Journal of Information Systems Engineering and Management
2025, 10(8s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Ridge Regressive Quadratic Multivalued Feature Matching Pursuit for Skill-based Employability Identification in Higher Education

Bijithra. N. C[1*], Dr. E. J. Thomson Fredrik[2]
[1*]Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India
[2]Professor, Department of Computer Technology, Karpagam Academy of Higher Education, Coimbatore, India
[1*]Corresponding author email: bijithranc@gmail.com, ORCID ID: 0009-0005-0361-3966

ARTICLE INFO

Received: 20 Oct 2024
Revised: 12 Dec 2024
Accepted: 24 Dec 2024

ABSTRACT

Skill-based employability identification involves evaluating student's skills and determining their suitability for a specific job or industry. Data mining techniques have been developed for predicting student employability based on certain skills. Skills identification is a crucial step for students in understanding employability. However, accurate and time-efficient prediction of student employability has become a pivotal focus for educational institutions. This paper introduces a novel approach using data mining techniques called Ridge Regressive Quadratic Multivalued Projection Matching Pursuit (RRQMPMP) to identify skill-based employability for students in higher education with better accuracy and minimum time consumption. The proposed RRQMPMP technique includes two major processes namely preprocessing and feature selection. First, the number of features and student data are collected from the dataset. Then the preprocessing steps are executed, including three processes namely missing data handling, duplicate data removal, and normalization to clean the input dataset. The Ridge Regressive Imputation Method is employed to handle missing data in the dataset. Subsequently, duplicate and non-duplicate data points are distinguished from the dataset using a Simple Matching Distance Measure. Finally, Quadratic Mean Feature Scaling is developed for the normalization process. With the preprocessed dataset, the feature selection step is performed by applying a Russell-Rao Index Multivalued Projection Matching Pursuit. Based on the Russell-Rao Similarity Index value, pertinent and impertinent features are identified. Finally, pertinent features are selected for skill-based student employability prediction to achieve higher accuracy and minimize time consumption as well as space complexity. An experimental evaluation is carried out with respect to accuracy, error rate, time complexity, and space complexity for different numbers of student data. The quantitatively analyzed results indicate that the performance of the proposed RRQMPMP technique increases the accuracy of skill-based student employability prediction with minimum time and space complexity compared to conventional methods.

Keywords: Skill-Based Employability prediction of students, Preprocessing, Ridge Regressive Imputation Method, Quadratic Mean Feature Scaling, Russell-Rao Index Multivalued Projection Matching Pursuit based Feature selection.

INTRODUCTION

The employability status of students has become an important focus in the area of education institutions and employee development. Several methods have been developed for student employability prediction analysis. A Light Gradient Boosting Classifier (LGBC) was developed in [1] for predicting student employability based on sensitive features. However, model did not consider the additional attributes like language skills, which further enhanced the accuracy of employability status predictions for graduates. A hybrid approach, known as the Deep Belief Network and Softmax Regression (DBN-SR), was developed in [2] to identify student employ ability through pre-processing and the removal of irrelevant attributes. However, the major challenges were the performance-related issues, specifically in time complexity. A new hybrid model, Linear Vector Quantization combined with AdaBoost optimization, was developed in [3] for identifying student academic performance and employability chances. A machine learning model was developed in [4] for assessing employment and entrepreneurship levels of college students. However, the efficient machine learning model was not employed to achieve optimal results. A new predictive approach was introduced in [5] using Machine Learning (ML) algorithms for identifying the information technology student's employability. A deep learning approach was developed in [6] to investigate and analyze the mechanisms that quantitatively enhance college student's employment capabilities. While the model successfully reduces the error rate, it is notable that the time consumption for student's employment capability analysis remained high. In [7], the primary objective was to analyze the employability assessment of college graduate students based on skills acquired in a training program. However, the accuracy performance of the employability assessment was not improved. The Support Vector Machine (SVM) was developed in [8] with the aim of student employability identification. However, it was observed that time complexity associated with student employability identification was not effectively minimized. Blended learning and employability assessments were conducted throughout the degree course was described in [9]. However, identification of employability based on skills assessment remained a challenging issue. Partial Least Squares Structural Equation Modeling (PLS-SEM) was developed in [10] to establish connections between employability and entrepreneurial targets and highlight significant impact of gender. However, it failed to discussion concerning the prediction of women's employability in higher education.

1.1 Important Contributions of the Paper

- A novel RRQMPMP is introduced to enhance the identification of student employability, incorporating two distinct processes namely preprocessing and feature selection.
- To minimize the time and space complexity of RRQMPMP technique, first step involves preprocessing dataset. This is achieved by introducing Ridge Regressive Imputation for handling missing data, Simple Matching Distance for removing duplicate data, Quadratic Mean Feature Scaling for normalization.
- RRQMPMP technique employs Russell-Rao Index Multivalued Projection Matching Pursuit for pertinent and impertinent feature selection. Pertinent features are identified through multivalued feature mapping, utilizing Russell-Rao Similarity Measure. This, in turn, enhances accuracy and minimizes error rate.

RELATED WORKS

Sequential mixed methods were developed in [11] to assess noticeable employability of college students with a project management degree at an Australian university. This method was utilized to recognize how universities support the development of perceived employability in the project management context. Findings of the method employed for career education early with higher qualification, but the time was higher. A simple method for developing technical skills and enhancing employability skills was introduced in [12]. For evaluating the dissimilarity in results, Quasi-Experiment research was utilized for five months. However, the space complexity was not considered. A novel method was developed in [13] to analyze recommendations for designing group-based assessments with the aim of enhancing the employability skills of project management graduates. Two groups were considered to include fifteen project management academics from Australian and UK universities. However, the viewpoint of students was not considered in developing recommendations and techniques. MICE education aims to offer cultured and suitable for the MICE industry. An investigation into employability skills assessment was conducted in [14] for MICE management. Service efficiency and service experience were improved. However, it failed to analyze employability skills specifically for undergraduate- and graduate-level students. The isomorphism as a heuristic device was developed in [15] for focused on developing employability skills among the young people. The process of institutional homogenization was investigated. Studio School model was visualized into policy instant, but, it failed to minimize time. Big data with a machine learning approach was developed in [16] for employability prediction. The feature selection algorithms were employed for constructing and selecting subsets of features with maximum precision. However, the model improvement and prevention of data over fitting were not achieved. An empirical approach for detecting the employability skills of graduates was developed in [17]. Senior undergraduate as well as postgraduate programs examined for students in time management skills. However, the analysis of accuracy was not assessed. A hybrid Deep Neural Network (DNN) method was introduced in [18] to identify student's employability. Supervised machine learning techniques were applied for predicting employment status to enhance accuracy. However, the time complexity associated with student's employability identification was not analyzed. An efficient method, called as optimization of bagging classifiers, was introduced in [19] for predicting student employability. The correlation heatmap was employed to investigate relationships among several variables and the target variable. The variables are normalized by the standard scaler function. The optimal parameters were determined with GridSearchCV. However, it was not focused on enhancing the performance of the model in the student employability prediction. A decision support system for enhancing student employability identification was introduced in [20] specifically designed for higher education organizations. Data mining techniques employed for estimating student's preparedness, but error rate was not minimized.

PROPOSED METHODOLOGY

The main aim of detecting employability skills as part of the learning process has become vital in the higher education system. The employability of graduates serves as a performance indicator for the higher education industry. The identification model for employability is required for students to validate their ability to get employment as well as support to improve their skills to achieve the preferred job profile. The data mining and Machine Learning techniques were majorly developed for employability prediction. Data Mining is a process of discovering and extracting valuable information from the dataset and transforming it into a reliable structure for further processing. Data Mining has applied in various applications to a greater extent, especially skill-based employability detection is a significant application. With the above-said motivation, a novel technique RRQMPMP is introduced for skill-based employability detection through sensitive feature selection. The selection of sensitive features is crucial for creating a more accurate employability model.

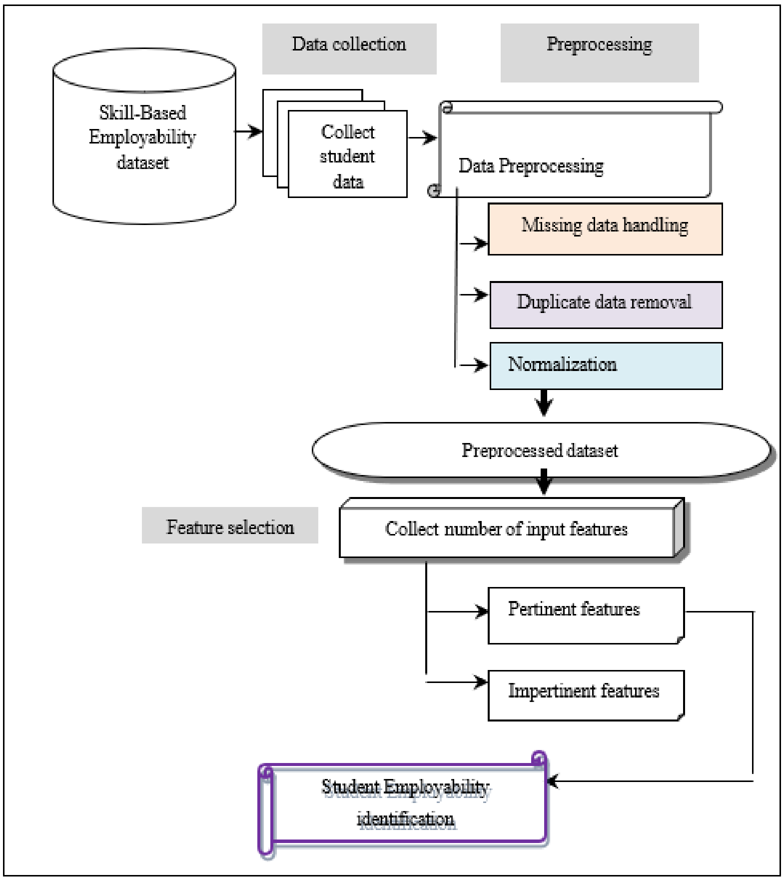

Figure 1 Process Fow of Proposed RRQMPMP Technique

Figure 1 illustrates the process flow of RRQMPMP techniques for accurately identifying skill-based student employability. The proposed RRQMPMP technique achieves student employability identification through student data collection, data pre-processing, and most favorable feature selection.

3.1 Data Collection Phase

The data collection phase is a fundamental component of the employability prediction model development process. This phase involves collecting the relevant and comprehensive data and to create a dataset https://www.kaggle.com/code/sayakghosh001/students-employability-dataset/input that summarizes the various aspects of a student's academic history and extracurricular skills. The dataset includes a collection of data stored in the form of rows and columns. Each column of a table represents a particular feature and each row indicates a given record of the data or number of instances.

3.2 Data Preprocessing

After data collection process, the proposed RRQMPMP technique undergoes a preprocessing step to obtain a structured dataset. Data preprocessing is a critical phase in the development of any Machine Learning model, involving the cleaning and transformation of raw datasets into a suitable format for analysis. This phase aims to enhance the quality of the dataset by addressing missing values, handling duplicates, and normalizing formats.

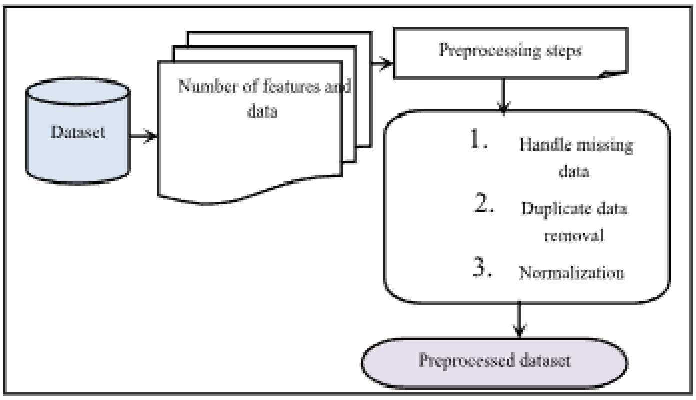

Figure 2 Structure of Data Preprocessing

Figure 2 depicts the structure of data preprocessing for detecting the employability identification of the students. The dataset includes a set of student data accumulated in the form of rows and columns. Each column of a table represents a particular feature and each row indicates a given record of the student data or instances. To start with the raw input dataset 'D' and arranged in the form of a matrix as below.

$$D = \begin{bmatrix} \boldsymbol{A_1} & \boldsymbol{A_2} & \dots & \boldsymbol{A_n} \\ d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \dots & \vdots \\ d_{m1} & d_{m2} & \dots & d_{mn} \end{bmatrix}, m = rows, n = columns \quad (1)$$

From above input matrix as given in (1), dataset 'D' includes 'n' column features or attributes $A_1, A_2, A_3 \ldots. A_n$and overall sample instances or student data 'd_{mn}'are presented in 'm row.

3.2.1 Ridge Regressive Imputation Method Based on Missing Data Handling

Handling missing data is a fundamental step in data preprocessing. Missing data refers to absence or lack of values in specific cells of a dataset. This absence occurs for various reasons due to data entry errors, equipment malfunction, or other factors. Effective handling of missing data is necessary to ensure the accuracy of student employability identification. RRQMPMP technique uses the Ridge Regressive Imputation Method to predict missing values based on the relationship between the missing values (i.e. dependent variable) and other independent variables (values of other variables). Ridge regression based imputing missing data is below.

$$d_m = \delta_o + \delta_1 d_1 + \delta_o + \delta_2 d_2 + \cdots + \delta_n d_n + \lambda \sum_{i=1}^{n} {\delta_i}^2 + E \quad (2)$$

Where, d_m denotes a missing value (i.e. dependent variable), $\delta_o$denotes an intercept term which is the value of the dependent variable when all independent variables have a value of zero, δ_1, δ_2, … δ_n denotes coefficients associated with the independent variables, λ denotes a regularization term to help prevent overfitting. Overfitting occurs when a model learns the training data too well and capturing noise, E indicates an error term which is measured as the difference between the actual values and the imputed (or predicted) values. The predicted values are imputed and handling the missing entries of the incomplete data in the dataset.

3.2.2 Duplicate Data Removal

Duplicate data removal is a preprocessing step in data mining to guarantee the quality and accuracy of the dataset by distinguishing the duplicate and non-duplicate data points. A duplicate record identification is carried out through exact matching involves comparing all data points in the specific attributes (or features) and removing the data points that have different values across all data points. A simple matching distance is used in the RRQMPMP technique to measure dissimilarity between the data points in the particular column.

$$\alpha_M = 1 - \left[\frac{{\vartheta_{ij}}^2}{m}\right] \quad (3)$$

Where, α_M indicates a simple matching distance, ϑ_{ij} denotes a distance between the data points in the particular column, m indicates the number of data points in the particular column.

$$\vartheta_{ij} = \sum_{i=1}^{m} \sum_{j=i+1}^{n} \left|d_i - d_j\right|^2 \quad (4)$$

Where ϑ_{ij} denotes a distance between the data points d_i and $d_j$in the particular column. The output of α_M is varied from 0 to 1. The value '1' indicates duplicate data points, while a similarity value of '0' indicates that the data points are categorized as non-duplicate. Through the analysis, the proposed technique accurately identifies duplicate data points with minimal processing time.

3.2.3 Data Normalization

The main aim of normalization is to transform the value of data in a common scale. RRQMPMP technique uses the quadratic mean feature scaling technique to normalize the input data points.

$$y = \frac{d - \mu_{qr}}{d_{mn} - d_{mx}} \quad (5)$$

$$\mu_{qr} = \sqrt{\frac{1}{n}\left({d_1}^2 + {d_2}^2 + {d_3}^2 + \cdots . + {d_n}^2\right)} \quad (6)$$

Where, y indicates normalized data results, d indicates an 'n' input data point of the feature, μ_{qr} denotes a quadratic mean, d_{mn}denotes minimum data point of the feature, d_{mx} indicates maximum data point of the feature. In this way, data normalization process is performed in efficient manner.

// Algorithm 1: Data Preprocessing
Input: dataset 'D', number of features and data points **Output**: Preprocessed dataset
Begin **Step 1:** Collect number of features and data points from the dataset 'D' **Step 2:** Formulate data in the form of row and column using (1) **Step 3: for each** feature and data points **Step 4:** Find the missing data using (2) **Step 5:** Detect the duplicate data using (3) **Step 6**: **if** (α_M = 1) **then** **Step 7**: identify duplicate data

Step 8: else
Step 9: identify non duplicate data
Step 10: end if
Step 11: Perform data normalization using (5) and (6)
Step 12: End for
Step 13: Return (preprocessed dataset)
End

3.3 Russell-Rao Index Multivalued Projection Matching Pursuit for Feature Selection

Feature selection is process of choosing a subset of relevant attributes from the original preprocessed dataset, aiming to enhance model performance and minimize time complexity. In this context, the RRQMPMP utilizes the Russell-Rao Similarity Index-based Projection matching Pursuit to select pertinent features and eliminate impertinent ones. Projection matching pursuit is a dimensionality reduction technique employed to identify the most informative projections in multidimensional space. For each projection, features are reduced based on their relevance. The main objective of Projection Matching Pursuit is to map pertinent features from high-dimensional space to low-dimensional space using the Russell-Rao Similarity Index. Multivalued mapping is also called as one-to-many mapping where each feature maps to multiple other features in the domain to identify the relevance. The Russell-Rao Similarity Index is a statistical method used to quantify the relevance between features. Relevant or pertinent features are selected for student employability identification.

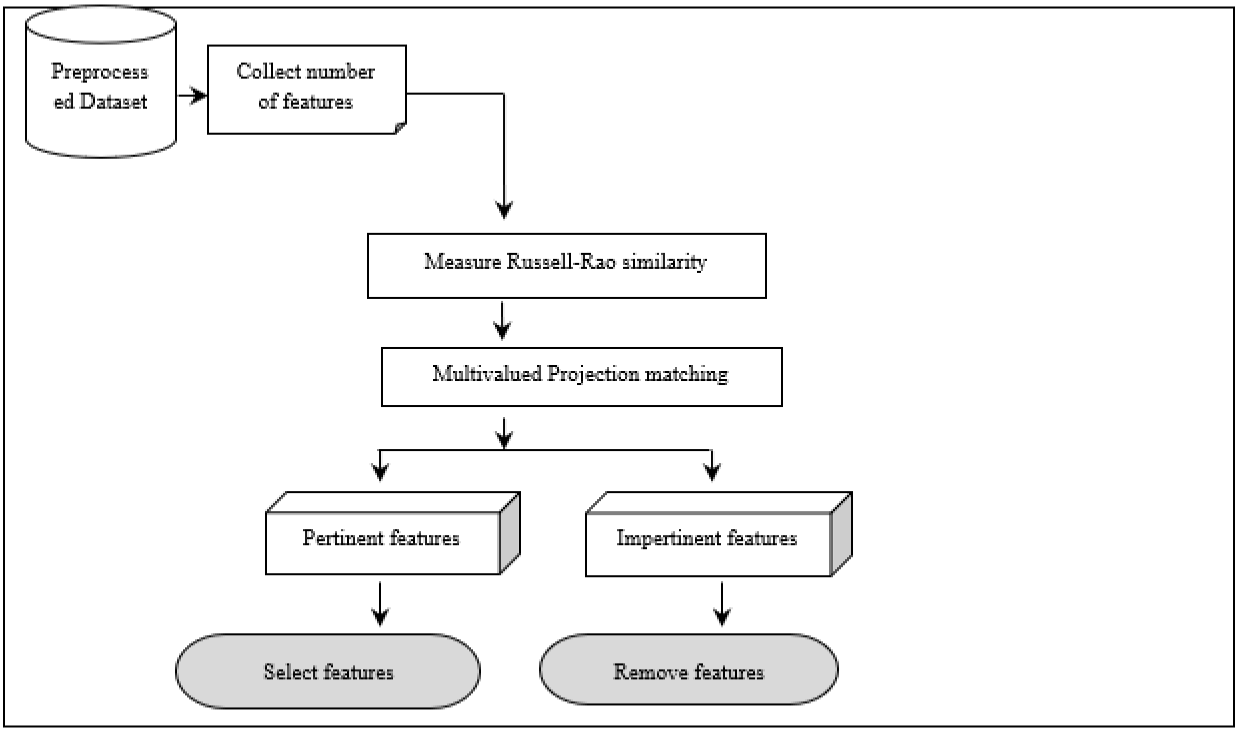

Figure 3 Flow process of Russell-Rao Index Multivalued Projection Matching Pursuit based Feature Selection

Figure 3 illustrates process of feature selection. First the number of features $A_1, A_2, A_3 \ldots A_n$ are considered as input. Let us consider the (A_i, A_j, RRS) in mapping space. A multivalued mapping is performed from one attribute A_i to many other attributes A_j denoted by,

$$f(x): A_i \rightarrow A_j \; where \; j = 1,,2,3 \ldots m \qquad (7)$$

Multivalued mapping satisfies the following similarity conditions. For each features performs one to many mappings based on Russel–Rao Similarity.

$$RRS = \sum_{i=1}^{n} \sum_{j=i+1}^{m} \left[\frac{A_i \cap A_j}{n}\right] \quad (8)$$

Where, RRS indicates a Russel–Rao Similarity, $A_i \cap A_j$ denotes a mutual dependence between the features A_i andA_j, n indicates a number of features in the given dataset. The similarity provides the outcomes range between 0 to 1. Based on the similarity value, the significant features are mapped by setting the threshold.

$$Y = \begin{cases} (RRS > T), \; PF \\ (RRS > T), \; IPF \end{cases} \quad (9)$$

Where Y denotes an output function that maps Pertinent Features (PF) when the similarity value (RRS) is greater than the Threshold (T), maps the Impertinent Features (IPF) when the similarity value (RRS) is lesser than the Threshold (T). In this way, the projection function maps the pertinent features and impertinent features from the high-dimensional space into low dimensional. Finally, it selects pertinent features and removes others. With selected pertinent features student employability is correctly detected with minimal time consumption.

// Algorithm 2: Russell-Rao Index Multivalued Projection Matching Pursuit
Input: Dataset 'D', features A = $\{A_1, A_2, A_3, \dots A_n\}$
Output: select pertinent features
Begin **Step 1:** collect number of features **from** Dataset 'D' **Step 2: for each** Features 'A_i' **Step 3: for each** Features 'A_j' **Step 4:** Perform multivalued mapping using (7) **Step 5:** Measure the **Russell-Rao similarity** 'RRS' **Step 6:** **If** **(**$RRS > T$**)** **then** **Step 7**: **Project** pertinent features **Step 8**: **else if (**$RRS < T$**)** **then** **Step 9**: **Project im**pertinent features **Step 10: End if** **Step 11:** **Select** pertinent features **Step 12: Remove im**pertinent features **Step 13**: **End for** **Step 14**: **End for** **End**

EXPERIMENTS

RRQMPMP and existing LGBM [1] and DBN-SR [2] are performed in the platform of python programming language using the Student's Employability Dataset. The implementation was conducted with system details of the system hardware and software like Windows 10 OS, core i3-4130 3.40GHZ Processor, and 8GB RAM.

4.1 Dataset Details

Student's Employability Dataset is used for conducting the experiment and it is taken from the https://www.kaggle.com/code/sayakghosh001/students-employability-dataset/input. The dataset was gathered from dissimilar university agencies in the Philippines. This dataset includes Mock job Interview Results of 2982 students which are scaled from 1 – 5 with their 10 features or attributes. The attribute description is listed in Table 1. First, the collected dataset is cleaned through missing data handling, duplicate data removal, and normalization in the preprocessing steps. Followed by, relevant or pertinent feature selection process is carried out to predict the Student's Employability.

Table 1 Attribute Description

S. No	Features or Attributes	Description
1	Name of Student	Student name
2	General Appearance	Scaled from 1 – 5 1. Low 2. Marginal 3. Average 4. Good 5. Superior
3	Manner Of Speaking	Scaled from 1 – 5 1. Low 2. Marginal 3. Average 4. Good 5. Superior
4	Physical Condition	Scaled from 1 – 5 1- Low 2- Marginal 3- Average 4- Good 5- Superior
5	Mental Alertness	Scaled from 1 – 5 1- Low 2- Marginal 3- Average 4- Good 5- Superior
6	self-confidence	Scaled from 1 – 5 1- Low 2- Marginal 3- Average 4- Good 5- Superior
7	Ability To Present Ideas	Scaled from 1 – 5 1- Low 2- Marginal 3- Average 4- Good 5- Superior
8	Communication Skills	Scaled from 1 – 5 1- Low 2- Marginal 3- Average

		4- Good 5- Superior
9	Student Performance Rating	Scaled from 1 – 5 1- Low 2- Marginal 3- Average 4- Good 5- Superior
10	Class	Indicates whether the student has 'Employable' or 'less Employable'

RESULTS AND DISCUSSION

The performance evaluation of RRQMPMP and existing methods [1] and [2] are conducted on different parameters including accuracy, error rate, time complexity, and space complexity. The performance results are presented by tables and graphical illustrations.

5.1 Accuracy:

It is measured by the number of instances or student data in student employability is correctly identified through feature selection. The accuracy is mathematically stated as given below.

$$AC = \sum_{i=1}^{n} \left(\frac{dci}{d_i}\right) * 100 \qquad (10)$$

Where AC indicates accuracy, d_i indicates number of instances or student data, dci indicates number of student data in student employability is correctly identified.

Table 2 Accuracy versus Number of Student Data

Number of Student Data	Accuracy (%)		
	RRQMPMP	LGBM	DBN-SR
250	90	87.6	85.6
500	90.2	88.4	86
750	89.2	86.93	84.66
1000	90.8	87	85
1250	89.6	85.6	84
1500	88.13	85.66	83.33
1750	88.85	86.57	82.85
2000	89.75	85	82.5
2250	90	87.11	83.11
2500	89.8	86.2	82

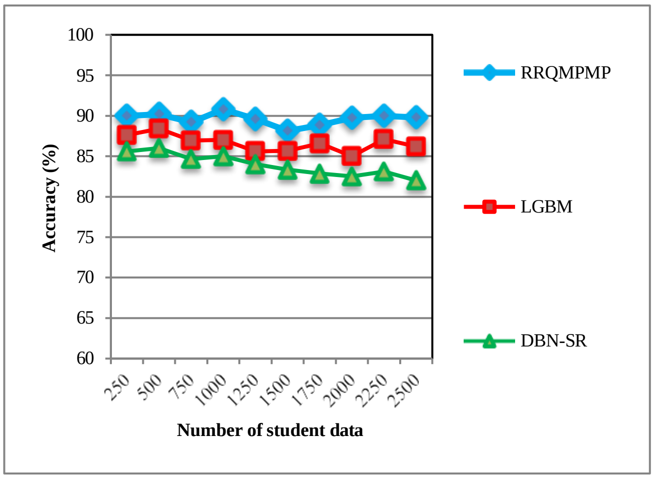

Figure 4 Performance Comparison of Accuracy

Figure 4 illustrates the graphical design depicting the accuracy of three methods, namely the proposed RRQMPMP technique, and two existing methods, namely LGBM [1] and DBN-SR [2], based on the number of student data points. On the horizontal axis, the number of student data ranges from 250, 500, 750 ... 2500, collected from the dataset and the vertical axis represents the corresponding accuracy results. The observed results reveal that the RRQMPMP technique enhances accuracy performance when compared to existing methods. Consider the input student data of 250 in the first iteration. By applying the RRQMPMP technique, results in an enhanced accuracy of employability identification were found to be 90%. In comparison, the accuracy of the existing methods [1] and [2] were found to be 87.6% and 85.6%, respectively. For each method, ten different sets of results were observed and compared. The average of these ten comparison results demonstrates that the accuracy of the RRQMPMP technique improves significantly by 4% and 7% compared to existing methods [1] and [2], respectively. This improvement is achieved though the efficient feature selection process employed by the RRQMPMP technique. With the preprocessed dataset, the RRQMPMP technique employs Russell-Rao Index Multivalued Projection Matching Pursuit for feature selection. The most relevant features for predicting student employability are accurately selected through the multivalued projection pursuit with the Russell-Rao Similarity Index. Features with higher similarity than the Threshold is selected as pertinent features for employability status prediction processes, resulting in higher accuracy.

5.2 Error rate:

It is measured by the number of instances or student data in student employability is incorrectly identified through feature selection. It is given below.

$$ET = \sum_{i=1}^{n} \left(\frac{dic}{d_i}\right) * 100 \qquad (11)$$

Where ET indicates error rate, d_i indicates a number of instances, dic indicates a number of instances or student data in which student employability is incorrectly identified.

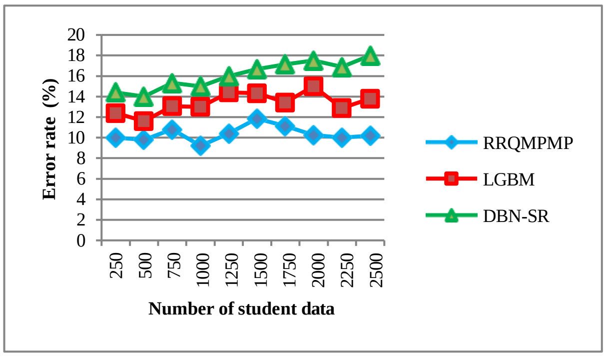

Figure 5 Performance Comparison of Error Rate

Figure 5 depicts the graphical results of the error rate in identifying the employability status of students using three methods namely the RRQMPMP technique and two existing methods, namely LGBM [1] and DBN-SR [2], based on a dataset of student information. In the figure, the horizontal axis represents the number of student data, while the vertical axis indicates the error rate output. The results illustrate that the error rate performance using the RRQMPMP technique is lower than that of existing methods [1] and [2]. Considering data from 250 students, the RRQMPMP technique achieved a 10% error rate. The other two existing methods [1] and [2] observed rates of 12.4% and 14.4%, respectively. A comparison of these results reveals that the RRQMPMP technique reduces the error rate by 22% and 35% compared to existing methods [1] and [2], respectively. This improvement is achieved by applying an RRQMPMP technique to select significant features and eliminate irrelevant ones from the dataset. By doing so, it effectively minimizes error rates associated with irrelevant features. Furthermore, the preprocessing step of the RRQMPMP technique handles the missing values and eliminates duplicate data from the dataset, contributing to the overall reduction in error rates in identifying student employability status.

5.3 Time complexity:

It is another significant performance metric to evaluate the performance of student employability identification. It is defined as the amount of time consumed by the algorithm for student employability identification through the feature selection process. This is mathematically formulated as follows:

$$T_{com} = \sum_{i=1}^{n} d_i * \ TM\,[SEI] \qquad (12)$$

Where, T_{com}denotes a time complexity, d_i indicates a number of instances or student data, $TM\,[SEI]$ represents time for student employability identification. It is measured in milliseconds (ms).

Table 3 Time Complexity versus Number of Student Data

Number of Student Data	Time Complexity (ms)		
	RRQMPMP	LGBM	DBN-SR
250	23.75	26.25	28
500	27.5	29.5	31
750	31.5	34.5	36
1000	35	40	42
1250	37.5	42.5	45
1500	39	45	46.5
1750	42	47.25	49
2000	44	50	52
2250	47.25	51.75	54
2500	50	52.5	55

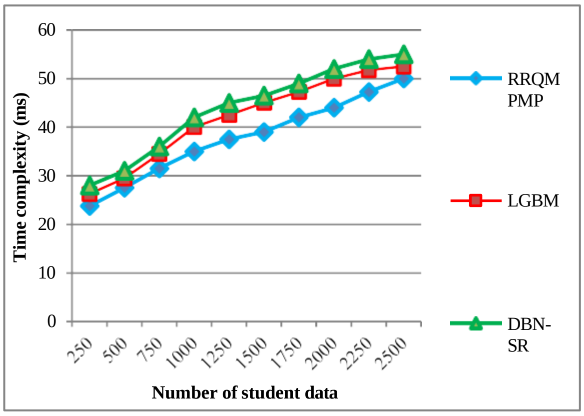

Figure 6 Performance Comparison of Time Complexity

Figure 6 depicts the performance results of time complexity versus the number of student data. As shown in the graph, the number of student data is represented on the 'x' axis, while the time complexity results are displayed on the 'y' axis. The experimental results for time complexity were obtained using three methods, namely RRQMPMP, LGBM [1], and DBN-SR [2]. The observed time complexity results indicate that the performance of the RRQMPMP technique is comparatively minimized compared to the existing methods. For instance, with the consideration of 250 data points, the time taken to perform employability identification was found to be 23.75ms for RRQMPMP, and it was 26.25m and 28ms for the existing methods [1] and [2], respectively. The observed results confirm that the RRQMPMP technique minimizes time complexity. After obtaining ten results, the overall time complexity of the RRQMPMP technique is compared to the existing results. The average of the ten results demonstrates that the proposed RRQMPMP technique reduces time consumption by 10% and 14% compared to [1] and [2], respectively. This is due to the fact that the RRQMPMP technique performs data preprocessing, including missing data handling, removing duplicates, and normalization. All three processes are employed for cleaning the dataset. Following this, the identification of student employability status involves relevant features instead of using the entire feature set, resulting in minimized time complexity.

5.4 Space Complexity:

It is another dimensionality reduction performance metric, specifically defined as the amount of storage space consumed by the algorithm during the feature selection process for identifying student employability. The computation is as follows,

$$SE_{com} = \sum_{i=1}^{n} d_i * \ SS\ [SEI] \qquad (13)$$

Where, SE_{com} denotes space complexity, d_i indicates number of instances or student data, $SS\ [SEI]$ represents storage space for student employability identification. It is measured in kilobytes (KB).

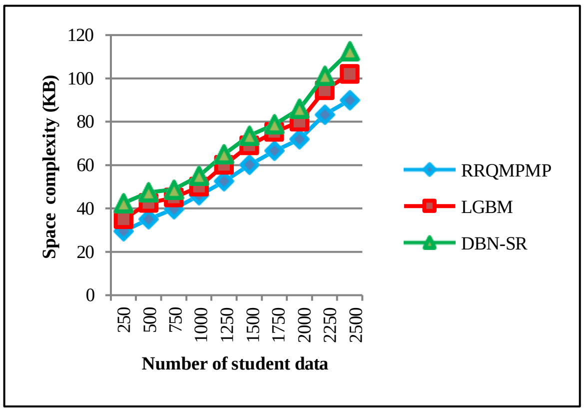

Figure 7 Performance Comparison of Space Complexity

The performance outcomes regarding space complexity are depicted in figure 7 for three methods namely RRQMPMP, LGBM [1], and DBN-SR [2], in relation to the number of student data ranging from 250 to 2500. Let us consider the scenario where the number of student data is 250 in the experimentation. The storage spaces required for predicting student employability were determined to be $29.5KB$ using the RRQMPMP technique and the space complexities of the other two methods [1] and [2] were measured at $35KB$ and$42.5MB$, respectively. The observed results indicate that the proposed RRQMPMP technique achieves lower space consumption than the existing methods. Averaging the results over ten experiments demonstrates that the overall performance of the space complexity of the RRQMPMP technique is considerably reduced by 12% and 20% compared to existing methods. This reduction is achieved due to the application of the feature selection process using the Russell-Rao Index Multivalued Projection Matching Pursuit. By removing irrelevant features and their associated data from the collected dataset, the space complexity is effectively decreased. Additionally, the utilization of the simple matching distance measure eliminates duplicate data, further minimizing the space complexity.

CONCLUSION

Educational data mining is a promising area for analyzing student's employability skills. This paper proposes accurate student employability identification through the assistance of an efficient RRQMPMP technique with minimum time. In data preprocessing, regression-based missing data handling is performed. Feature scaling is carried out through the quadratic mean. Quantitatively analyzed results indicate that RRQMPMP technique exhibits improved performance achieve the 6% of improved accuracy and 28%, 12% of minimum error, time complexity, as well as 16% of space complexity compared to existing techniques.

REFERENCES

[1] Oumaima Saidani, Leila Jamel Menzli , Amel Ksibi , Nazik Alturki , And Ala Saleh Alluhaidan, "Predicting Student Employability Through the Internship Context Using Gradient Boosting Models", IEEE Access, Volume 10, 2022, Pages 46472 – 46489.

[2] Anita Bai and Swati Hira, "An intelligent hybrid deep belief network model for predictingstudents employability", Soft Computing, Elsevier, Volume 25, 2021, Pages9241-9254.

[3] K. Subhash Bhagavan,J. Thangakumar, D. Venkata Subramanian, "Predictive analysis of student academic performance and employability chances using HLVQ algorithm", Journal of Ambient Intelligence and Humanized Computing, Springer, Volume 12, 2021, Pages3789-3797. https://doi.org/10.1007/s12652-019-01674-8

[4] Shixiao Li, "The Use of Machine Learning Model in the Evaluation of College Students' Employment and Entrepreneurship Level", Wireless Communications and Mobile Computing, Hindawi, Volume 2022, July 2022, Pages 1-10.

[5] Gehad ElSharkawy, Yehia Helmy, Engy Yehia, "Employability Prediction of Information Technology Graduates using Machine Learning Algorithms", International Journal of Advanced Computer Science and Applications (IJACSA), Volume 13, Issue 10, 2022, Pages359-367.

[6] XiangminMeng,GuoyanRen,and Wenjun Huang, "A Quantitative Enhancement Mechanism of University Students' Employability and Entrepreneurship Based on Deep Learning in the Context of the Digital Era", Scientific Programming, Hindawi, Volume 2021, November 2021, Pages 1-12.

[7] Pilar Laguna-Sánchez, Pilar Abad, Concepción de la Fuente-Cabrero and Rocío Calero, "A University Training Programme for Acquiring Entrepreneurial and Transversal Employability Skills, a Students' Assessment", Sustainability, Volume 12, Issue 3, 2020, Pages 1-17.

[8] Cherry D. Casuat, Enrique D. Festijo, Alvin Sarraga Alon, "Predicting Students' Employability using Support Vector Machine: A SMOTE-Optimized Machine Learning System", International Journal of Emerging Trends in Engineering Research, Volume 8, Issue 5, 2020, Pages 2101 – 2106.

[9] Teresa Crew and Olivia Martins, "Students' views and experiences of blended learning and employability in a post-pandemic context", Social Sciences & Humanities Open, Elsevier, Volume 8, Issue 1, 2023, Pages 1-8.

[10] Jos´e Manuel Santos-Ja´en, Patricia P. Iglesias-Sanchez, Carmen Jambrino-Maldonado, "The role of gender and connections between entrepreneurship and employability in higher education", The International Journal of Management Education, Elsevier, Volume 20, Issue 3, November 2022, Pages 1-15.

[11] Guinevere Gilbert and, Michelle Turner, Omid Haass, "Working up to work: Perceived employability of students commencing a project management degree", Project Leadership and Society, Elsevier, Volume 3, December 2022, Pages 1-9.

[12] Pallavi Gupta, Ambarish Dattaa, Satyanarayan Kothe, “Developing employability skills in vulnerable youth: Designing logic model framework and outcome evaluation using quasi-experiment”, World Development Sustainability, Elsevier, Volume 2, 2023, Pages 1-12.

[13] Roksana Jahan Tumpa, Samer Skaik, Miriam Ham, Ghulam Chaudhry, “Enhancing project management graduates’ employability through group assessment innovations: An empirical study”, Project Leadership and Society, Elsevier, Volume 4, 2023, Pages 1-14.

[14] Xiao Liu, Randy Seevers, Hongyi Lin, “Employability skills for MICE management in the context of ICTs”, PLoS ONE, Volume 17, Issue 7, Pages1-16.

[15] James Robson, Ashmita Randhawa and Ewart Keep, “Employability skills in mainstreameducation: Innovations in schooling andinstitutional isomorphism”, British Educational Research Journal, Volume 48, Issue 1, 2022, Pages 120–136, DOI: 10.1002/berj.3756

[16] DrishtySobnath, Tobiasz Kaduk, Ikram Ur Rehman, Olufemi Isiaq, “Feature Selection for UK Disabled Students’ Engagement Post Higher Education: A Machine Learning Approach for a Predictive Employment Model”, IEEE Access, Volume 8, 2020, Pages 159530 – 159541.

[17] Barbara A. Stewart, “An empirical approach to identifying employability skills required of graduates in the environmental sciences”, Industry and Higher Education, Volume 35, Issue 2, 2020, Pages1–13.

[18] Tianyi Sun and Zheng He, “Developing intelligent hybrid DNN model for predictingstudents’ employability – A Machine Learning approach”, Journal of Education, Humanities and Social Sciences, Volume 18, 2023, Pages 235-248.

[19] Minh-Thanh Vo, Trang Nguyen and Tuong Le, “OPT-BAG Model for Predicting Student Employability”, Computers, Materials & Continua, Volume 76, Issue 2, 2023, Pages 1555-1568.

[20] Maria Elisa Linda Taeza-Cruz and Marifel Grace Capili-Kummer, “Decision Support System to Enhance Students’ Employability using Data Mining Techniques for Higher Education Institutions”, International Journal of Computing and Digital Systems, Volume 13, Issue 1, 2023, Pages 1253-1262.