10.31341/jios.48.1.7
Open Access
UDC 004.021:004.65
Original Scientific Paper

A Study on Comparative Analysis of Feature Selection Algorithms for Students Grades Prediction

Muhammad Arham Tariq *arhamtariq99@gmail.com*
Department of Computer Science
University of Central Punjab, Lahore, Pakistan

Abstract

Education data mining (EDM) applies data mining techniques to extract insights from educational data, enabling educators to evaluate their teaching methods and improve student outcomes. Feature selection algorithms play a crucial role in improving classifier accuracy by reducing redundant features. However, a detailed and diverse comparative analysis of feature selection algorithms on multiclass educational datasets is missing. This paper presents a study that compares ten different feature selection algorithms for predicting student grades. The goal is to identify the most effective feature selection technique for multi-class student grades prediction. Five classifiers, including Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), Gradient Boosting (GB), and k-Nearest Neighbors (KNN), are trained and tested on ten different feature selection algorithms. The results show that SelectFwe(SFWE-F) performed best, achieving an accuracy of 74.3% with Random Forests (RT) across all ten feature selection algorithms. This algorithm selects features based on their relationship with the target variable while controlling the family-wise error rate.
Keywords: Classification Models, Educational Data Mining, Feature Selection, Multi-Class Datasets, Student Performance

1. Introduction

To The field of education has witnessed a surge in the availability of large volumes of student data in recent years [1]. This has led to a growing interest in the use of data mining techniques to extract meaningful insights and improve educational outcomes. Educational data mining (EDM) involves the use of these techniques to analyze student data and identify patterns that can inform instructional practices and improve student outcomes. Feature selection algorithms play a critical role in this process by reducing redundant features and increasing classifier accuracy. Feature selection is an essential step in educational data mining as it helps to identify the most relevant and informative features from a large set of variables. It enables researchers to improve the accuracy of predictive models and identify important factors that influence student learning outcomes. By selecting the most relevant features, educational institutions can also optimize their resources and tailor their interventions to specific student needs.

However, a detailed and diversified comparative analysis of feature selection algorithms for multi-class student grades prediction has been lacking. The main contribution of the research is a detailed and diversified comparative analysis of ten feature selection algorithms on multi-class educational datasets for predicting student grades, which had previously been missing in the literature. This paper aims to fill this gap by presenting a comparative analysis of ten different feature selection algorithms. The goal is to identify the most effective feature selection technique for predicting student grades. To evaluate the performance of these algorithms, five different classifiers, including Support Vector Machines (SVM) and Decision Trees (DT), are trained and tested on the ten feature selection algorithms.

The ten feature selection algorithms used are SelectKBest (SKBF), SelectKBest (SKBC), SelectKBest (SKBM), SelectPercentile (SPF), SelectFpr (SFPRF), SelectFwe (SFWEF), SelectFdr (SFDRF), VarianceThreshold (VT), GenericUnivariateSelect (GUSF), and GenericUnivariateSelect (GUS-M). Each algorithm is evaluated with five different machine learning models using a pipeline and 10K-fold cross-validation.

2. Literature Review

In this section, a review of relevant studies and literature that have investigated the prediction of academic performance using various machine learning algorithms and feature selection techniques is presented. These studies shed light on different approaches and methodologies employed in the field and provide valuable insights into the performance of these methods.

Jalota, C., & Agrawal, R. 2021[2] conducted a study on comparative analysis of correlation feature selection and wrapper-based feature selection on educational datasets. The results of the research claims that J48 algorithm gain high accuracy by combining with correlation feature selection. Kumar et al.,2022 [3] utilized five different classification algorithms combined with three different feature selection algorithms to build a model for student academic performance prediction. Feature selection algorithms utilized in the approach were Correlation Attribute Evaluator, Information Gain Attribute, Gain Ratio. The highest accuracy achieved was 83.33% with the help of Decision tree and Gain ratio. Agrusti, F, 2020 [4] presents a deep learning-based approach for predicting university dropout, with a case study conducted at Roma Tre University. The approach uses a long short-term memory (LSTM) neural network to model the temporal dependencies of students' academic performance and predict their likelihood of dropping out, achieving promising results. Acharya, A., & Sinha, D,2014 [5] utilized student academic performance dataset having 309 instances and 14 features. Main objective of the research conducted was to find best feature selection algorithm among three different types. The types were Filter based, Wrapper Based learning and Correlation based Algorithm. The results claimed Correlation Based Feature Selection provides best result. Nidhi, Kumar, &Agarwal, 2021 [6] presents a comparative analysis of different heterogeneous ensemble learning algorithms using feature selection techniques for predicting academic performance of students. The study evaluates the performance of six different algorithms on a dataset of student performance indicators, comparing their accuracy, precision, recall, and F1 score. The results show that the Random Forest algorithm with feature selection outperformed the other algorithms in terms of accuracy and other evaluation metrics.Anuradha,& Velmurugan,2015 [7] conducted a study that uses a dataset of student information, including demographics, socio-economic status, and past academic performance, to train and test several algorithms, including Decision Tree, Naive Bayes, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). The paper evaluates the algorithms' performance in terms of accuracy, sensitivity, and specificity. The results show that SVM outperforms the other methods in terms of overall performance, with an accuracy of 89.2%, compared to Decision Tree (87.1%), Naive Bayes (84.3%), and KNN (82.7%).

Enaro, A. O., & Chakraborty, S,2020 [8] used a dataset of student information, including demographic, behavioral, and academic features, to evaluate the performance of four feature selection algorithms: Correlation-based Feature Selection (CFS), Information Gain (IG), ReliefF, and Chi-Square. The paper evaluates the algorithms' performance in terms of classification accuracy using three different classifiers: J48 Decision Tree, Naive Bayes, and Multilayer Perceptron (MLP) Neural Network. The results show that CFS outperforms the other algorithms in terms of classification accuracy across all three classifiers, followed by ReliefF and IG, while Chi-Square performs the worst. Bakker, T el al.,2023 [9] proposed a study that s uses a dataset of 44 autistic students and 87 non-autistic students, including demographic and academic performance features, to compare the performance of three machine learning algorithms: Decision Tree (DT), Artificial Neural Network (ANN), and Support Vector Machine (SVM). The results indicate that the SVM algorithm outperformed the other two algorithms in predicting academic success, achieving an accuracy of 89%. Hamdi et al,2022 [10] new feature selection method based on the Chicken Swarm Optimization algorithm (CSO) combined with machine learning techniques for predicting student academic performance. The authors use a dataset consisting of demographic and academic performance features of 470 students to evaluate the performance of the proposed method. The authors compare the performance of the proposed CSO-based feature selection method with four other feature selection methods, including Random Forest, Correlation-based Feature Selection (CFS), Chi-square (Chi2), and Information Gain (IG). The authors evaluate the algorithms' performance in terms of accuracy, precision, recall, and F1-score. The results show that the proposed CSO-based feature selection method outperforms the other four methods in terms of accuracy, achieving an accuracy of 84.89%. The study also identifies the most significant predictors of student performance, including demographic information and academic achievement. The findings of this study suggest that the proposed CSO-based feature selection method can be a useful tool for feature selection in predicting student performance and can contribute to the development of targeted interventions to improve academic outcomes in higher education.

Huynh-Cam et al,2022 [11] The study aimed to predict the academic performance of international students, students with disabilities, and local students using machine learning algorithms based on their admission profiles and first-semester grades.

Results showed that SVM was the best model for predicting academic performance of students with disabilities, while RF was best for local students. The most important features were numbers of required and elective credits, source of living expenses, and parental occupation and income. This study can help institutions take early measures to improve the academic performance of students and attract more international students. Najm et al.2022 [12] presents a roadmap for implementing an educational data mart based on historical data from Alexandria Private Elementary School in Iraq. The data mart is constructed and an OLAP cube is used for OLAP operations and reports. Nine algorithms are used for OLAP mining and clustering with expectation maximization is found to have the highest accuracy (96.76%) for predicting student performance and grades. This study can help academic institutions make informed decisions based on historical data.

3. Background

3.1. SelectKBest

The SelectKBest [13] algorithm in scikit-learn, which is used for feature selection in machine learning. However, they differ in the statistical tests that they use for selecting the top K features from a dataset. Here's a brief explanation of each algorithm:

3.1.1. SelectKBest with chi-squared test (SKBF)

This algorithm [14] uses the chi-squared test to evaluate the independence of each feature and the target variable. It selects the top K features with the highest chi-squared scores, indicating the strongest relationship with the target variable.

3.1.2. SelectKBest with ANOVA F-test (SKBC)

This algorithm uses the ANOVA F-test [15] to evaluate the difference in means between groups of samples. It selects the top K features with the highest F-scores, indicating the greatest difference in means between the groups.

3.1.3. SelectKBest with mutual information (SKBM)

This algorithm uses mutual information to evaluate the dependence between each feature and the target variable. It selects the top K features with the highest mutual information scores, indicating the strongest dependence with the target variable.

The main difference between these algorithms is the statistical test that they use for feature selection. The chi-squared test is useful for categorical data, the ANOVA F-test is useful for continuous data with categorical targets, and mutual information is useful for any type of data.

In summary, SelectKBest with chi-squared test (SKBF), SelectKBest with ANOVA F-test (SKBC), and SelectKBest with mutual information (SKBM) are all variations of the SelectKBest algorithm that use different statistical tests for feature selection. The choice of algorithm depends on the type of data and the problem being solved.

3.2. SelectPercentile

SelectPercentile [16] is a feature selection technique in machine learning that, like SelectKBest, is used to select the top features from a dataset. However, instead of selecting a fixed number of features, SelectPercentile selects a specified percentage of the total number of features.

The SelectPercentile algorithm is based on univariate statistical tests, similar to SelectKBest. It evaluates the relationship between each feature and the target variable and selects the top features based on a specified statistical test. The difference is that instead of selecting a fixed number of features, SelectPercentile selects the top features based on a specified percentage.

The SelectPercentile algorithm works by first computing a score for each feature using the specified statistical test. It then selects the top percentage of features with the highest scores. For example, if a user specifies a percentile of 10%, SelectPercentile will select the top 10% of features with the highest scores.

3.3. SelectFpr

SelectFpr [17] is a feature selection technique in machine learning that is used to select features based on a specified false positive rate. It is a part of the Select family of feature selection methods in scikit-learn library in Python.

The SelectFpr algorithm works by selecting the features that have a false positive rate lower than the specified threshold. It uses a statistical test, such as the chi-squared test or ANOVA F-test, to evaluate the relationship between each feature and the target variable, and then selects the features that meet the specified false positive rate threshold.

3.4. SelectFwe

SelectFwe[18] is a feature selection technique in machine learning that is used to select features based on a specified family-wise error rate. It is a part of the Select family of feature selection methods in scikit-learn library in Python.

The family-wise error rate (FWER) is a statistical measure that indicates the probability of making at least one false discovery among all the discoveries made by a model. In the context of feature selection, it is the probability of selecting at least one irrelevant feature.

3.5. SelectFdr

SelectFdr[19] is a feature selection technique in machine learning that is used to select features based on a specified false discovery rate. It is a part of the Select family of feature selection methods in scikit-learn library in Python.

The false discovery rate (FDR) is a statistical measure that indicates the proportion of false discoveries among all the discoveries made by a model. In the context of feature selection, it is the proportion of irrelevant features selected by the model.

The SelectFdr algorithm works by selecting the features that have a false discovery rate lower than the specified threshold. It uses a statistical test, such as the chi-squared test or ANOVA F-test, to evaluate the relationship between each feature and the target variable, and then selects the features that meet the specified false discovery rate threshold.

3.6. Variance Threshold

Variance Threshold [20] is a feature selection technique used in machine learning to remove features with low variance from a dataset. Variance is a measure of how much a feature's values vary from the mean value. Features with low variance may not be useful in predicting the output and can be removed to simplify the model and reduce overfitting.

The idea behind Variance Threshold is that if a feature's variance is below a certain threshold, it is likely that the feature has almost constant values across all samples and will not provide much information to the model. Therefore, it is safe to remove such features.

3.7. GenericUnivariateSelect

GenericUnivariateSelect [21] is a feature selection algorithm provided by the scikit-learn library in Python. It is a type of univariate feature selection method, which means that it evaluates the importance of each feature independently and selects the best ones based on a statistical test or a score function.

The algorithm takes three parameters:

- **score_func:** This is a function that is used to calculate the score of each feature. The available score functions in scikit-learn include ANOVA F-value, mutual information, chi2, and others.
- **mode:** This parameter determines how the features are selected based on their scores. It can be set to 'k_best' to select the top k features with the highest scores, 'percentile' to select the features above a certain percentile of the score distribution, or 'false_discovery_rate' to select the features with the lowest false discovery rate.
- **param:** This parameter is used to specify the number of features to select in the case of 'k_best' mode, the percentile of features to select in the case of

'percentile' mode, or the target false discovery rate in the case of 'false_discovery_rate,mode.

The GUS-F and GUS-M are two variants of the GenericUnivariateSelect class, which use different methods for selecting the features. In summary, GUS-F selects features based on their scores in univariate statistical tests, while GUS-M selects features based on a mutual information score between each feature and the target variable. The choice between these two variants depends on the nature of the data and the research question being addressed.

4. Methods

In this section, detailed description of dataset, research methodology, data preparation steps, and the evaluation process conducted to analyze feature selection algorithms for student academic performance are added.

4.1. Dataset

Hamtini, T.,2015 [22] This dataset was acquired from the Kalboard 360 learning management system (LMS), which grants users access to educational materials as long as they are connected to the internet. Additionally, this system monitors students' progress, including their interactions with educational content, such as how often they read or view it. With the help of this data, our objective is to uncover the factors associated with students' academic success. The dataset pertains to a group of students who took part in an educational program. This dataset comprises 480 data points and 16 attributes. Each data point corresponds to a specific student, while each attribute relates to a particular characteristic or trait of that student. It includes three categories denoting Low-Level, Middle-Level, and High-Level student performance. The dataset incorporates demographic data such as gender, nationality, and class section, as well as academic performance indicators, including grades in various subjects, attendance, and quiz/exam scores. Furthermore, it encompasses data related to the student's learning styles, including their preferred method of learning, approach to learning, and motivation levels. This dataset is valuable in investigating correlations between academic performance and various demographic and learning style aspects. It can also assist in predicting student performance or highlighting areas that require educational interventions. Fig. 1 depicts the distribution of the three categories across the dataset.

4.2. Data Pre-Processing

Yu, N,2018 [23] Data Pre-processing is one of initial major step to be performed in the machine learning process. It can contain many phases. Because the dataset lacks anomalies, noise, or missing values, these steps were disregarded. Apart from that feature encoding was adopted. Feature encoding is the process of converting categorical variables in a dataset into numerical values that can be used in statistical models or machine learning algorithms.

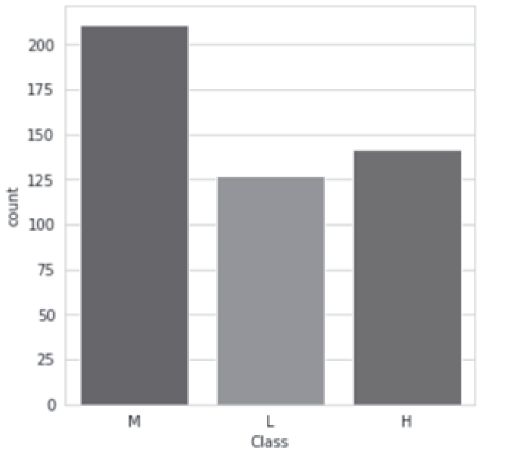

Figure 1. Dataset Classes Distribution

4.3. Features Selection

In this phase, ten different feature selection algorithms SelectKBest(SKBF), SelectKBest(SKBC),SelectKBest(SKBM),SelectPercentile(SPF),SelectFpr(SFPRF), SelectFwe(SFWEF),SelectFdr(SFDRF),VarianceThreshold(VT),GenericUnivariateSelect(GUSF),GenericUnivariateSelect(GUS-M) are applied. All of the feature selection algorithms are applied and evaluated differently with the help of different classification algorithms.

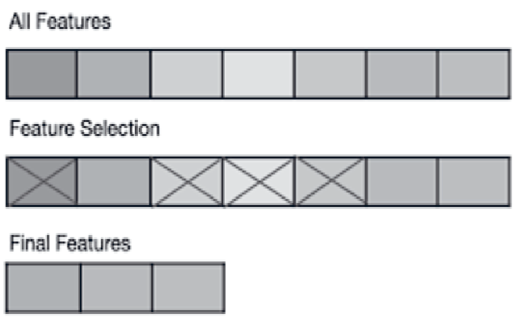

Figure 2. Working Demonstration of Feature Selection Algorithm

4.4. Hyperparameter Tunning

Hyperparameter tuning is a crucial step in the machine learning pipeline and can greatly impact the accuracy and generalization of the resulting model. In this phase of the configured approach four different classification models named as SVM, DT, RT and GB are applied and tunned with the help of the parameters mentioned in the Table 1. Fig 3 represent complete working model of hyperparameter tunning.

Classifier Name	Parameters
KNN	'params': {'n_neighbors': [3, 5, 7]}
DT	'params': {'kernel': ['linear', 'rbf'], 'C': [1, 10]
RT	{ params': {'n_estimators': [50, 100], 'max_depth': [5, 10], 'min_samples_split': [2, 5], 'min_samples_leaf': [1, 2]}
GB	{'n_estimators': [50, 100], 'learning_rate': [0.01, 0.1]}

Table 1. Classifiers Parameter

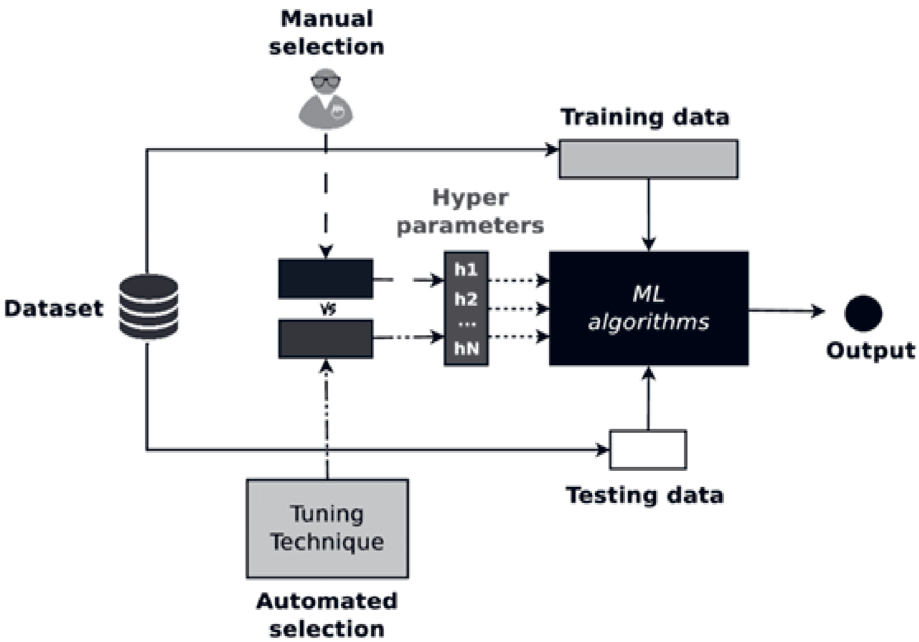

Figure 3. Working Demonstration of Hyperparameter Tunning [24]

4.5. Model Evaluation

Different evaluation metrics are applied in order to check the overall performance of the machine learning classifiers utilized in the configured approach.

The evaluation metrics adopted can be termed as:

Accuracy:

The formula for accuracy is:

Accuracy = (Number of Correct Predictions) / (Total Number of Predictions)

Precision:

The formula for precision is:

Precision = (Number of True Positive Predictions) / (Number of True Positive Predictions + Number of False Positive Predictions

Recall:

The formula for recall is:

Recall = (Number of True Positive Predictions) / (Number of True Positive Predictions + Number of False Negative Predictions)

F1-Score:

The formula for F1 score is:

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

5. Results and Discussion

In this Section detailed results of the classifiers combined with different feature selection algorithms will be explained. Fig. 4 & 5 represents the mean accuracies and F1-scores of the classifiers. On the X axis there are 5 different classifiers and Y-axis represents 10 different feature selection algorithms. The results are generated using 10 K fold cross validation. Table 2-4 represents the classifiers Recall, Precision and F1-Score. Table 5 comprises of précised mean accuracies.

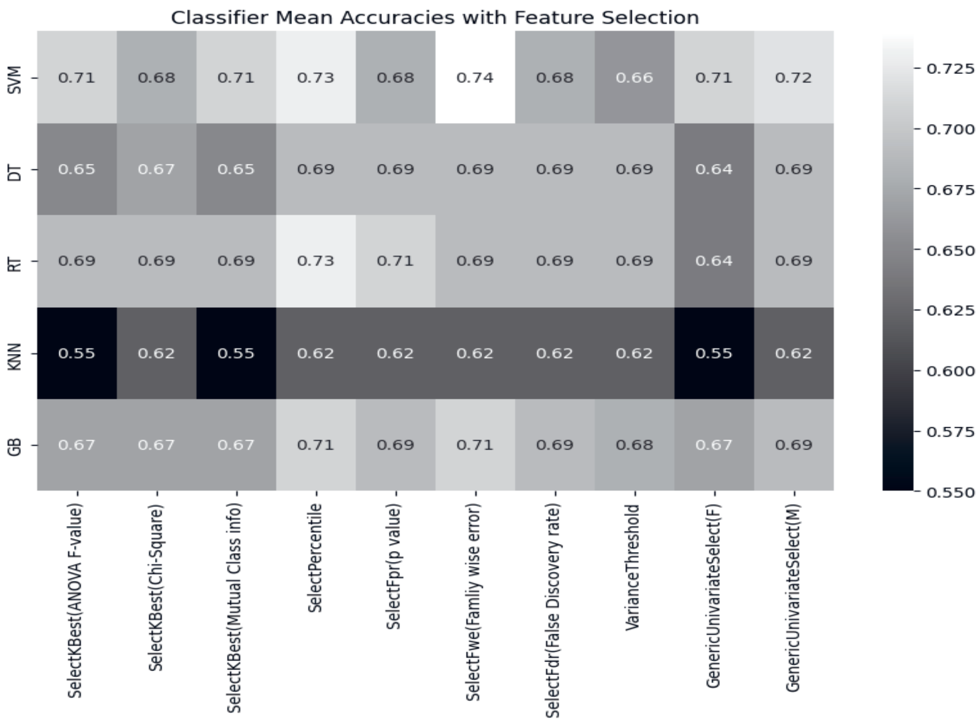

Figure 4. Accuracy of the Classifiers from Feature Selection Algorithms

Feature Selection Algorithm	SVM	DT	RT	KNN	GB
SelectKBest(ANOVA F-value)	0.69	0.72	0.71	0.65	0.74
SelectKBest(Chi-Square)	0.71	0.67	0.75	0.68	0.75
SelectKBest(Mutual Class info)	0.69	0.64	0.77	0.65	0.74

SelectPercentile	0.76	0.70	0.79	0.68	0.75
SelectFpr(p value)	0.69	0.65	0.81	0.68	0.78
SelectFwe(Famliy wise error)	0.76	0.72	0.81	0.68	0.75
SelectFdr(False Discovery rate)	0.69	0.65	0.79	0.68	0.78
VarianceThreshold	0.75	0.72	0.84	0.68	0.80
GenericUnivariateSelect(F)	0.69	0.72	0.70	0.65	0.74
GenericUnivariateSelect(M)	0.72	0.69	0.77	0.68	0.77

Table 2. Recall of Classifiers with Feature Selection

Feature Selection Algorithm	SVM	DT	RT	KNN	GB
SelectKBest(ANOVA F-value)	0.71	0.73	0.73	0.65	0.76
SelectKBest(Chi-Square)	0.74	0.68	0.76	0.68	0.76
SelectKBest(Mutual Class info)	0.71	0.64	0.77	0.65	0.76
SelectPercentile	0.77	0.70	0.81	0.68	0.77
SelectFpr(p value)	0.70	0.67	0.81	0.68	0.80
SelectFwe(Famliy wise error)	0.77	0.72	0.83	0.68	0.78
SelectFdr(False Discovery rate)	0.70	0.66	0.80	0.68	0.80
VarianceThreshold	0.77	0.72	0.86	0.69	0.82
GenericUnivariateSelect(F)	0.71	0.73	0.71	0.65	0.76
GenericUnivariateSelect(M)	0.73	0.69	0.781	0.69	0.80

Table 3. Precision of Classifiers with Feature Selection

Feature Selection Algorithm	SVM	DT	RT	KNN	GB
SelectKBest(ANOVA F-value)	0.69	0.72	0.71	0.64	0.73
SelectKBest(Chi-Square)	0.70	0.66	0.74	0.67	0.74
SelectKBest(Mutual Class info)	0.69	0.64	0.76	0.64	0.73
SelectPercentile	0.76	0.70	0.79	0.67	0.74
SelectFpr(p value)	0.69	0.65	0.81	0.67	0.78
SelectFwe(Famliy wise error)	0.76	0.72	0.80	0.67	0.74
SelectFdr(False Discovery rate)	0.69	0.64	0.79	0.67	0.78
VarianceThreshold	0.75	0.71	0.84	0.68	0.80
GenericUnivariateSelect(F)	0.69	0.72	0.70	0.64	0.73
GenericUnivariateSelect(M)	0.72	0.69	0.76	0.68	0.77

Table 4. F1-Score of Classifiers with Feature Selection

Feature Selection Algorithm	SVM	DT	RT	KNN	GB
SelectKBest(ANOVA F-value)	0.71	0.65	0.69	0.55	0.67
SelectKBest(Chi-Square)	0.68	0.67	0.69	0.62	0.67
SelectKBest(Mutual Class info)	0.71	0.65	0.69	0.55	0.67
SelectPercentile	0.73	0.69	0.73	0.62	0.71
SelectFpr(p value)	0.68	0.69	0.71	0.62	0.69
SelectFwe(Famliy wise error)	0.74	0.69	0.69	0.62	0.71
SelectFdr(False Discovery rate)	0.68	0.69	0.69	0.62	0.69

VarianceThreshold	0.66	0.69	0.69	0.62	0.68
GenericUnivariateSelect(F)	0.71	0.64	0.64	0.55	0.67
GenericUnivariateSelect(M)	0.72	0.69	0.69	0.62	0.69

Table 5. Accuracy of Classifiers with Feature Selection

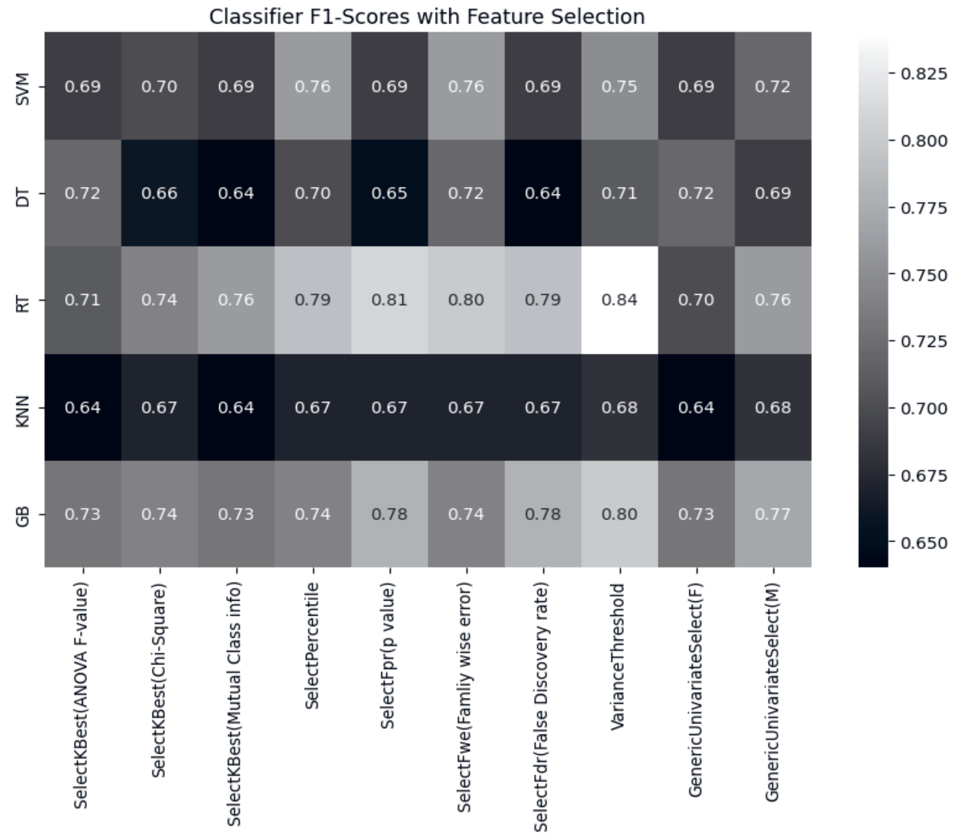

Figure 5. F1-Score of the Feature selection algorithms with the five Machine learning classifiers

SelectFwe (SFWE-F) is a feature selection algorithm that selects features based on their relationship with the target variable while controlling the family-wise error rate. In this study, SFWE-F performed best among the ten feature selection algorithms evaluated, achieving the highest accuracy of 74.3% when combined with Random Forests (RT) across all ten feature selection algorithms. Random Forest is a tree-based ensemble learning method that constructs multiple decision trees and combines their outputs to make predictions. It is known for its robustness against noise and high accuracy in both classification and regression tasks.

The reason why the combination of SFWE-F and Random Forests achieved the highest accuracy is that SFWE-F removes irrelevant features while controlling the family-wise error rate. By doing so, it identifies the most relevant features that have a strong correlation with the target variable, which in this case is the student grades. Random Forests, on the other hand, is able to capture the nonlinear relationships between features and the target variable. When SFWE-F is combined with Random Forests, it helps to reduce the number of irrelevant features, thereby improving the accuracy of the model. Additionally, Random Forests are known to handle noise and overfitting well, which makes them an excellent choice for educational datasets where noise and irrelevant features may be present.

6. Conclusion

Predicting timely student academic performance has become a major concern now a days. Machine Learning Models accurate predictions can benefit educational stakeholders in many capacities. Like other datasets Mostly educational datasets by nature can contains noise and less contributing features. Feature Selection/Removal algorithms plays vital role in building a machine learning model. There are many such algorithms presented. Keeping this in mind, the configured research main objective was to find best performing feature selection algorithm. Previously there were many approaches configured for this purpose. But a detailed and diversified in depth analysis was missing. For this the purpose, a significant pool of features selection algorithm was evaluated using different classifiers. The results claimed that Combination of the SVM with SelectFwe(Family wise error) proved to be effective with an accuracy of 74.3%.

This is due to the fact that the SelectFwe algorithm was able to select the most informative features while controlling the false discovery rate. Additionally, SVM is a powerful classification model that can perform well on a variety of datasets. Taking future aspects under consideration, evaluation of different under sampling and noise reduction algorithms along with multiple feature selection algorithms can be effective. Lastly replacement of traditional machine learning models with deep learning models can produce more predictive power.

References

[1] Guleria, P., & Sood, M.Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling. Education and Information Technologies, 2023,28(1), 1081-1116.Available From: https://link.springer.com/article/10.1007/s10639-022-11152-y.

[2] Jalota, C., & Agrawal, R. Feature selection algorithms and student academic performance: A study. In International Conference on Innovative Computing and Communications: Proceedings of ICICC, 2020, Volume 1 (pp. 317-328). Springer Singapore.

[3] Kumar, M., Sharma, B., & Handa, D. Building predictive model by using data mining and feature selection techniques on academic dataset. IJMECS,2022, 14, 16-29.

[4] Agrusti, F., Mezzini, M., & Bonavolontà, G. Deep learning approach for predicting university dropout: A case study at Roma Tre University. Journal of e-Learning and Knowledge Society,2020, 16(1), 44-54.

[5] Acharya, A., & Sinha, D. Application of feature selection methods in educational data mining. International Journal of Computer Applications, 2014,103(2).

[6] Mm Nidhi, N., Kumar, M., & Agarwal, S. Comparative Analysis of Heterogeneous Ensemble Learning using Feature Selection Techniques for Predicting Academic Performance of Students. In 2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST), IEEE,2021 ,pp. 212-217.

[7] Anuradha, C., & Velmurugan, T. A comparative analysis on the evaluation of classification algorithms in the prediction of students performance. Indian Journal of Science and Technology, 2015,8(15), 1-12.

[8] Enaro, A. O., & Chakraborty, S. Feature selection algorithms for predicting students academic performance using data mining techniques. Int. J. Sci. Technol.Res,2020,9(4).

[9] Bakker, T., Krabbendam, L., Bhulai, S., Meeter, M., & Begeer, S. Predicting academic success of autistic students in higher education. Autism,2023, 13623613221146439.

[10] Hamdi, M., Hilali-Jaghdam, I., Khayyat, M. M., Elnaim, B. M., Abdel-Khalek, S., & Mansour, R. F. Chicken Swarm-Based Feature Subset Selection with Optimal Machine Learning Enabled Data Mining Approach. Applied Sciences, 2022,12(13), 6787.

[11] Huynh-Cam T-T, Chen L-S, Huynh K-V. Learning Performance of International Students and Students with Disabilities: Early Prediction and Feature Selection through Educational Data Mining. Big Data and Cognitive Computing. 2022; 6(3):94. https://doi.org/10.3390/bdcc6030094.

[12] Najm, I. A., Dahr, J. M., Hamoud, A. K., Alasady, A. S., Awadh, W. A., Kamel, M. B., & Humadi, A. M. OLAP Mining with Educational Data Mart to Predict Students' Performance. Informatica,2022. 46(5).

[13] Goel, M., Sharma, A., Chilwal, A. S., Kumari, S., Kumar, A., & Bagler, G. Machine learning models to predict sweetness of molecules. Computers in Biology and Medicine, 2023.152, 106441.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 2011.12:2825–2830.

[15] Luepsen, H. . ANOVA with binary variables: the F-test and some alternatives. Communications in Statistics-Simulation and Computation,2023, 52(3), 745-769.

[16] Yu, L., & Liu, H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In Proceedings of the 20th international conference on machine learning 2003,ICML-03,pp. 856-863.

[17] Mahmood, S., Ghosh, S., Vazquez, M. J., Bista, R. G., Abdollahi, H., & Jain, A. N. A machine learning approach to predicting gene essentiality in cancer cell lines. Frontiers in Genetics,2022, 13, 760. doi: 10.3389/fgene.2022.810329.

[18] Khan, A. M., Ali, N. N., & Mahmood, A. Z.. Classification of breast cancer histology images using transfer learning and ensemble methods. Sensors, 19(23), 2019,5149. doi: 10.3390/s19235149.

[19] Saad, T. M., Abdel-Nasser, M. M., & Aziz, A. A. Brain tumor classification using convolutional neural networks for MRI images. Journal of Medical Systems, 2018,42(7), 129. doi: 10.1007/s10916-018-0983-3.

[20] Olsen, J. K., Kallimani, J. M., & D'Mello, M. J. Predicting performance on concept inventory questions using student-generated question features. In Proceedings of the 9th International Conference on Educational Data Mining 2016,pp. 462-465.

[21] Perera, R., & Jayawardena, Predicting student grades using machine learning: A comparison of feature selection techniques. In Proceedings of the International Conference on Advanced ICT for Education (ICAICTE) 2022,pp. 70-77.

[22] Amrieh, Elaf Abu, Thair Hamtini, and Ibrahim Aljarah. "Preprocessing and analyzing educational data set using X-API for improving student's performance." In 2015 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT), pp. 1-5. IEEE, 2015..

[23] Yu, N., Li, Z., & Yu, Z.. Survey on encoding schemes for genomic data representation and feature learning—from signal processing to machine learning. Big Data Mining and Analytics, 2018,1(3), 191-210.

[24] Mantovani, R. "Use of meta-learning for hyperparameter tuning of classification problems." PhD diss., University of Sao Carlos, 2018.