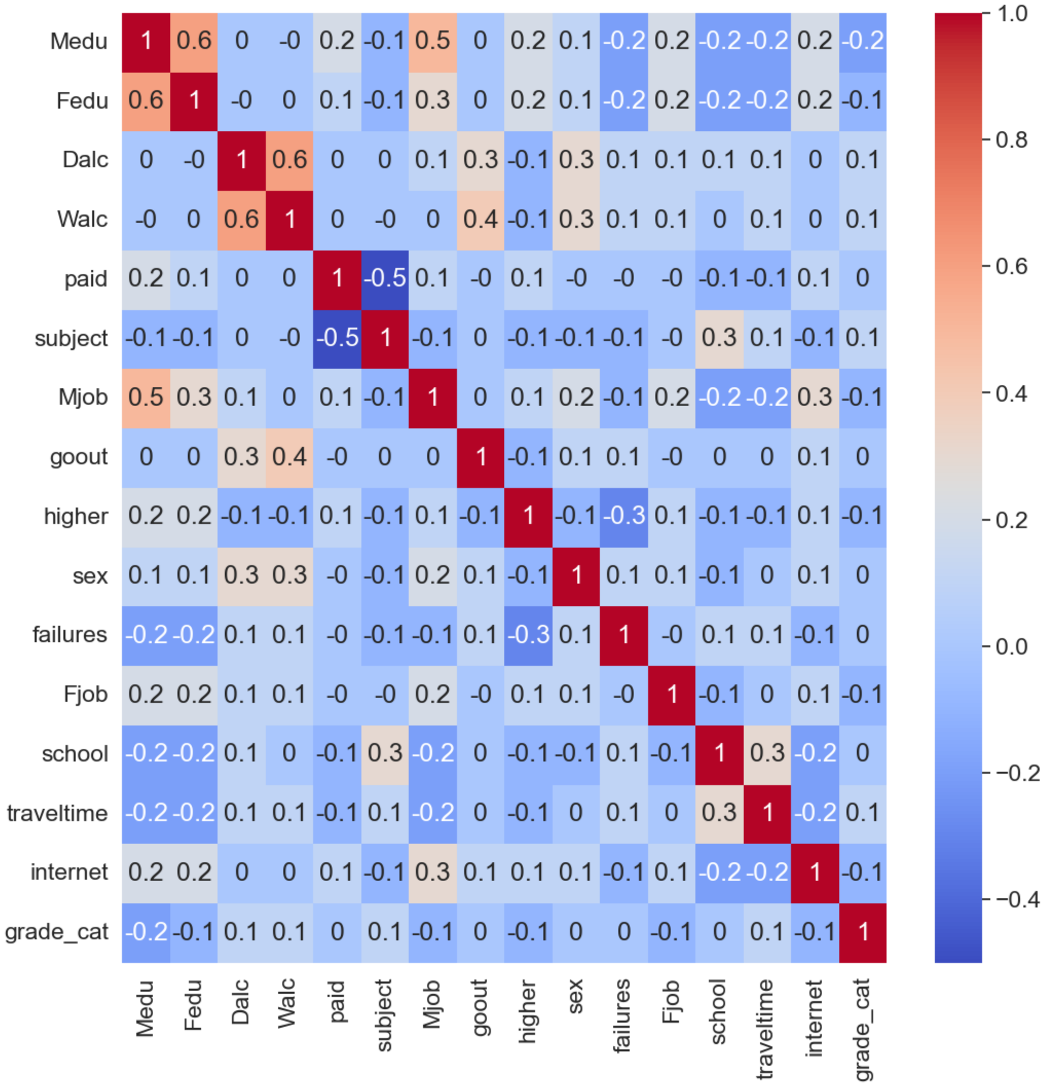


Figure 4.5: Correlation heatmap of dataset SSP

ilarity in parental occupations. Another key insight is the positive correlation between going out (goout) and weekend alcohol consumption (Walc) at 0.4, which could imply social dynamics influencing drinking habits among students.

A significant negative correlation is observed between paid and subject at -0.5, which might indicate that students taking paid classes do not always correlate with subject, suggesting nuances in the effectiveness or necessity of these classes. The relationship between travel time (traveltime) and school choice (school) is positively marked at 0.3, indicating students attending a specific school might face longer commutes. Additionally, there is correlation between desires for higher education (higher) and internet access (internet) at 0.1, pointing towards the role of digital connectivity in educational aspirations. This matrix, therefore, provides a multifaceted view of the factors affecting students' academic and personal lives, highlighting the influence of family background, personal habits, and educational support on their experiences and outcomes.

The correlation matrix depicted in Figure 4.6 analyzes various factors related to higher education students' performance (HESP), showcasing how different variables interact within an academic context. Gender shows a good correlation with expected GPA (0.3), suggesting demographic patterns in educational pathways. A negative correlation between scholarship status and age (-0.3) could indicate younger students are more likely to receive scholarships. Work experience positively correlates with gender (0.2) and negatively with scholarship (-0.2), highlighting potential financial independence or necessity among male students.

Activities show a negative correlation with expected GPA (-0.3), suggesting extracurricular engagement might impact academic expectations. Partner status has a slight positive effect on salary (0.1) and living arrangements (0.3), indicating relationship status may influence financial and living conditions. Transport and living conditions share a positive correlation (0.3), reflecting geographical and socio-economic factors affecting student life.

Notably, cumulative GPA (CUML_GPA) positively correlates with expected GPA (EXP_GPA, 0.7), indicating alignment between students' performance expectations and outcomes. However, a strong negative correlation between classroom engagement and grade (-0.3) suggests that active classroom participation does not always correlate with higher grades, a counterintuitive finding that warrants further investigation.

Overall, this matrix reveals complex interdependencies among socio-demographic factors, educational background, personal life, and academic performance in higher education, highlighting the multifaceted nature of student success.

The correlation matrix for the WOC2 dataset, presented in Figure 4.7, provides insight into the relationships between various academic assessments and the final class performance of students. The matrix highlights that none of the initial quizzes (Quiz01) show a significant correlation with the final class performance, suggesting that these early assessments may not be predictive of final outcomes. However, there is a notable progression in the significance of assignments and exams over time towards class performance.

Assignment01 shows a moderate correlation with Assignment02 (0.6) and Assign-

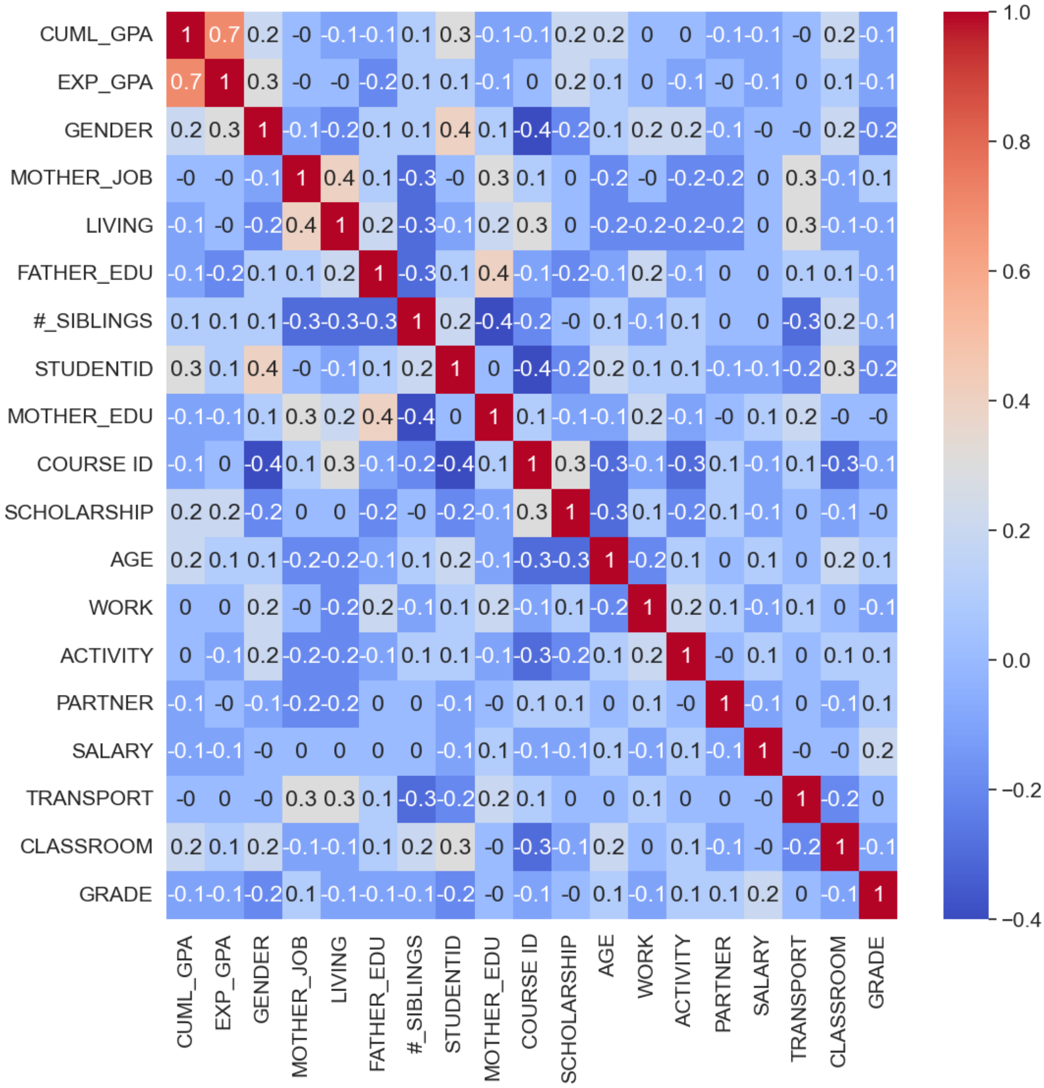


Figure 4.6: Correlation heatmap of dataset HESP

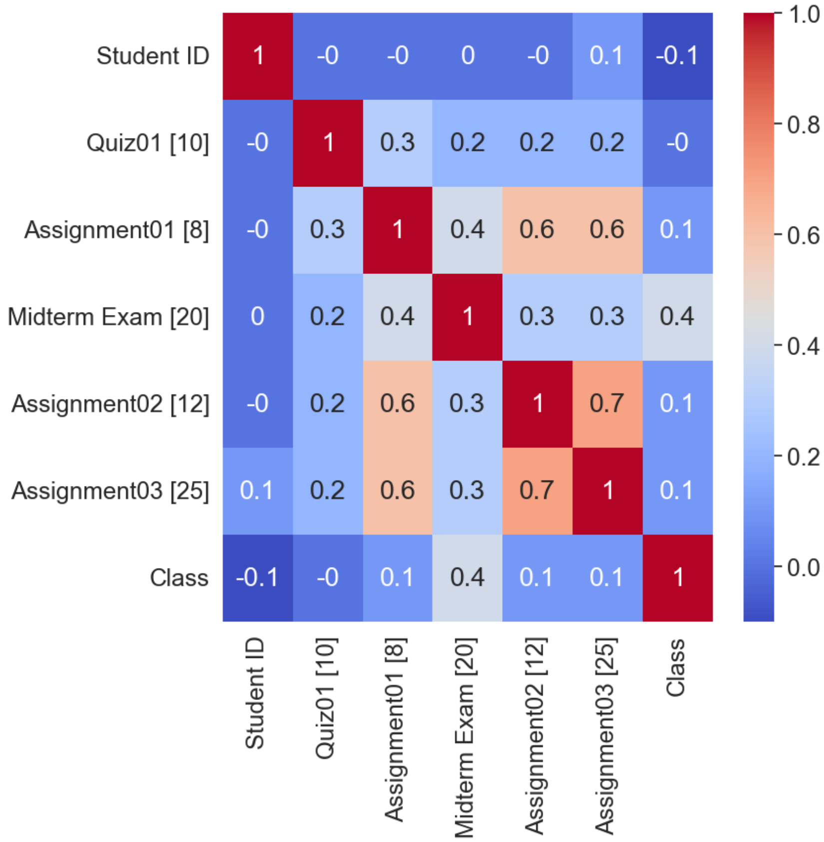


Figure 4.7: Correlation heatmap of dataset WOC2

ment03 (0.6), indicating a consistency in students' performance across assignments. The Midterm Exam holds a stronger correlation (0.4) with the final class, highlighting its predictive value for final performance. Interestingly, Assignment03, which has the highest weight (25), also shows a strong correlation with Assignment02 (0.7) and a moderate correlation (0.1) with the final class performance, underlining the importance of consistent performance in higher-weight assignments for overall success.

The matrix also suggests that while quizzes and early assignments may not directly predict the final class performance, they are closely related to each other, hinting at an underlying pattern of student engagement and understanding. This nuanced view provided by the correlation matrix underscores the complexity of academic performance, where later, more heavily weighted assessments have a more pronounced impact on the final class grade, reflecting the cumulative nature of learning and assessment in the educational process.

4.2.5 Dataset Heterogeneity and Formatting Challenges

The datasets under consideration showcased heterogeneity in terms of features, with a diverse range spanning demographic, behavioral, academic, and health-related variables. However, not all datasets were homogeneously formatted for classification tasks. Variations in the number of target classes and class imbalances were identified, necessitating a nuanced approach in the subsequent data preprocessing stage to address these challenges effectively. The insights gleaned from this extensive EDA not only facilitated a profound understanding of the dataset's intricacies but also laid the foundation for informed decisions in subsequent stages of the study. The interplay of statistical analyses, visualizations, and semantic explorations during the EDA phase enriched the dataset's interpretability, contributing significantly to the overall efficacy and accuracy of the subsequent machine-learning endeavors.

4.3 Data Preprocessing

The data preprocessing phase in this study involved a meticulous series of steps aimed at refining the dataset to enhance its suitability for machine learning applications. Each step was strategically designed to address specific challenges, ensuring the robustness and reliability of subsequent analyses.

4.3.1 Handling Categorical Data

A critical aspect of the data preprocessing process was the treatment of categorical data. To enable the seamless integration of the dataset with machine learning algorithms, categorical variables were encoded using Label Encoders. This transformation facilitated the representation of categorical information in a numerical format, allowing for effective utilization in predictive models.

4.3.2 Stratified Data Splitting

To ensure robust model evaluation, the dataset was divided into training and testing sets with an 80-20 split ratio. Stratified sampling was employed during this process to maintain the distribution of target classes in both the training and testing sets. This approach is particularly valuable when dealing with imbalanced datasets, as it preserves the proportionality of class instances, preventing skewed model performance assessments.

4.3.3 Standardisation

Standardization was applied to bring all features to a common scale, mitigating the impact of differing measurement units. This step is crucial for algorithms that are sensitive to the scale of input features, ensuring that each feature contributes proportionally to the model's learning process. Standardizing the data fosters improved convergence and stability in machine learning algorithms.

4.3.4 Mitigating Class Imbalances with ADASYN

Class imbalances, where certain classes have significantly fewer instances, pose challenges in classification tasks. To address this, we employ the Adaptive Synthetic Sampling (ADASYN) technique. ADASYN dynamically generates synthetic samples for the minority class based on local density, effectively balancing class distribution and preventing model bias towards the majority class. This approach, as presented by He et al. [64], adapts to the difficulty level of learning for each minority class instance, improving classification by reducing bias and shifting the decision boundary towards challenging examples.

4.3.5 Transforming SSP and HESP Datasets with the Five-Number Method

The SSP and HESP datasets underwent a unique transformation using the five-number method to address specific characteristics. In this method, the class counts within these datasets were divided into three categories: "Low," "Mid," and "High." This categorization aimed to provide a nuanced representation of the class distribution, facilitating a more granular evaluation of model performance across different levels of class representation. The adoption of the five-number method in these datasets was driven by the need for a tailored approach to handling class distributions, acknowledging the nuances inherent in the SSP and HESP datasets. The intricate combination of these preprocessing techniques laid the groundwork for subsequent stages of the study, ensuring that the dataset was optimized in terms of structure, scale, and class distribution for effective machine learning model development and evaluation.

4.4 Feature Selection Methods

The feature selection phase is integral to optimizing machine learning models by streamlining the input space, enhancing efficiency, and mitigating the risk of overfitting. In this section, we delve into a diverse set of feature selection algorithms, each designed to strategically identify and retain the most informative attributes. Recursive Feature Elimination (RFE) systematically removes less impactful features, refining the model's input set. Forward Selection (FS) incrementally builds the feature subset based on predictive performance, iteratively enhancing model accuracy. Genetic Algorithm (GE) employs evolutionary principles to evolve a subset of features, mimicking the process of natural selection to arrive at an optimal feature configuration. Additionally, we introduce our novel contribution, the Adaptive Feature Selection Algorithm (AFSA), which adapts dynamically to the dataset's characteristics, offering a tailored and efficient approach to feature selection. Through a comparative exploration of these methods, we aim to unravel their efficacy in enhancing model performance and contributing to the overall success of the machine learning endeavors in this study.

4.4.1 Recursive Feature Elimination (RFE)

Recursive Feature Elimination (RFE) stands as a powerful feature selection algorithm designed to iteratively enhance model performance by systematically eliminating the least informative features. The fundamental principle underlying RFE involves a recursive process that begins with the entire feature set. In each iteration, a model is trained on the current set of features, and the features are then ranked based on a specified criterion, often tied to their contribution to predictive accuracy. Subsequently, the feature with the lowest ranking is systematically removed. This recursive cycle continues until the algorithm converges to the optimal subset of features, determined by predefined criteria or through cross-validation. The algorithm's iterative nature allows it to capture the evolving dynamics of feature importance, ensuring that the most relevant attributes are retained in the final subset. This adaptability makes RFE particularly valuable in scenarios where the initial feature set might be extensive, containing redundant or irrelevant information. The iterative refinement process not only contributes to computational efficiency but also guards against overfitting by focusing on the most informative features. Algorithmically, the RFE process involves several key steps, as outlined in Algorithm 4.1. The initial model is trained on the complete feature set, and at each iteration, the least informative feature is identified and eliminated. The algorithm converges when the specified criteria for feature subset optimization are met. Through this nuanced process, RFE serves as a sophisticated tool for enhancing the efficiency and predictive power of machine learning models by strategically paring down the feature space to its most informative components.

4.4.2 Forward Selection (FS)

Forward Selection (FS) emerges as a dynamic feature selection algorithm, distinguished by its incremental approach to constructing a feature subset through the iterative addition of the most informative features. Unlike backward elimination strategies, FS initiates with an empty feature set and progressively augments it until a predetermined termination criterion is satisfied. This criterion is typically tied to optimizing model performance, whether through maximizing predictive accuracy or adhering to a specific evaluation metric. The fundamental premise of FS lies in its adaptability and ability to navigate through the feature space, strategically incorporating attributes that contribute most significantly to the model's efficacy. This stepwise refinement process ensures that the algorithm converges to a feature subset optimized for the given task, steering clear of redundant or less impactful attributes. Algorithm 4.2 outlines the procedural steps of the Forward Selection algorithm. It commences with an empty feature set, and at each iteration, evaluates the impact of adding each remaining feature. The feature that maximizes the specified criterion is selected for inclusion, and this process iterates until the termination criterion is met. Forward Selection is particularly advantageous in scenarios where the initial feature set is extensive, and computational efficiency is a priority. By strategically building the feature subset based on the most informative attributes, FS offers a tailored and efficient

Algorithm 4.1: Recursive Feature Elimination (RFE)

Input: Input DataFrame, Target Column, Model, Number of Folds for Cross-Validation (cv)

Output: Selected Features

```
X ← Input DataFrame
y ← Target column of DataFrame
selected_features ← set of all features
n ← number of features in selected_features
optimal_score ← 0
optimal_feature_subset ← ∅
while true do
    worst_feature ← None
    for feature in selected_features do
        candidate_features ← selected_features − feature
        X_train_fs ← X[candidate_features]
        cv_accuracy_fs ← cross_val_score(Model, X_train_fs, y, cv, 'accuracy')
        avg_accuracy ← mean(cv_accuracy_fs)
        if avg_accuracy > optimal_score then
            optimal_score ← avg_accuracy
            optimal_feature_subset ← candidate_features
            worst_feature ← feature
        end if
    end for
    if worst_feature is not None then
        selected_features ← selected_features − worst_feature
        n ← n - 1
        if n ≤ 1 then
            break
        end if
    else
        break
    end if
end while
```

approach to feature selection, contributing to the overall optimization of machine learning models.

Algorithm 4.2: Forward Selection (FS)

Input: Input DataFrame, Target Column, Model, Number of Folds for Cross-Validation (cv)

Output: Selected Features

```
X ← Input DataFrame
y ← Target column of DataFrame
selected_features ← ∅
n ← number of features in X
optimal_score ← 0
optimal_feature_subset ← ∅
while true do
  best_feature ← None
  for feature in X do
    if feature not in selected_features then
      candidate_features ← selected_features + feature
      X_train_fs ← X[candidate_features]
      cv_accuracy_fs ← cross_val_score(Model, X_train_fs, y, cv, 'accuracy')
      avg_accuracy ← mean(cv_accuracy_fs)
      if avg_accuracy > optimal_score then
        optimal_score ← avg_accuracy
        best_feature ← feature
      end if
    end if
  end for
  if best_feature is not None then
    selected_features ← selected_features + best_feature
    n ← n - 1
    if n ≤ 0 then
      break
    end if
  else
    break
  end if
end while
```

4.4.3 Genetic Algorithm (GA)

A genetic algorithm for feature selection, presented in Algorithm 4.3, is a heuristic optimization technique inspired by the process of natural selection and evolution. It is employed to automatically determine the most relevant subset of features from a given dataset for use in machine learning models. The primary objective is to enhance model performance, reduce overfitting, and enhance interpretability by identifying and retaining the most informative features while discarding irrelevant or redundant ones. In this algorithm, a population of potential feature subsets is initialized, typically represented as binary strings, where each bit corresponds to the inclusion or exclusion of a specific feature. The algorithm then evaluates the fitness of these subsets by training and cross-validating a machine-learning model with each feature combination. Fitness is typically measured in terms of the model's performance on a specified metric, such as accuracy or F1-score. The fittest feature subsets, those that yield the best model performance, are selected to form a new generation. Genetic operations like crossover and mutation are applied to the selected individuals to create a new population. Crossover combines two feature subsets to produce offspring with a mixture of their parent's characteristics. Mutation introduces small random changes to the feature subsets to maintain genetic diversity. This iterative process continues for a fixed number of generations, allowing the algorithm to converge towards a feature subset that optimizes model performance. Ultimately, the algorithm outputs the best-selected features, which can then be used to train a final machine-learning model with improved generalization and efficiency.

Algorithm 4.3: Genetic Algorithm for Feature Selection

Input: Input DataFrame, Target Column, Population Size, Number of Generations, Mutation Rate, Model, Number of Folds for Cross-Validation (cv)
Output: Best Feature Subset

$X \leftarrow$ Input DataFrame
$y \leftarrow$ Target column of DataFrame
$n \leftarrow$ number of features in X
population $\leftarrow$ randomly initialize population with binary feature selection masks
for *generation* $\leftarrow$ 1 to $Number\ of\ Generations$ **do**
  scores $\leftarrow$ empty list
  for *individual* in *population* **do**
    $selected\_features \leftarrow$ features selected based on the individual's mask
    $X\_train\_fs \leftarrow$ X[selected_features]
    cv_accuracy $\leftarrow$ cross_val_score(Model, X_train_fs, y, cv, 'accuracy')
    Append *cv_accuracy* to *scores*
  end for
  best_indices $\leftarrow$ indices of the top-performing individuals
  best_population $\leftarrow$ individuals at *best_indices*
  best_individual $\leftarrow$ individual with the highest accuracy among *best_population*
  new_population $\leftarrow$ *best_population*
  while size of *new_population* $<$ Population Size **do**
    $parent1, parent2 \leftarrow$ randomly select two individuals from *best_population*
    crossover_point $\leftarrow$ randomly select a crossover point
    child $\leftarrow$ combine $parent1$ and $parent2$ using crossover at *crossover_point*
    Apply mutation with probability Mutation Rate to *child* (flip some bits)
    Append *child* to *new_population*
  end while
  population $\leftarrow$ *new_population*
end for
$best\_selected\_features \leftarrow$ features selected based on *best_individual*'s mask
return $best\_selected\_features$

4.4.4 Adaptive Feature Selection Algorithm (AFSA)

The Adaptive Feature Selection Algorithm (AFSA), as delineated in Figure 4.8 and Algorithm 4.4, represents a sophisticated approach to feature selection, combining multiple ranking methods to ascertain the relevance of each feature. The algorithm employs a diverse array of feature ranking methods, including Information Gain, Chi-Square Test, Mutual Information, Relief, and Gini Importance. Each feature's relevance is quantified through these methods, followed by the normalization of ranks and computation of an average rank for individual features. For the normalization of ranks, we used min-max normalization. The core of AFSA lies in its iterative feature selection process, where features are evaluated and selected based on their rank order. The algorithm begins with the top-ranked feature and progressively adds features to the subset, each time assessing the model's cross-validation accuracy. A unique aspect of this process is the incorporation of the *is_improvement_good_enough* function, which introduces a dynamic threshold for determining the significance of accuracy improvement upon the inclusion of new features. This function operates under the premise that as the algorithm proceeds down the list of ranked features, the criterion for improvement becomes less stringent. It begins with a 5% threshold for top-ranked features, decrementing by 0.5% for each subsequent feature. However, the minimum threshold is set at 1%, ensuring that only meaningful improvements contribute to the retention of a feature. This dynamic thresholding serves a dual purpose: it not only seeks to enhance the model's performance by including impactful features but also ensures the compactness of the feature set by prioritizing significant improvements, especially for higher-ranked features. In essence, AFSA is designed to reduce dataset dimensionality by adaptively selecting features that offer the most information. This combination of diverse feature ranking methods with an iterative, threshold-based selection strategy enables AFSA to conduct a comprehensive assessment of feature relevance. Such an approach not only enhances the accuracy of machine learning models but also contributes to their efficiency by focusing on the most influential features, thereby streamlining the predictive process.

4.4.5 Delineating AFSA from Forward Selection: A Novelty Perspective

To address the distinctiveness of the Adaptive Feature Selection Algorithm (AFSA) from the traditional Forward Selection (FS) approach, it is imperative to underscore the foundational and operational differences that underscore AFSA's novelty and its contribution to the existing corpus of knowledge on feature selection methodologies. While both strategies aim at enhancing model performance through judicious feature selection, the mechanisms they employ to achieve this objective exhibit significant divergences.

Methodological Distinction: The crux of Forward Selection lies in its incremental nature, beginning with an empty set and progressively adding features that maximize a performance criterion until no further improvements can be made. This process, although systematic, does not incorporate multiple feature ranking metrics or a dynamic

Algorithm 4.4: Adaptive Feature Selection Algorithm (AFSA)

```
Input: X_train, X_test, y_train, y_test, model
Output: Selected Features, Evaluation Metrics
Define function is_improvement_good_enough(last-_accuracy, current_accuracy, feature_position):
    threshold ← max(0.01, 0.05 - 0.005 × feature_position)
    return current_accuracy - last_accuracy > threshold
Define function get_feature_rank(DataFrame, TargetColumn):
    Calculate feature scores using Information Gain, Chi-Square Test, Mutual Information, Relief, and Gini Importance
    Rank the features based on feature scores
    Normalize each rank and calculate the average rank
    Sort features based on the average rank in descending order
    return sorted list of features
sorted_features ← get_feature_rank(Input DataFrame, Target Column)
selected_features ← empty list
best_accuracy ← 0
for feature in sorted_features do
  selected_features.append(feature)
  X_train_fs ← X_train[selected_features]
  X_test_fs ← X_test[selected_features]
  cv_accuracy ← mean(cross_val_score(model, X_train_fs, y_train, 'accuracy'))
  if cv_accuracy > best_accuracy and is_improvement_good_enough(best_accuracy, cv_accuracy, idx) then
    best_accuracy ← cv_accuracy
  else
    selected_features.remove(feature)
  end if
end for
Train the model with selected features and evaluate on the test set
Calculate evaluation metrics including accuracy, F1 score, precision, and recall for both cross-validation and test sets
Record execution time and list of selected features
return selected_features, evaluation metrics
```

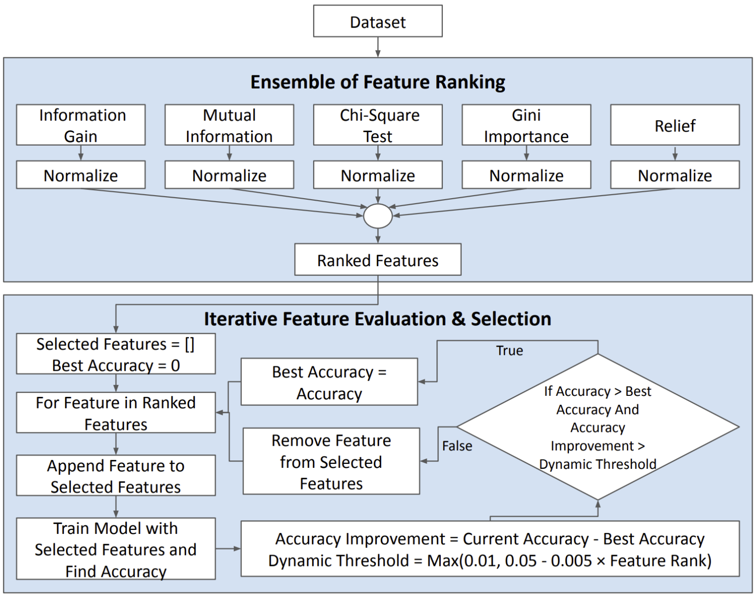


Figure 4.8: Flowchart of Adaptive Feature Selection Algorithm (AFSA).

thresholding mechanism for feature inclusion. Conversely, AFSA integrates a preliminary comprehensive feature ranking phase, employing multiple evaluation metrics such as Information Gain, Chi-Square Test, Mutual Information, Relief, and Gini Importance. This multifaceted ranking not only provides a robust basis for feature evaluation but also enhances the algorithm's adaptability to varied data characteristics.

Adaptive Selection Process: Beyond the initial ranking, AFSA distinguishes itself through an adaptive feature inclusion strategy. Unlike FS, which adheres to a static metric for feature addition, AFSA utilizes a dynamic thresholding function (*is_improvement_good_enough*) to determine the significance of including a new feature based on its position in the ranked list. This adaptability allows AFSA to maintain a balance between model complexity and performance, ensuring the inclusion of features that offer substantial improvement and discarding those with marginal benefits.

Computational Efficiency: The operational efficiency of AFSA further sets it apart from FS. By leveraging the ranked features list and dynamically adjusting the inclusion criteria, AFSA minimizes unnecessary model evaluations, reducing computational time. This efficiency is particularly crucial in big data contexts, where the volume of features can significantly impact the feasibility of feature selection processes.

Contribution to Knowledge: The amalgamation of these features—multi-criteria ranking, adaptive thresholding, and enhanced computational efficiency—positions AFSA as a significant advancement over traditional FS. It not only broadens the methodological toolkit available for feature selection but also introduces a nuanced approach that addresses some of the limitations inherent in existing strategies. Through these innovations, AFSA contributes to the body of knowledge by offering a versatile, efficient, and effective solution for feature selection, applicable across a diverse array of machine learning tasks and data contexts.

In summary, the distinction between AFSA and FS is not merely a matter of implementation details but a fundamental difference in approach, efficiency, and adaptability. By providing a detailed exposition of these differences, we aim to clarify the unique contributions of AFSA to the field of feature selection in machine learning.

4.4.6 Rationale Behind AFSA's Hybrid Approach

The inception of AFSA was driven by a pivotal goal: to amalgamate the efficiency and expediency inherent in filter-based feature selection approaches with the precision and performance optimization characteristic of more iterative methodologies, such as Forward Selection and Recursive Feature Elimination. This hybrid strategy stems from a critical observation: while filter-based methods excel in reducing training time due to their simplicity and independence from learning algorithms, they often do not account for feature dependencies or interaction effects, potentially compromising predictive accuracy. Conversely, methods like Forward Selection, though highly effective in optimizing model performance through exhaustive feature evaluation, can be prohibitively time-consuming and computationally intensive, especially for datasets with a high-dimensional feature space.

The AFSA framework was thus designed to bridge this gap, leveraging the speed of filter-based approaches through an initial feature ranking phase that employs a diverse set of metrics, including Information Gain, Chi-Square Test, Mutual Information, Relief, and Gini Importance. This phase serves to quickly reduce the dimensionality of the dataset, focusing subsequent analysis on features most likely to impact model performance. The iterative selection process that follows, incorporating a dynamic threshold for accuracy improvement, mirrors the thoroughness of more computationally demanding methods. By adjusting the improvement threshold based on feature ranking, AFSA ensures that early additions must meet a higher standard of performance enhancement, reflecting their presumed greater impact, while allowing for more flexibility as the selection process progresses. This nuanced approach ensures that AFSA remains both swift and effective, marrying the rapid training times of filter-based methods with the accuracy and model performance enhancements typical of iterative selection techniques.

Empirical evidence from our experiments substantiates the efficacy of this approach. Training time comparisons between AFSA and traditional methodologies underscore the significant reductions in computational overhead achieved by AFSA, without sacrificing model accuracy. These findings are particularly compelling, demonstrating that AFSA provides a viable pathway to achieving high-performance machine learning models in a fraction of the time required by conventional methods. Through this innovative amalgamation of speed and precision, AFSA represents a significant advancement in the field of feature selection, offering a scalable and efficient solution for high-dimensional data analysis.

4.4.7 Theoretical Computational Efficiency Analysis

To thoroughly evaluate the computational efficiency of our proposed Adaptive Feature Selection Algorithm (AFSA) relative to established feature selection methods such as Recursive Feature Elimination (RFE), Forward Selection (FS), and Genetic Algorithms (GA), we embark on a theoretical analysis of their respective time complexities. This analysis not only clarifies the inherent computational demands of each algorithm but also positions AFSA within the context of big data applications, highlighting its potential efficiency advantages.

Recursive Feature Elimination (RFE): The time complexity of RFE is influenced by the number of features n and the iterative process of feature elimination. Assuming a constant time complexity $O(1)$ for each model training and evaluation, RFE exhibits a worst-case time complexity of $O(n^2)$. This quadratic complexity arises from the cumulative training and evaluation steps across n iterations, indicative of a significant computational burden for large feature sets.

Forward Selection (FS): FS operates in a manner akin to RFE but in reverse, adding features incrementally. Under the same assumption of constant evaluation time, FS's time complexity similarly approximates to $O(n^2)$. This reflects the iterative addition and evaluation of features, rendering FS computationally intensive, especially as the number of features grows.

Genetic Algorithm (GA): The computational complexity of GA hinges on the population size p, the number of generations g, and the fitness evaluation process, predominantly the model training for each feature subset. Assuming a linear complexity for fitness evaluation, the overall complexity of GA can be expressed as $O(p \cdot g \cdot n)$, where n denotes the number of features. This suggests that GA's computational demand scales with the size of the feature set and the parameters governing the evolutionary process.

Adaptive Feature Selection Algorithm (AFSA): The Adaptive Feature Selection Algorithm (AFSA) employs a distinctive hybrid approach, initially leveraging multiple ranking methods for a comprehensive evaluation of features, followed by an adaptive, iterative selection process. The algorithm's complexity primarily stems from two components: the feature ranking and the selection phases. Given the absence of feature reduction or filtering in the initial ranking, and considering the use of multiple ranking methods, the complexity of this phase can be approximated as $O(n \log n)$, assuming the ranking algorithms employed operate in logarithmic time with respect to the number of features n. The subsequent selection phase, characterized by a single iteration over the ranked features without nested loops, implies a linear complexity, $O(n)$. Therefore, AFSA's overall complexity is more accurately described by $O(n \log n + n)$, which simplifies to $O(n \log n)$. This adjustment underscores AFSA's efficiency, particularly highlighting its capability to process large feature sets with a linearly scalable selection phase, thereby minimizing the computational overhead typically associated with feature selection in high-dimensional datasets.

Through this theoretical exploration, it becomes evident that while RFE and FS face scalability challenges due to their quadratic complexities, and GA's efficiency is contingent upon evolutionary parameters, AFSA's design inherently aims at reducing computational overhead. This analysis highlights AFSA's potential to deliver a balanced and efficient approach to feature selection in the realm of big data.

4.4.8 Comparative Analysis of AFSA

Unlike traditional methods, AFSA integrates a multifaceted ranking mechanism and an adaptive threshold-based selection strategy, distinguishing itself through the amalgamation of robustness and efficiency in feature selection.

AFSA's methodology significantly deviates from singular metric approaches like RFE and FS by employing a diverse array of feature ranking methods (Information Gain, Chi-Square Test, Mutual Information, Relief, and Gini Importance). This multiplicity ensures a comprehensive evaluation of feature relevance, mitigating the risk of bias toward specific data characteristics. Consequently, AFSA facilitates a more informed and versatile feature selection process, enhancing model performance across varied datasets.

Moreover, AFSA introduces an adaptive feature selection strategy that dynamically adjusts the inclusion threshold based on the incremental value each feature adds to model accuracy. This approach starkly contrasts with the static nature of FS and the exhaustive retraining in RFE, where features are added or eliminated without adjusting for the diminishing returns on model performance. By implementing a decreasing threshold for accuracy improvement, AFSA optimizes the selection process, effectively balancing between model accuracy and feature set compactness without necessitating extensive retraining at each iteration.

The computational complexity of AFSA, while initially seeming higher due to the utilization of multiple ranking methods, is offset by its strategic iterative selection process. The early termination criteria based on dynamic thresholding significantly reduce the number of model retraining instances, especially in big data contexts. This adaptive mechanism ensures that only features contributing meaningful improvements are considered, thereby minimizing computational overhead. In contrast, traditional methods like RFE may involve numerous retraining steps as they do not adjust selection criteria based on the feature's order or potential impact, leading to higher computational demands.

In summary, AFSA represents a significant advancement in feature selection methodologies by providing a balanced approach that not only offers competitive accuracy and efficient feature identification but also addresses the critical aspect of the computational burden. This enhanced comparative analysis underscores AFSA's theoretical and practical superiority, particularly in handling large-scale datasets, thereby affirming its contribution to the field of machine learning.

4.5 Cross Validation

Cross-validation stands as an indispensable technique in the assessment of machine learning models, playing a pivotal role in gauging their performance robustness and generalisability. This methodology involves the systematic partitioning of the dataset into training and validation sets, with a primary objective of ensuring that every data point is utilized for validation at some point during the evaluation process. This not only enhances the reliability of performance metrics but also guards against potential biases that may arise from a static train-test split. A widely adopted form of cross-validation is the k-fold cross-validation, a process that entails dividing the dataset into k subsets or folds. In each iteration of the cross-validation loop, the model is trained on k-1 folds and validated on the remaining one. This cyclic process is repeated k times, with each fold serving as the validation set exactly once. The results are then averaged across all iterations, providing a comprehensive and unbiased estimate of the model's generalization performance. In the context of this study, where the prediction of student performance serves as a focal point, the application of 5-fold cross-validation was integral. The choice of a 5-fold setup was deliberate, balancing the need for rigorous assessment with computational efficiency. By systematically rotating through different subsets of the dataset during the validation phase, the cross-validation strategy employed in this study contributes to the robustness of model evaluations. Additionally, the use of cross-validation acts as a safeguard against overfitting, ensuring that the models generalize well to unseen data and align with the overarching goals of accuracy and reliability in the prediction of student outcomes.

4.6 Classifiers

In this study, we adopted a selection of classifiers based on the comprehensive survey conducted by Xiao et al. [65] in the field of educational data mining (EDM). Xiao et al. provided a valuable list of widely used models in EDM research, laying the foundation for our model selection. However, the specific criteria for our selection were defined with our specific scenario in mind. Our primary considerations were the prevalence of models in EDM, their lightweight nature for iterative feature selection algorithms, and the exclusion of ensemble models for a fair comparison. The selected classifiers encompass a diverse set of models commonly employed in EDM research. Decision Trees (DT), K-Nearest Neighbours (KNN), Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machines (SVM) were chosen for their widespread usage and compatibility with our defined criteria. DT and NB offer interpretability and swift training and prediction speeds. KNN, known for its simplicity and noise insensitivity, was chosen with consideration for its suitability in smaller datasets. LR, a classical regression algorithm, provides fast training and ease of understanding. SVM, effective in classification, rounds out our selected models. The model selection process, informed by the survey findings of Xiao et al. [65], focused on models widely used in EDM research. The chosen classifiers meet our specific criteria of prevalence, lightweight nature, and exclusion of ensemble models. This strategic selection ensures that our comparative analysis provides meaningful insights into the predictive performance of individual classifiers in the context of student performance prediction using adaptive feature selection algorithms.

4.7 Evaluation

The assessment of the machine learning models employed in this study was conducted using a diverse set of performance metrics, including accuracy, F1-score, precision, recall, and the feature reduction factor (FRF), as detailed in Equation 4.1. These metrics collectively offered a comprehensive and nuanced evaluation of the model's efficacy in predicting student performance across various scenarios. The adoption of a 5-fold cross-validation approach ensured a robust and reliable assessment of model performance, mitigating the impact of data partitioning on the evaluation outcomes. Accuracy, a fundamental metric, provided an overarching measure of the models' correctness in predictions. The F1-score, precision, and recall offered insights into the models' ability to balance true positives, false positives, and false negatives, providing a more nuanced understanding of performance. The feature reduction factor (FRF) played a distinctive role in quantifying the impact of feature selection on model performance. As defined in Equation 4.1, the FRF represented the ratio of the total feature count to the selected feature count. This metric served as a valuable indicator of the efficiency of feature selection algorithms, offering insights into the degree of dimensionality reduction achieved.

$$\mathrm{FRF} = \frac{\text{Total feature count}}{\text{Selected feature count}} \tag{4.1}$$

The adoption of such a comprehensive set of metrics, coupled with the meticulous 5-fold cross-validation approach, contributed to a thorough evaluation of the machine learning models. The outcomes of this evaluation not only validated the models' predictive capabilities but also provided actionable insights for enhancing educational outcomes in the realm of educational data mining.

Chapter 5

Results Discussion

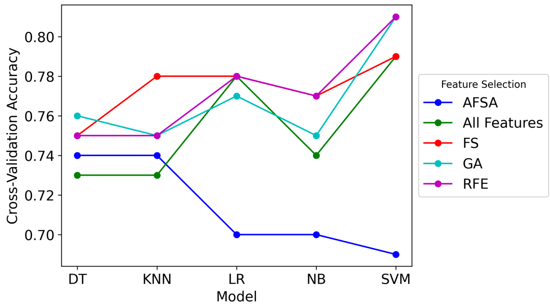
In this section, we present the outcomes of our experiments and engage in a comprehensive discussion of the results. Our focus is on evaluating the performance of the proposed Adaptive Feature Selection Algorithm (AFSA) in comparison to other state-of-the-art feature selection algorithms. We conducted experiments across five diverse machine-learning models and four distinct student performance prediction datasets.

5.1 Performance Metrics and Comparison

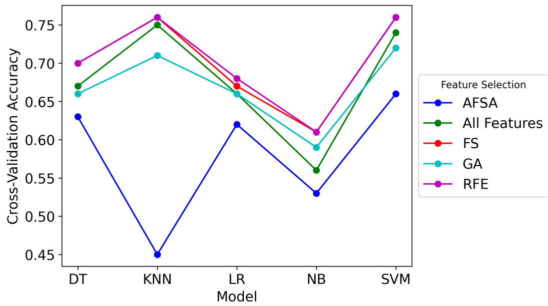
We assessed the effectiveness of various feature selection algorithms by employing cross-validation accuracy, F1 score, precision, recall, feature reduction factor, and training time as key metrics. The comparison included AFSA alongside other widely used feature selection methods. Figure 5.1 illustrates the cross-validation accuracy for each permutation, organized by feature selection algorithms, across different machine learning models and datasets. Notably, the forward selection (FS) algorithm consistently demonstrated a leading accuracy, with AFSA closely following suit. Other feature selection methods exhibited marginal differences in performance. Concerning the feature reduction factor, Figure 5.2 demonstrates AFSA's superiority, yielding an average reduction factor of 5.71. In contrast, the best-performing alternative, FS, achieved only 1.8. This underscores AFSA's effectiveness in streamlining datasets without compromising predictive accuracy. The training time analysis, presented in Figure 5.3, positions AFSA as the top-performing feature selection algorithm, trailing only behind the scenario where no feature selection is applied. AFSA exhibited an average training time of 1.24 seconds, outperforming other algorithms such as Recursive Feature Elimination (RFE), which required 5.81 seconds on average.

5.2 Dataset-Specific Performance

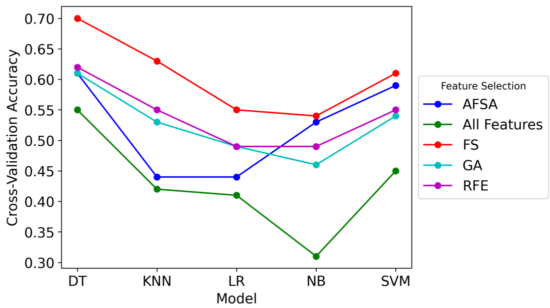
To delve deeper into the performance variations, we present specific evaluations for each dataset. Table 5.1 showcases the performance metrics for the XAPI dataset across five machine-learning models. Notably, the Genetic Algorithm (GA) excelled for Decision



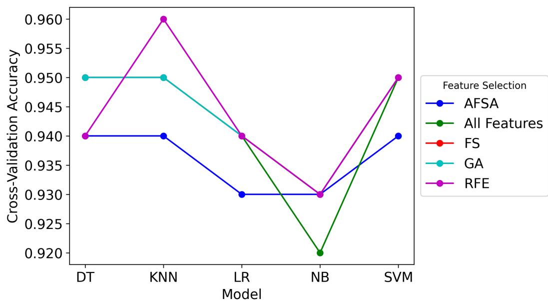
(a) Feature Selection with ML models on XAPI Dataset.



(b) Feature Selection with ML models on SSP Datasets

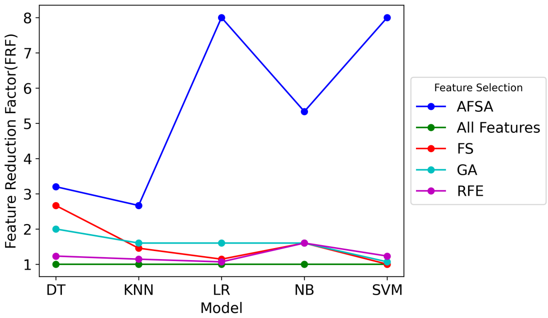


(c) Feature Selection with ML models on HESP Dataset

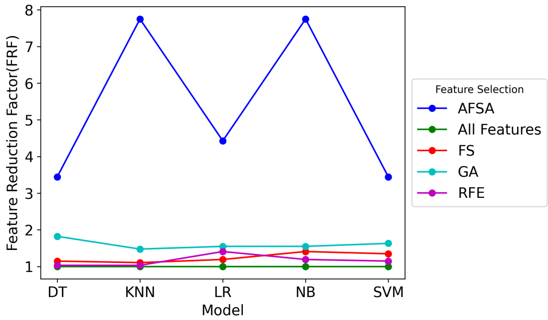


(d) Feature Selection with ML models on WOC2 dataset

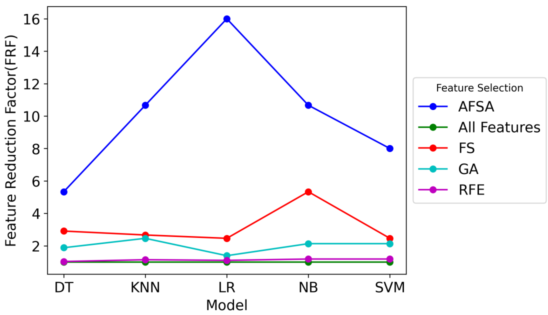
Figure 5.1: Cross-Validation Accuracy Grouped by Feature Selection with Five ML Models on Four Datasets.



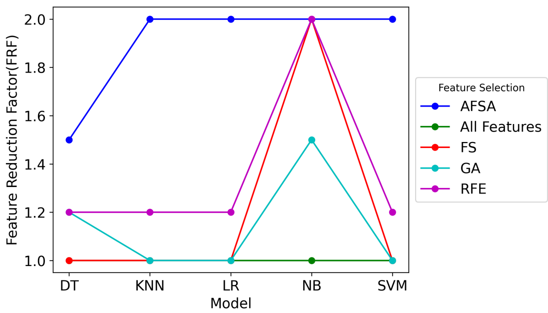
(a) Feature Selection Algorithms with Five models on XAPI Dataset.



(b) Feature Selection Algorithms with Five models on SSP Dataset.

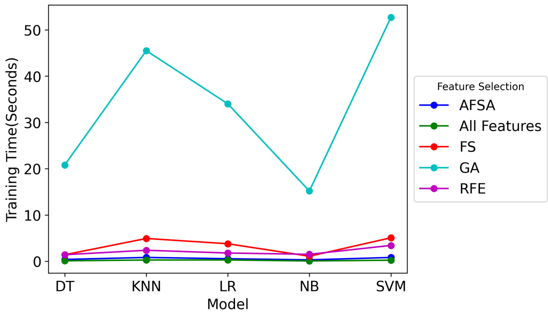


(c) Feature Selection Algorithms with Five models on HESP Dataset

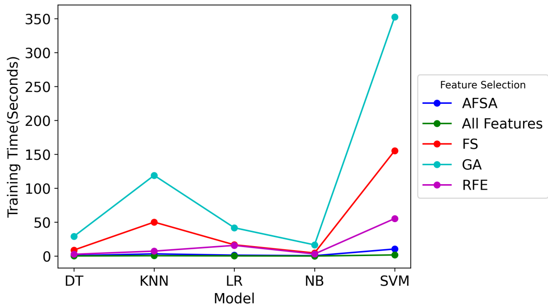


(d) Feature Selection Algorithms with Five models on WOC2 Dataset

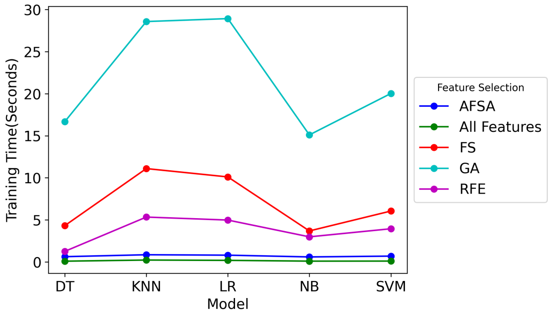
Figure 5.2: Feature Reduction Factor Grouped by Feature Selection Algorithms with Five ML Models and Four Datasets.



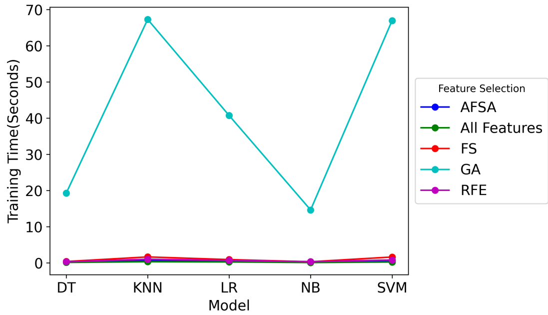
(a) Training time grouped by feature selection algorithms for 5 different machine learning models and dataset XAPI



(b) Training time grouped by feature selection algorithms for 5 different machine learning models and dataset SSP



(c) Training time grouped by feature selection algorithms for 5 different machine learning models and dataset HESP



(d) Training time grouped by feature selection algorithms for 5 different machine learning models and dataset WOC2

Figure 5.3: Training time grouped by feature selection algorithms for 5 different machine learning models and 4 datasets

Table 5.1: Evaluation Metrics on XAPI Dataset.

Model	Feature Selection	Accuracy	F1 Score	Precision	Recall	FRF	Training Time(s)
DT	AFSA	0.74	0.75	0.75	0.75	3.2	0.4
DT	All Features	0.73	0.73	0.74	0.74	1	0.11
DT	FS	0.75	0.75	0.76	0.75	2.67	1.41
DT	GA	0.76	0.76	0.76	0.76	2	20.83
DT	RFE	0.75	0.75	0.76	0.75	1.23	1.43
KNN	AFSA	0.74	0.75	0.75	0.75	2.67	0.84
KNN	All Features	0.73	0.72	0.73	0.74	1	0.28
KNN	FS	0.78	0.77	0.78	0.78	1.45	4.93
KNN	GA	0.75	0.74	0.75	0.75	1.6	45.51
KNN	RFE	0.75	0.74	0.74	0.75	1.14	2.38
LR	AFSA	0.7	0.69	0.71	0.7	8	0.56
LR	All Features	0.78	0.78	0.78	0.78	1	0.29
LR	FS	0.78	0.78	0.78	0.78	1.14	3.79
LR	GA	0.77	0.77	0.77	0.77	1.6	34.02
LR	RFE	0.78	0.78	0.78	0.78	1.07	1.79
NB	AFSA	0.7	0.69	0.71	0.71	5.33	0.31
NB	All Features	0.74	0.74	0.74	0.74	1	0.08
NB	FS	0.77	0.76	0.77	0.77	1.6	1.09
NB	GA	0.75	0.75	0.75	0.75	1.6	15.19
NB	RFE	0.77	0.77	0.77	0.78	1.6	1.51
SVM	AFSA	0.69	0.69	0.72	0.71	8	0.83
SVM	All Features	0.79	0.79	0.8	0.8	1	0.24
SVM	FS	0.79	0.79	0.8	0.8	1	5.08
SVM	GA	0.81	0.81	0.81	0.81	1.07	52.72
SVM	RFE	0.81	0.81	0.81	0.81	1.23	3.43

Trees, while Forward Selection (FS) proved optimal for k-Nearest Neighbours. Logistic Regression demonstrated consistent performance across feature selection methods. Support Vector Machines achieved peak accuracy with GA but demonstrated nuanced dependencies on feature selection techniques. Similar analyses were conducted for datasets WOC2, SSP, and HESP, presented in Tables 5.2, 5.3, and 5.4, respectively. These tables provide a comprehensive overview of each model's performance, emphasizing the influence of feature selection techniques on accuracy, precision, recall, and F1 Score, alongside varying training times.

5.3 Comparative Accuracy Analysis

In this part, we delve into a comprehensive comparison of the Adaptive Feature Selection Algorithm (AFSA) against various established feature selection methodologies across multiple models and datasets. Our objective is to elucidate the extent to which accuracy variations manifest, alongside examining the statistical significance of these differences.

5.3.1 Variation in Accuracy

The analysis reveals notable findings regarding the performance of AFSA in comparison to its counterparts:

- **Maximum Difference in Accuracy:** The apex of AFSA's performance superiority is marked by a maximum accuracy difference of +0.22, indicating scenarios where AFSA outperforms other methods by a margin of 22%.

Table 5.2: Evaluation Metrics on WOC2 Dataset.

Model	Feature Selection	Accuracy	F1 Score	Precision	Recall	FRF	Training Time(s)
DT	AFSA	0.94	0.94	0.94	0.94	1.5	0.2
DT	All Features	0.95	0.95	0.95	0.95	1	0.11
DT	FS	0.95	0.95	0.95	0.95	1	0.38
DT	GA	0.95	0.95	0.95	0.95	1.2	19.27
DT	RFE	0.94	0.94	0.94	0.94	1.2	0.32
KNN	AFSA	0.94	0.94	0.94	0.94	2	0.65
KNN	All Features	0.95	0.95	0.96	0.95	1	0.32
KNN	FS	0.95	0.95	0.96	0.95	1	1.65
KNN	GA	0.95	0.95	0.96	0.95	1	67.31
KNN	RFE	0.96	0.96	0.96	0.96	1.2	1.01
LR	AFSA	0.93	0.94	0.94	0.94	2	0.4
LR	All Features	0.94	0.94	0.94	0.94	1	0.25
LR	FS	0.94	0.94	0.94	0.94	1	0.91
LR	GA	0.94	0.94	0.94	0.94	1	40.74
LR	RFE	0.94	0.94	0.94	0.94	1.2	0.69
NB	AFSA	0.93	0.93	0.93	0.93	2	0.16
NB	All Features	0.92	0.92	0.92	0.92	1	0.09
NB	FS	0.93	0.93	0.93	0.93	2	0.32
NB	GA	0.93	0.93	0.93	0.93	1.5	14.64
NB	RFE	0.93	0.93	0.93	0.93	2	0.35
SVM	AFSA	0.94	0.94	0.94	0.94	2	0.49
SVM	All Features	0.95	0.95	0.95	0.95	1	0.24
SVM	FS	0.95	0.95	0.95	0.95	1	1.62
SVM	GA	0.95	0.95	0.95	0.95	1	66.98
SVM	RFE	0.95	0.95	0.95	0.95	1.2	0.8

Table 5.3: Evaluation Metrics on SSP Dataset.

Model	Feature Selection	Accuracy	F1 Score	Precision	Recall	FRF	Training Time(s)
DT	AFSA	0.63	0.63	0.64	0.64	3.44	0.74
DT	All Features	0.67	0.67	0.68	0.67	1	0.19
DT	FS	0.7	0.7	0.7	0.7	1.15	8.9
DT	GA	0.66	0.66	0.67	0.66	1.82	29.08
DT	RFE	0.7	0.7	0.7	0.7	1.03	2.64
KNN	AFSA	0.45	0.46	0.59	0.5	7.75	3.26
KNN	All Features	0.75	0.73	0.77	0.75	1	0.55
KNN	FS	0.76	0.75	0.78	0.76	1.11	50.09
KNN	GA	0.71	0.69	0.73	0.71	1.48	118.96
KNN	RFE	0.76	0.75	0.78	0.76	1.03	7.31
LR	AFSA	0.62	0.61	0.64	0.62	4.43	1.37
LR	All Features	0.66	0.65	0.68	0.66	1	0.28
LR	FS	0.67	0.67	0.69	0.67	1.19	16.62
LR	GA	0.66	0.66	0.68	0.66	1.55	41.72
LR	RFE	0.68	0.68	0.7	0.68	1.41	15.75
NB	AFSA	0.53	0.5	0.61	0.54	7.75	0.57
NB	All Features	0.56	0.54	0.68	0.57	1	0.09
NB	FS	0.61	0.6	0.63	0.61	1.41	4.56
NB	GA	0.59	0.58	0.64	0.59	1.55	16.52
NB	RFE	0.61	0.6	0.65	0.61	1.19	3.09
SVM	AFSA	0.66	0.65	0.67	0.66	3.44	10.42
SVM	All Features	0.74	0.74	0.77	0.74	1	1.7
SVM	FS	0.76	0.76	0.78	0.76	1.35	155.29
SVM	GA	0.72	0.72	0.74	0.72	1.63	352.4
SVM	RFE	0.76	0.76	0.78	0.76	1.15	55.27

Table 5.4: Evaluation Metrics on HESP Dataset.

Model	Feature Selection	Accuracy	F1 Score	Precision	Recall	FRF	Training Time(s)
DT	AFSA	0.61	0.64	0.65	0.64	5.33	0.63
DT	All Features	0.55	0.55	0.57	0.54	1	0.09
DT	FS	0.7	0.7	0.72	0.7	2.91	4.31
DT	GA	0.61	0.61	0.62	0.62	1.88	16.67
DT	RFE	0.62	0.62	0.63	0.62	1.03	1.26
KNN	AFSA	0.44	0.53	0.54	0.54	10.67	0.86
KNN	All Features	0.42	0.38	0.44	0.43	1	0.23
KNN	FS	0.63	0.63	0.67	0.64	2.67	11.1
KNN	GA	0.53	0.53	0.55	0.52	2.46	28.58
KNN	RFE	0.55	0.52	0.6	0.55	1.14	5.33
LR	AFSA	0.44	0.52	0.53	0.52	16	0.81
LR	All Features	0.41	0.4	0.42	0.41	1	0.19
LR	FS	0.55	0.53	0.54	0.55	2.46	10.11
LR	GA	0.49	0.49	0.5	0.49	1.39	28.93
LR	RFE	0.49	0.47	0.48	0.48	1.1	4.98
NB	AFSA	0.53	0.54	0.54	0.57	10.67	0.6
NB	All Features	0.31	0.28	0.34	0.31	1	0.1
NB	FS	0.54	0.55	0.59	0.54	5.33	3.69
NB	GA	0.46	0.44	0.43	0.46	2.13	15.1
NB	RFE	0.49	0.49	0.49	0.49	1.19	2.98
SVM	AFSA	0.59	0.63	0.65	0.64	8	0.69
SVM	All Features	0.45	0.42	0.44	0.44	1	0.1
SVM	FS	0.61	0.58	0.67	0.6	2.46	6.05
SVM	GA	0.54	0.49	0.49	0.52	2.13	20.03
SVM	RFE	0.55	0.52	0.54	0.54	1.19	3.95

- **Minimum Difference in Accuracy:** Conversely, the nadir of AFSA's performance is observed at a -0.31 difference, underscoring instances where AFSA lags behind alternative methods by 31% in terms of accuracy.
- **Standard Deviation:** The standard deviation of the accuracy differences between AFSA and the comparative algorithms stands at approximately 0.082, reflecting the variability and consistency of AFSA's performance across diverse scenarios.

5.3.2 Statistical Significance Analysis

A detailed statistical significance analysis, employing t-tests to compare the accuracy discrepancies between AFSA and each of the benchmarked feature selection algorithms, yielded the following p-values:

- **All Features (no selection):** A p-value of 0.83 suggests no statistically significant difference in performance when juxtaposing AFSA with the baseline approach of utilizing all features.
- **Feature Selection (FS):** With a p-value of 0.17, the accuracy differential between AFSA and FS does not reach statistical significance, albeit being relatively closer to the threshold of significance.
- **Genetic Algorithm (GA):** The p-value of 0.46 indicates the absence of a statistically significant accuracy difference between AFSA and GA.

- **Recursive Feature Elimination (RFE):** A p-value of 0.32 similarly denotes a lack of statistically significant difference in accuracy between AFSA and RFE.

The comprehensive comparative analysis underscores that while variations in accuracy between AFSA and other feature selection methodologies are observed, these variations generally do not translate to statistically significant differences. Such discrepancies could be attributed more to the specific characteristics inherent to the datasets or models rather than to an intrinsic superiority of one feature selection method over another. The absence of statistically significant differences, especially when compared to the conventional approach of employing all features, suggests that the choice of feature selection method should be informed by considerations beyond mere accuracy metrics; factors such as computational efficiency, model interpretability, and domain-specific requirements emerge as critical determinants in selecting an appropriate feature selection strategy.

5.4 Comparative Analysis of Training Times

The training time efficiency of the Adaptive Feature Selection Algorithm (AFSA) has been analyzed in comparison to other prevalent feature selection methods across a variety of models and datasets. We aim to evaluate the extent of variability in training times and assess the statistical significance of these differences.

5.4.1 Variability in Training Times

Upon comparing AFSA's training times with those of other algorithms, several key observations were made:

- **Maximum Difference in Training Time:** The most favorable scenario for AFSA showed it outpacing another method by approximately +8.71 seconds, indicating its potential for faster performance in optimal conditions.

- **Minimum Difference in Training Time:** Conversely, the most substantial lag observed for AFSA, compared to another method, amounted to -341.98 seconds, highlighting instances where AFSA's training time significantly exceeded that of its competitors.

- **Standard Deviation:** The calculated standard deviation of the training time differences between AFSA and other algorithms stood at 44.44 seconds, emphasizing the wide range of variability in training efficiency across different settings.

5.4.2 Statistical Significance of Training Time Differences

A closer statistical examination of the training time differences between AFSA and each benchmarked algorithm yielded the following p-values:

- **All Features (no selection):** A p-value of 0.074 suggests a marginally significant difference, with AFSA generally exhibiting slower training times than when all features are used, albeit this being close to the conventional significance threshold.
- **Feature Selection (FS):** The observed p-value of 0.104 indicates a lack of statistically significant difference in training times between AFSA and FS.
- **Genetic Algorithm (GA):** With a p-value of 0.007, there is a statistically significant difference favoring AFSA's training time efficiency over GA.
- **Recursive Feature Elimination (RFE):** A p-value of 0.113 signals no statistically significant difference in training times when comparing AFSA to RFE.

The comprehensive analysis underscores the significant variation in training times when AFSA is juxtaposed with other feature selection methodologies. Notably, AFSA demonstrates a statistically significant efficiency advantage over the Genetic Algorithm, marking it as a potentially time-efficient choice under certain conditions. However, the broad standard deviation across comparisons signals that both dataset characteristics and model selection play pivotal roles in influencing training time outcomes. Consequently, in the selection of a feature selection method, considerations must extend beyond mere accuracy implications to include potential training time efficiencies or constraints, especially in contexts where time is of the essence.

5.5 Distinguishing Feature Selection in Machine Learning and Deep Learning

In the realm of predictive modeling, it is essential to distinguish between the methodologies employed by traditional machine learning and deep learning frameworks, particularly concerning feature selection. Traditional machine learning models benefit substantially from explicit feature selection techniques, which aim to enhance model performance by reducing dimensionality, improving interpretability, and mitigating overfitting. In contrast, deep learning models, through their multi-layered architectures, inherently learn to identify and extract relevant features directly from raw data, bypassing the need for external feature selection processes. This inherent capability of deep learning models to autonomously learn complex representations makes them uniquely distinct from traditional approaches where feature selection plays a critical role. Our focus on refining feature selection for traditional machine learning models stems from this distinction, aiming to optimize model efficiency and performance in applications where explicit feature engineering is indispensable and where computational resources or data availability might limit the applicability of deep learning approaches.

Table 5.5: Grouped by Feature Selection and Summarised by the Average of the Evaluation Metric for Feature Selection Algorithms.

Feature Selection	Accuracy	F1 Score	Precision	Recall	FRF	Training Time(s)
AFSA	0.69	0.7	0.72	0.71	5.71	1.24
All Features	0.7	0.69	0.72	0.7	1	0.28
FS	0.76	0.75	0.77	0.76	1.8	14.6
GA	0.73	0.72	0.73	0.73	1.58	52.26
RFE	0.74	0.73	0.75	0.74	1.23	5.81

Table 5.6: Grouped by Model and Summarised by the Average of the Evaluation Metric for Models.

Model	Accuracy	F1 Score	Precision	Recall	FRF	Training Time (s)
DT	0.75	0.75	0.75	0.75	1.83	5.45
KNN	0.72	0.72	0.75	0.73	2.22	17.56
LR	0.71	0.71	0.72	0.71	2.53	10.21
NB	0.68	0.67	0.7	0.68	2.64	4.05
SVM	0.75	0.74	0.76	0.75	2.09	36.92

5.6 Summarized Insights

Table 5.5 consolidates the performance metrics across all datasets, highlighting the relative effectiveness of different feature selection algorithms. FS emerges as the top-performing method, while AFSA demonstrates competitive performance, particularly in terms of recall and feature reduction factors. Table 5.6 summarises the performance metrics for different machine learning models, showcasing Decision Trees as the top-performing model. Notably, Naive Bayes exhibits efficient training times despite slightly lower overall performance. Table 5.7 offers insights into dataset-specific challenges and opportunities. The WOC2 dataset stands out with the highest average accuracy, while SSP and XAPI demonstrate competitive performance with varying computational demands.

In conclusion, the experimental results underscore the efficacy of AFSA in feature selection for student performance prediction. While FS remains a top-performing method, AFSA provides a valuable alternative with significant feature reduction and efficient training times. The dataset-specific nuances highlight the importance of tailoring feature selection strategies to individual characteristics, offering valuable insights for practitioners in educational data analytics.

Table 5.7: Grouped by Dataset and Summarised by the Average of the Evaluation Metric for Datasets.

Dataset	**Accuracy**	**F1 Score**	**Precision**	**Recall**	**FRF**	**Training Time (s)**
HESP	0.52	0.52	0.55	0.53	3.49	6.69
SSP	0.66	0.66	0.7	0.67	2.07	35.89
WOC2	0.94	0.94	0.94	0.94	1.32	8.8
XAPI	0.76	0.75	0.76	0.76	2.17	7.96

Chapter 6

Conclusion & Future Work

6.1 Summary of Findings

In this study, we have introduced an Adaptive Feature Selection Algorithm (AFSA) for predicting student performance. The comprehensive evaluation, encompassing diverse machine learning models and student performance datasets, unveiled AFSA's superior performance compared to existing methods. Our study not only highlights AFSA's superiority in feature reduction but also its ability to uphold a competitive edge in predictive accuracy. This has been substantiated through rigorous evaluations.

6.1.1 Achievements of AFSA

Throughout our study, AFSA has proven itself as a powerful tool in educational data analytics by achieving an average feature reduction factor (FRF) of 5.71. This notable decrease in the number of features required for accurate predictions showcases AFSA's efficiency in processing large educational datasets, which is crucial for managing resources in educational institutions and improving computational performance.

6.1.2 Comparison with Existing Methods

The comprehensive comparative analysis undertaken in this study elucidates AFSA's enhanced capability to streamline complex datasets effectively. Despite the forward selection (FS) algorithm demonstrating high accuracy, AFSA remains competitive, offering a balanced approach with the additional advantage of significant feature reduction. This makes AFSA a preferable choice in scenarios where both dimensionality reduction and accuracy are paramount.

6.1.3 Implications for Educational Data Mining

AFSA's robustness and versatility suggest its applicability across a broad spectrum of educational settings. It can facilitate highly personalized learning paths and can be integral in designing effective educational interventions that are tailored to the needs of diverse student populations. This could revolutionize how educational data is utilized for continuous improvement in teaching strategies and student outcomes.

Our evaluations across various datasets highlight AFSA's consistent performance and reliability. This adaptability confirms its potential as a standard tool in educational analytics, capable of handling different types of educational environments and learning modalities.

6.2 Future Research Directions

The promising results from this initial study open up several pathways for further enhancements and explorations of AFSA.

6.2.1 Algorithm Enhancement

Integration of Feature Ranking Methods

Future iterations of AFSA could see the integration of a variety of feature ranking methods. This would potentially cater to different analytical needs and preferences, thereby broadening the algorithm's applicability and enhancing its precision.

Iterative Selection Strategy

Optimizing the iterative selection strategy could lead to a more refined and efficient algorithm. This could further reduce the computational demands of the feature selection process, enhancing AFSA's scalability to even larger datasets.

6.2.2 Hybrid Feature Selection Approaches

The exploration of hybrid approaches, combining AFSA with other robust feature selection methods, could yield a superlative model. This model would not only retain the strengths of individual approaches but also mitigate their weaknesses, leading to unprecedented performance in feature selection tasks.

6.2.3 Hyper-parameter Optimization

Delving into hyper-parameter tuning could significantly affect AFSA's performance. This exploration would aim to establish a set of optimal parameters that maximize the algorithm's efficiency and accuracy across various scenarios and datasets.

6.2.4 Understanding Model-Data Interactions

There is a critical need to examine the nuanced interactions between different feature selection techniques and the array of machine learning models. Such an investigation would aim to identify synergies that could lead to the development of more robust educational prediction models.

6.3 Broader Implications and Applications

6.3.1 Generalisability of AFSA

Testing AFSA on a broader scale, involving more diverse and extensive datasets, would provide deeper insights into its generalizability and effectiveness across different educational systems and cultures.

6.3.2 Feature Relationships and Student Outcomes

Further research into how specific features influence student outcomes could lead to more targeted and effective educational interventions. This could also assist policymakers and educators in understanding critical factors that drive student success.

6.4 Concluding Remarks

AFSA stands out as a significant innovation in the field of educational data mining. Its capability to efficiently reduce feature dimensionality while maintaining high predictive accuracy positions it as a valuable tool for data-driven decision-making in education. The broad implications of AFSA's application promise not only enhanced educational outcomes but also more efficient administrative strategies.

Acknowledgment

This research is funded by Institute for Advanced Research Publication Grant of United International University, Ref. No.: IAR-2024-Pub-015.

References

[1] Syeda Farjana Shetu, Mohd Saifuzzaman, Nazmun Nessa Moon, Sharmin Sultana, and Ridwanullah Yousuf. Student's performance prediction using data mining technique depending on overall academic status and environmental attributes. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2020, Volume 2*, pages 757–769. Springer, 2021.

[2] Han Xue and Yanmin Niu. Multi-output based hybrid integrated models for student performance prediction. *Applied Sciences*, 13(9):5384, 2023.

[3] Mohammad Sabri, Mohammad Zahid, Nazatul Aini Abd Majid, Siti Aishah Hanawi, Nur Izzati Mohd Talib, and Ariff Imran Anuar Yatim. Prediction model based on continuous data for student performance using principal component analysis and support vector machine. *TEM Journal*, 12(2), 2023.

[4] MRM VeeraManickam, M Mohanapriya, Bishwajeet K Pandey, Sushma Akhade, SA Kale, Reshma Patil, and M Vigneshwar. Map-reduce framework based cluster architecture for academic student's performance prediction using cumulative dragonfly based neural network. *Cluster Computing*, 22(Suppl 1):1259–1275, 2019.

[5] Xueliang Zhang, Jiawei Liu, Chi Zhang, Dongyan Shao, and Zhiqiang Cai. Innovation performance prediction of university student teams based on bayesian networks. *Sustainability*, 15(3):2335, 2023.

[6] Koushik Roy, Huu-Hoa Nguyen, and Dewan Md Farid. Impact of dimensionality reduction techniques on student performance prediction using machine learning. *CTU Journal of Innovation and Sustainable Development*, 15:93–101, October 2023.

[7] Koushik Roy, Huu-Hoa Nguyen, and Dewan Md. Farid. Impact of dimensionality reduction techniques on student performance prediction using machine learning. In *International Conference on Intelligent Systems and Data Science (ISDS-2023)*, pages 1–15, Can Tho University (CTU), Vietnam, November 2023.

[8] V Ganesh Karthikeyan, P Thangaraj, and S Karthik. Towards developing hybrid educational data mining model (hedm) for efficient and accurate student performance evaluation. *Soft Computing*, 24(24):18477–18487, 2020.

[9] Liana Maria Crivei, Gabriela Czibula, George Ciubotariu, and Mariana Dindelegan. Unsupervised learning based mining of academic data sets for students' performance analysis. In *2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pages 000011–000016. IEEE, 2020.

[10] Al Farissi, Halina Mohamed Dahlan, et al. Genetic algorithm based feature selection with ensemble methods for student academic performance prediction. In *journal of physics: Conference series*, volume 1500, page 012110. IOP Publishing, 2020.

[11] Kingsley Okoye, Arturo Arrona-Palacios, Claudia Camacho-Zuñiga, Joaquín Alejandro Guerra Achem, Jose Escamilla, and Samira Hosseini. Towards teaching analytics: a contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification. *Education and Information Technologies*, pages 1–43, 2022.

[12] ES Vinoth Kumar, S Appavu alias Balamurugan, and S Sasikala. Multi-tier student performance evaluation model (mtspem) with integrated classification techniques for educational decision making. *International Journal of Computational Intelligence Systems*, 14(1):1796–1808, 2021.

[13] Hamza Turabieh, Sana Al Azwari, Mahmoud Rokaya, Wael Alosaimi, Abdullah Alharbi, Wajdi Alhakami, and Mrim Alnfiai. Enhanced harris hawks optimization as a feature selection for the prediction of student performance. *Computing*, 103:1417–1438, 2021.

[14] Aya Nabil, Mohammed Seyam, and Ahmed Abou-Elfetouh. Prediction of students' academic performance based on courses' grades using deep neural networks. *IEEE Access*, 9:140731–140746, 2021.

[15] Mesfer Al Duhayyim, Radwa Marzouk, Fahd N Al-Wesabi, Maram Alrajhi, Manar Ahmed Hamza, and Abu Sarwar Zamani. An improved evolutionary algorithm for data mining and knowledge discovery. *CMC-COMPUTERS MATERIALS & CONTINUA*, 71(1):1233–1247, 2022.

[16] Lina Gao, Zhongying Zhao, Chao Li, Jianli Zhao, and Qingtian Zeng. Deep cognitive diagnosis model for predicting students' performance. *Future Generation Computer Systems*, 126:252–262, 2022.

[17] Guiyun Feng, Muwei Fan, and Yu Chen. Analysis and prediction of students' academic performance based on educational data mining. *IEEE Access*, 10:19558–19571, 2022.

[18] Khaledun Nahar, Boishakhe Islam Shova, Tahmina Ria, Humayara Binte Rashid, and AHM Saiful Islam. Mining educational data to predict students performance: A comparative study of data mining techniques. *Education and Information Technologies*, 26(5):6051–6067, 2021.

[19] Saud Altaf, Waseem Soomro, and Mohd Izani Mohamed Rawi. Student performance prediction using multi-layers artificial neural networks: A case study on educational data mining. In *Proceedings of the 2019 3rd International Conference on Information System and Data Mining*, pages 59–64, 2019.

[20] Steven Lehr, Hong Liu, Sean Kinglesmith, Alex Konyha, Natalia Robaszewska, and Jacob Medinilla. Use educational data mining to predict undergraduate retention. In *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, pages 428–430. IEEE, 2016.

[21] Juan L Rastrollo-Guerrero, Juan A Gómez-Pulido, and Arturo Durán-Domínguez. Analyzing and predicting students' performance by means of machine learning: A review. *Applied sciences*, 10(3):1042, 2020.

[22] Sumyea Helal, Jiuyong Li, Lin Liu, Esmaeil Ebrahimie, Shane Dawson, Duncan J Murray, and Qi Long. Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*, 161:134–146, 2018.

[23] Otgontsetseg Sukhbaatar, Kohichi Ogata, and Tsuyoshi Usagawa. Mining educational data to predict academic dropouts: a case study in blended learning course. In *TENCON 2018-2018 IEEE region 10 conference*, pages 2205–2208. IEEE, 2018.

[24] Drishty Sobnath, Tobiasz Kaduk, Ikram Ur Rehman, and Olufemi Isiaq. Feature selection for uk disabled students' engagement post higher education: a machine learning approach for a predictive employment model. *IEEE Access*, 8:159530–159541, 2020.

[25] Aimad Qazdar, Brahim Er-Raha, Chihab Cherkaoui, and Driss Mammass. A machine learning algorithm framework for predicting students performance: A case study of baccalaureate students in morocco. *Education and Information Technologies*, 24:3577–3589, 2019.

[26] Hanan Abdullah Mengash. Using data mining techniques to predict student performance to support decision making in university admission systems. *Ieee Access*, 8:55462–55470, 2020.

[27] Deepti Aggarwal, Sonu Mittal, and Vikram Bali. Significance of non-academic parameters for predicting student performance using ensemble learning techniques. *International Journal of System Dynamics Applications (IJSDA)*, 10(3):38–49, 2021.

[28] Mohammad Hasan and Mohamed Aly. Get more from less: a hybrid machine learning framework for improving early predictions in stem education. In *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 826–831. IEEE, 2019.

[29] Ammar Almasri, Rami S Alkhawaldeh, and Erbuğ Çelebi. Clustering-based emt model for predicting student performance. *Arabian Journal for Science and Engineering*, 45:10067–10078, 2020.

[30] Joy Dhar and Asoke Kumar Jodder. An effective recommendation system to forecast the best educational program using machine learning classification algorithms. *Ingénierie des Systèmes d Inf.*, 25(5):559–568, 2020.

[31] Elaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah. Mining educational data to predict student's academic performance using ensemble methods. *International journal of database theory and application*, 9(8):119–136, 2016.

[32] Juan David Garcia and Anastasija Skrita. Predicting academic performance based on students' family environment: Evidence for colombia using classification trees. *Psychology, Society & Education*, 11(3):299–311, 2019.

[33] Rafaella Leandra Souza do Nascimento, Ricardo Batista das Neves Junior, Manoel Alves de Almeida Neto, and Roberta Andrade de Araújo Fagundes. Educational data mining: An application of regressors in predicting school dropout. In *Machine Learning and Data Mining in Pattern Recognition: 14th International Conference, MLDM 2018, New York, NY, USA, July 15-19, 2018, Proceedings, Part II 14*, pages 246–257. Springer, 2018.

[34] Bo Wu, Shaojie Qu, Yin Ni, Yemin Zhou, Pengxiang Wang, and Qiwen Li. Predicting student performance using weblogs. In *2019 14th International Conference on Computer Science & Education (ICCSE)*, pages 616–621. IEEE, 2019.

[35] Bashir Khan Yousafzai, Maqsood Hayat, and Sher Afzal. Application of machine learning and data mining in predicting the performance of intermediate and secondary education level student. *Education and Information Technologies*, 25:4677–4697, 2020.

[36] Hamza Turabieh. Hybrid machine learning classifiers to predict student performance. In *2019 2nd international conference on new trends in computing sciences (ICTCS)*, pages 1–6. IEEE, 2019.

[37] Muhammad Wafi, Umar Faruq, and Ahmad Afif Supianto. Automatic feature selection for modified k-nearest neighbor to predict student's academic performance. In *2019 International Conference on Sustainable Information Engineering and Technology (SIET)*, pages 44–48. IEEE, 2019.

[38] Moke Xu, Yu Liang, and Wenjun Wu. Predicting honors student performance using rbfnn and pca method. In *Database Systems for Advanced Applications: DASFAA 2017 International Workshops: BDMS, BDQM, SeCoP, and DMMOOC, Suzhou, China, March 27-30, 2017, Proceedings 22*, pages 364–375. Springer, 2017.

[39] Charoula Angeli, Sarah K Howard, Jun Ma, Jie Yang, and Paul A Kirschner. Data mining in educational technology classroom research: Can it make a contribution? *Computers & Education*, 113:226–242, 2017.

[40] Javier Bravo-Agapito, Claire Frances Bonilla, and Isaac Seoane. Data mining in foreign language learning. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(1):e1287, 2020.

[41] Paul Joseph M Estrera, Pamela E Natan, Babe Grrece T Rivera, and Faith B Colarte. Student performance analysis for academic ranking using decision tree approach in university of science and technology of southern philippines senior high school abstract. *International Journal of Engineering and Technology*, 3(5):147–153, 2017.

[42] M Ramaswami and R Bhaskaran. A study on feature selection techniques in educational data mining. *arXiv preprint arXiv:0912.3924*, 2009.

[43] January D Febro. Utilizing feature selection in identifying predicting factors of student retention. *International Journal of Advanced Computer Science and Applications*, 10(9), 2019.

[44] Maryam Zaffar, Manzoor Ahmed Hashmani, KS Savita, and Syed Sajjad Hussain Rizvi. A study of feature selection algorithms for predicting students academic performance. *International Journal of Advanced Computer Science and Applications*, 9(5), 2018.

[45] Isabelle Guyon, Steve Gunn, Masoud Nikravesh, and Lofti A Zadeh. *Feature extraction: foundations and applications*, volume 207. Springer, 2008.

[46] Beatriz Remeseiro and Veronica Bolon-Canedo. A review of feature selection methods in medical applications. *Computers in biology and medicine*, 112:103375, 2019.

[47] Inaki Inza, Pedro Larranaga, Rosa Blanco, and Antonio J Cerrolaza. Filter versus wrapper gene selection approaches in dna microarray domains. *Artificial intelligence in medicine*, 31(2):91–103, 2004.

[48] Yap Bee Wah, Nurain Ibrahim, Hamzah Abdul Hamid, Shuzlina Abdul-Rahman, and Simon Fong. Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy. *Pertanika Journal of Science & Technology*, 26(1), 2018.

[49] Ron Kohavi and George H John. Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2):273–324, 1997.

[50] Sebastian Okser, Tapio Pahikkala, Antti Airola, Tapio Salakoski, Samuli Ripatti, and Tero Aittokallio. Regularized machine learning in the genetic prediction of complex traits. *PLoS genetics*, 10(11):e1004754, 2014.

[51] Makiko Yoshida and Asako Koike. Snpinterforest: a new method for detecting epistatic interactions. *BMC bioinformatics*, 12:1–10, 2011.

[52] Zhi Wei, Wei Wang, Jonathan Bradfield, Jin Li, Christopher Cardinale, Edward Frackelton, Cecilia Kim, Frank Mentch, Kristel Van Steen, Peter M Visscher, et al. Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease. *The American Journal of Human Genetics*, 92(6):1008–1012, 2013.

[53] Raid Alzubi, Naeem Ramzan, Hadeel Alzoubi, and Abbes Amira. A hybrid feature selection method for complex diseases snps. *IEEE Access*, 6:1292–1301, 2017.

[54] Li Ma, Alon Keinan, and Andrew G Clark. Biological knowledge-driven analysis of epistasis in human gwas with application to lipid traits. *Epistasis: Methods and Protocols*, pages 35–45, 2015.

[55] Gina M D'Angelo, D Chandrasekhra Rao, and C Charles Gu. Combining least absolute shrinkage and selection operator (lasso) and principal-components analysis for detection of gene-gene interactions in genome-wide association studies. In *BMC proceedings*, volume 3, pages 1–5. BioMed Central, 2009.

[56] Nicholas Pudjihartono, Tayaza Fadason, Andreas W Kempa-Liehr, and Justin M O'Sullivan. A review of feature selection methods for machine learning-based disease risk prediction. *Frontiers in Bioinformatics*, 2:927312, 2022.

[57] Thierry Schüpbach, Ioannis Xenarios, Sven Bergmann, and Karen Kapur. Fastepistasis: a high performance computing solution for quantitative trait epistasis. *Bioinformatics*, 26(11):1468–1469, 2010.

[58] Silke Szymczak, Joanna M Biernacka, Heather J Cordell, Oscar González-Recio, Inke R König, Heping Zhang, and Yan V Sun. Machine learning in genome-wide association studies. *Genetic epidemiology*, 33(S1):S51–S57, 2009.

[59] Casey S Greene, Daniel S Himmelstein, Jeff Kiralis, and Jason H Moore. The informative extremes: using both nearest and farthest individuals can improve relief algorithms in the domain of human genetics. In *European conference on evolutionary computation, machine learning and data mining in bioinformatics*, pages 182–193. Springer, 2010.

[60] Shouheng Tuo, Haiyan Liu, and Hao Chen. Multipopulation harmony search algorithm for the detection of high-order snp interactions. *Bioinformatics*, 36(16):4389–4398, 2020.

[61] Paulo Cortez. Student Performance. UCI Machine Learning Repository, 2014. DOI: https://doi.org/10.24432/C5TG7T.

[62] Nevriye Yılmaz and Boran Sekeroglu. Student performance classification using artificial intelligence techniques. In *International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions*, pages 596–603. Springer, 2019.

[63] MohammadNoor Injadat, Abdallah Moubayed, Ali Bou Nassif, and Abdallah Shami. Multi-split optimized bagging ensemble model selection for multi-class educational data mining. *Applied Intelligence*, 50:4506–4528, 2020.

[64] Haibo He, Yang Bai, Edwardo A Garcia, and Shutao Li. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, pages 1322–1328. Ieee, 2008.

[65] Wen Xiao, Ping Ji, and Juan Hu. A survey on educational data mining methods used for predicting students' performance. *Engineering Reports*, 4(5):e12482, 2022.