Journal of Universal Computer Science, vol. 28, no. 9 (2022), 898-915
submitted: 21/1/2022, accepted: 30/4/2022, appeared: 28/9/2022

English Teaching in Artificial Intelligence-based Higher Vocational Education Using Machine Learning Techniques for Students' Feedback Analysis and Course Selection Recommendation

Xin Ma
(School of Foreign Languages, Zhengzhou Normal University, Zhengzhou, 450045, China
https://orcid.org/0000-0002-6004-4326, focusmaxine@163.com)

Abstract: Higher vocational education is a self-contained method of higher education that is aligned with global productivity and economic development. Its goal is to develop talented workers who contribute significantly to the economy and industry. Teaching analysis, teaching strategy, teaching practice, and assessment are all part of the course design process in high vocational education. Teaching assessment is one of the most effective methods for improving the quality of course teaching among teaching processes. This research proposes novel techniques in English teaching based on artificial intelligence for course selection based on students' feedback. Here, the dataset has been collected based on the students' feedback on courses for Higher Vocational Education in English teaching. This dataset has been processed to remove invalid data, missing values, and noise. The processed data features have been dimensionality reduction integrated with K-means neural network. And the extracted features have been classified with higher accuracy using recursive elimination-based convolutional neural network. Based on this feedback data classification, recommendation for courses in Higher Vocational Education in English teaching has been suggested. The experimental analysis shows various students' feedback dataset validation and training in terms of accuracy of 96%, precision of 92%, recall of 93%, RMSE of 68%, and computational time of 65%.

Keywords: Higher vocational education, artificial intelligence, course selection, students' feedback, English teaching
Categories: H.5.1, L.3.2, I.2.7, I.2.8, I.6.4, I.7.0, M.7
DOI: 10.3897/jucs.94160

1 Introduction

Artificial intelligence (AI) and digitalization are transforming the way we work, live, communicate, learn, and play. Individuals are increasingly encountering advanced technology such as AI in their daily lives, whether they are aware of it or not, in interactions as different as applying for a loan as well as reading through social media, of which may have a dramatic impact on their lives. [Farahabadi et al., 2021]. Many jobs are being touched by human creativity as well as innovation, which is making it increasingly possible to generate value from new technology two decades into the twenty-first century. AI is increasingly being incorporated into as well as impacting industrial and agricultural processes, services, value chains, and workplace organizations, with more national policies concentrating on it. AI can enhance people's lives, but it also poses ethical and societal concerns, including employment creation and obsolescence [Reddy et al. 2020]. It is a source of social as well as political friction, and it has the potential to exacerbate disparities within and across countries.

Furthermore, a rising body of data suggests that factors like technique as well as globalization are either 'polarising' the workforce into high- and low-skilled occupations or 'hollowing out' the demand for intermediate-level skills. Intermediate-skilled workers are particularly vulnerable due to the repetitive nature of the activities they frequently perform and the fact that techniques that potentially replace them, such as AI and robotics, can save companies a lot of money [Sarveniazi, 2014]. However, its vital to note that this only applies to intermediate talents as we know them now. The permanence of intermediate occupations is revealed via analyses conducted by occupation rather than the wage percentiles, indicating "the shifting character of intermediate positions" [Zhang et al. 2017]. Equality, as well as openness of education, are important reform markers for modernizing education reform. Use of AI methods can help reach this reform goal more effectively [Perez and Tah, 2020]. In China, science and technology innovation are the primary drivers of MODE reform [Sadr et al., 2019], [Venkateswaran, 2022]. China's AI speech recognition system has now surpassed the world's best, with a 97 percent accuracy rate. With the arrival of the fourth scientific as well as method revolution, China should make full use of more mature AI methods, innovate TECHNIQUES and implement REFORM to facilitate the inefficient and stupid CONDITION [Chen and Jeong, 2007].
The contribution of this paper is as follows:

1. To propose a novel technique in English teaching based on artificial intelligence in course selection based on students' feedback
2. The dataset has been collected based on the student's feedback on courses for Higher Vocational Education in English teaching
3. To extract features using dimensionality reduction integrated with K-means Neural network.
4. To classify the extracted features using recursive elimination-based convolutional neural network.
5. To classify the feedback data, the recommendation for courses in Higher Vocational Education in English teaching has been suggested.

2 Related Works

There are various effective ways in NLP that help in educational contexts, such as the function of empirical data, corpora, and other linguistic features that are important and effective in the language learning process. Reference [Sadr et al., 2019] addressed new prospects for improving natural language processing (NLP) and its utility in developing educational tools such as reading and writing materials. The use of linguistics in the classroom can help students manage and deal with reading and writing challenges. This can be accomplished by examining syntactic and morphological parameters. Motivation in language acquisition is a powerful tool that may also improve students' educational practices and academic achievement [Venkateswaran, 2022]. DL research has seen a significant increase in activity in recent years. In [National body, 2021], popular architectural models and training techniques were used to introduce DL in NNs briefly. Because of the rise in data volume and processing power, neural networks with increasingly complicated topologies have got a lot of interest. They have been used in

a variety of sectors [Radic et al., 2020]. A considerable number of research investigations focused on practical applications, yielding many study findings.
DL methods have been utilized in image analysis, text analysis, speech recognition, and other fields, providing solutions to various real-world problems [Mosavi et al., 2020]. The basic knowledge of transfer learning, the numerous types of methodologies utilized to accomplish transfer learning, and how transfer learning was being used in many subfields of medical image analysis were all evaluated by the author [Ilic et al., 2020]. The review demonstrates that recent developments in DL, particularly advances in transfer learning, have enabled the identification, classification, and quantification of specific patterns from many medical pictures [Taylor et al., 2021]. The authors [Dhawan and Batra, 2021] introduced the Char-CNNS (Character-level CNN with Shortcuts) method to provide an automated approach for determining whether the material in social media comprises cyberbullying. To assist people in comprehending how DL methods are tailored to meet the challenge of speaker recognition, the work [Xiao and Yi, 2020] presented a new technique to extract speaker features by developing CNN filters linked to the speaker. Researchers have employed DL methods to build efficient processes to calculate students' learning state as well as behaviour [Bucea-Manea-Tonis et al. 2020] [Bernardo et al., 2021] [Du, 2021]. The success of teaching and learning depends on interaction. Researchers use DL methods to create ways to assist and encourage students' learning enthusiasm because of their strength in natural language generation as well as processing [Dong and Tsai, 2021].

3 System Model

This section proposes novel techniques in English teaching based on AI in course selection based on students' feedback. Here, the dataset has been collected based on the students' feedback on courses for Higher Vocational Education in English teaching. This dataset has been processed to remove invalid data, missing values, and noise. The processed data features have been dimensionality reduction integrated with K-means Neural network. And the extracted features have been classified with higher accuracy using a recursive elimination-based convolutional neural network. Based on this feedback data classification, recommendation for courses in Higher Vocational Education in English teaching has been suggested. The overall proposed architecture is shown in Figure 1.

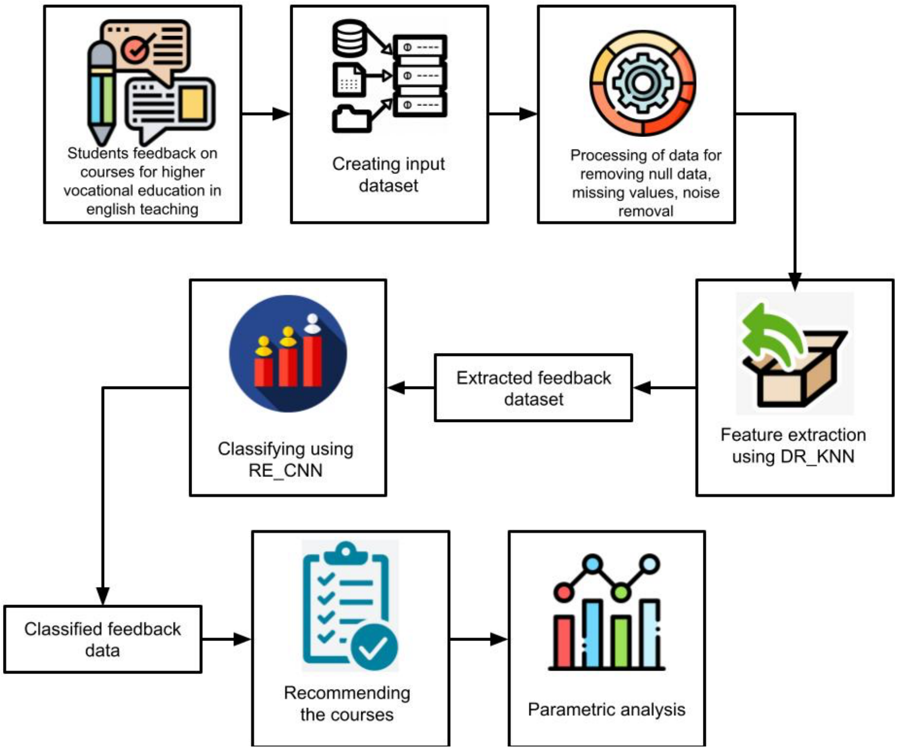


Figure 1: Overall proposed architecture

4 Dimensionality Reduction Integrated with K-means Neural Network

Let w = [w1,...wn] be the coefficients of the first principal component, which is a linear combination of X. eq. (1) in matrix form:

$$U_1 = w^T X$$

$$\text{var}\ (U_1) = \text{var}\ (w^T X) = w^T S w \quad (1)$$

Where S is X's n × n sample covariance matrix, Clearly, increasing the amount of w can make var(U1) arbitrarily big. As a result, we select w to maximize $w^T S w$ while requiring w to be of unit length in Eq (2).

$$\max w^T Sw \text{subject to } w^T w = 1 \quad (2)$$

In eq. (3), a Lagrange multiplier α1 of 1 is introduced to solve this optimization problem:

$$L(w, \alpha) = w^T S w - \alpha_1 (w^T w - 1) \tag{3}$$

Standardization of raw data: In eq. (4), the raw data should have unit variance and zero mean

$$x_j^i = \frac{x_i^i - \tilde{x}_j}{\sigma_j} \forall j \tag{4}$$

In eq. (5), calculate the raw data's covariance matrix as follows:

$$\sum = \frac{1}{m} \sum_i \text{ n}(x_i)(x_i)^T, \sum \in R^{n*n} \tag{5}$$

Data must first be standardized before the covariance matrix can be calculated. To do so, we use eq. (6) to calculate the primary vector of empirical mean:

$$U_m = \frac{1}{n} \sum_{i=1}^{n} X_{[m,i]} \tag{6}$$

On matrix lines, the empirical mean would be used. Distance matrix with the mean would then be calculated as eq (7)

$$\mathrm{B = X - uh,} \tag{7}$$

In each of the entries, h is a vector with a size of 1 × n and a value of 1. Equation (8) would be used to produce a covariance matrix Σ with m × m dimensions:

$$\sum = E[B \otimes B] = E[B \cdot B^*] = \frac{1}{n} B \cdot B^* \tag{8}$$

Consider the random vector $X' = [X_1, X_2, \cdots, X_n]$ and assume that it exhibits matrix covariance Σ with exceptional values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. Take a look at the linear compositions in eq. (9):

$$\begin{cases} Y_1 = l_1' X = l_{11} X_1 + l_{21} X_2 + \cdots + l_{n1} X_n, \\ Y_2 = l_2' X = l_{12} X_1 + l_{22} X_2 + \cdots + l_{n2} X_n, \\ \vdots \\ Y_n = l_n' X = l_{1n} X_1 + l_{2n} X_2 + \cdots + l_{nn} X_n. \end{cases} \tag{9}$$

$$\text{var } (Y_i) = l_i' \textstyle\sum_i, \text{ cov } (Y_i, Y_k) = l_i' \textstyle\sum_k, \; i, k = 1,2, \cdots, n$$

Calculate the covariance matrix's eigenvectors and eigenvalues as provided in eq (10)

$$u^T \sum = \lambda \mu$$

$$U = \begin{bmatrix} 1 & | & | \\ u_1 & u_2 \ldots & u_n \\ 1 & | & | \end{bmatrix} \tag{10}$$

It is necessary to project the raw data onto a k-dimensional subspace: Pick up the covariance matrix's top k eigenvectors. These will be the data's new original foundations. The relevant vector's calculation is provided in Eq (11)

$$x_i^{\,new} = \begin{bmatrix} u_1^T x^i \\ u_2^T x^i \\ :::: \\ :::: \\ u_k^T x^i \end{bmatrix} \in R^k \tag{11}$$

For given data $\{x_i\}_{i=1}^N$ let $D = [d_{ij}]$be pairwise Euclidean matrix whose entry ij, and d shows Euclidean distance between high-dimensional data points i-x multidimensional scaling finds a linear mapping P such that enhances the cost function in eq. (12):

$$\psi(Y) := \sum_{l,j} \left(d_{ij}^2 - \|y_i - y_j\|^2 \right) \tag{12}$$

The Euclidean distance between the low-dimensional data points y_i and $y_j$is $\|y_i - y_j\|$ with $\|v_j\|^2 = 1$ for all column vector v_j of P, and y_j is constrained to be $x_i A$. The eigen decomposition of the Gram matrix T G XX =, X x = I, is demonstrated to produce the minimum of the cost function a (Y). Actually, we may get the Gram matrix by computing eq. (13): double-centering pairwise squared Euclidean distance matrix

$$g_y = -\frac{1}{2}\left(d_{lj}^2 - \frac{1}{n}\sum_l d_d^2 - \frac{1}{n}\sum_l d_{jl}^2 + \frac{1}{n^2}\sum_{l=m} d_{lm}^2 \right) \tag{13}$$

Scatters are measured by utilizing scatter matrices. Consider r class C_i each including n_i points $x_j^i \in \mathbb{R}^l$ and set $X = [\hat{C}_1, \cdots, \hat{C}_r] \in \mathbb{R}^{l \times n}$, where $\hat{C}_i = [x_1^i, \cdots, x_{n_i}^i]$ and $n = \sum_{i=1}^r n_j$. Let $\overline{x'} = \frac{1}{n_i}\sum_{j-r}^n x_j'$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^n x'$

Now define 3 scatter matrices:

Between-class scatter matrix $S_b := \sum_{i=1}^r n_i (\overline{x^y} - \bar{x})(\overline{x^2} - \bar{x})^T$,

Within-class scatter matrix $S_w := \sum_{i=1}^r \sum_{j-1}^{n_i} (x_j' - \overline{x^t})(x_j' - \overline{x^\iota})^T$,

Total scatter matrix $S_t := \sum_{i=1}^r \sum_{j=1}^{n_1} (x_j - \bar{x})(x_j - \bar{x})^T$. LDA is a strategy for solving Eq. (14)'s optimization problem:

$$\arg \max_{U \in \mathbb{R}^{6=}} \frac{|U^T S_b U|}{|U^T S_w U|} \tag{14}$$

Let $X = x_{i_{iel}^n}^n$ and $Y = y_{i,4}^n$ be two data set of n points in $\mathbb{R}^p$ and $\mathbb{R}^q$, associated with them have two matrices by eq. (15):

$$A_X = [x_1 - \bar{x}, \cdots, x_n - \bar{x}] \in \mathbb{R}^{p \times n}, A_y = [y_1 - \bar{y}, \cdots, y_n - \bar{y}] \in \mathbb{R}^{q \times n} \tag{15}$$

where $\tilde{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\tilde{y} = \frac{1}{n}\sum_{i=-}^{n} y_i$ are means of x_i and y_i s.

CCA is, in fact, a strategy for solving the following optimization problem in Eq. (16):

$$\arg\max_{U_X,V_y} \frac{U_X^T A_X A_y^T U_y}{\sqrt{U_X^T A_X A_X^T U_X}\sqrt{U_Y^T A_y A_y^T U_Y}} \tag{16}$$

which can be modified as eq. (17)

$$\arg\max_{U_X U_Y} U_X^T A_X A_Y^T U_Y, U_X^T A_X A_X^T U_X = 1, U_Y^T A_Y A_Y^T U_Y = 1 \tag{17}$$

Assuming that the solution to the previous optimization issue is the pair of projective directions (U_x^*, U_y^*), we can identify another pair of projective orders by solving eq. (18)

$$\arg\max_{U_X,U_Y} U_X^T A_X A_Y^T U_Y, U_X^* A_X A_X^T U_X^* = U_Y^{*T} A_Y A_Y^T U_Y^* = 0, U_X^T A_X A_X^T U_X = U_Y^T A_Y A_Y^T U_Y = 1 \tag{18}$$

We get an m-dimensional specification of the linear combination of these vector solutions by repeating the previous technique m-1 times. In reality, by solving the paired eigenvalue problem eq. (19), we can obtain this m-dimensional space:

$$A_X A_Y^T (A_Y A_Y^T)^{-1} U_X = \lambda A_X A_X^T U_X, A_Y A_X^T (A_X A_X^T)^{-1} U_X = \lambda A_y A_y^T U_y \tag{19}$$

and eigenvectors $\left(U_X^{(i)}, U_Y^{(i)}\right), i = 1, \cdots, m$ corresponding to m largest eigenvalues are pairs of projective directions. Hence from eq. (20)

$$\left\{U_x^{(i)T} A_x, i = 1, \cdots, m\right\} \text{ and } \left\{U_y^{(t)r} A_Y, i = 1, \cdots, m\right\} \tag{20}$$

The ith row and jth columns of a matrix $\mathbf{X} = [x_{ij}]$ are given as xi and xj. Moreover refer to Frobenius norm, 2-norm, 1-norm, and 2,1-norm of a matrix X as

$\| \mathbf{X} \|_F = \sqrt{\sum_i \|\mathbf{x}^i\|_2^2} = \sqrt{\sum_j \|\mathbf{x}_j\|_2^2}$, $\| \mathbf{X} \|_2 = \sqrt{\sum_i \sum_j |x_{ij}|^2}$, $\| \mathbf{X} \|_1 = \sum_i \sum_j |x_{ij}|$, and $\| \mathbf{X} \|_{2.1} = \sum_i \|\mathbf{x}^i\|_2 = \sum_i \sqrt{\sum_j \mathbf{x}_{ij}^2}$. Transpose, trace operator and inverse of a matrix X are denoted by $\mathbf{X}^T$, Tr ($\mathbf{X}$), and $\mathbf{X}^{-1}$. As seen in Eq. (21) the least-squares loss function is as follows:

$$\min_{\mathbf{W}} \sum_{i=1}^{m} \|\mathbf{X}^T \mathbf{w}_i - \mathbf{y}_i\|_2^2 = \min_{\mathbf{W}} \|\mathbf{X}^T \mathbf{W} - \mathbf{Y}\|_F^2 \tag{21}$$

||.||F denotes Frobenius matrix norm. Equation (22)'s optimization function is clearly convex, smooth and the ideal weight matrix W is W∗ = (XXT) −1XY. In real-world applications, however, XXT is not always invertible. To achieve this, the smooth regularisation term is given a conventional objection function, such as a 2-norm.

$$\min_{\mathbf{W}} \|\mathbf{X}^T\mathbf{W} - \mathbf{Y}\|_F^2 + \rho \parallel \mathbf{W} \parallel_2^2 \tag{22}$$

We take the derivative of each row wi(1 ≤ i ≤ m)and then set it to zero, as shown in Eq. (23):

$$\mathbf{XX}^T\mathbf{w}_i - \mathbf{Xy}^i + \rho_1\mathbf{D}_i\mathbf{w}_i + \rho_2\grave{\mathbf{D}}\mathbf{w}_i + \rho_3\mathbf{XLX}^T\mathbf{w}_i = 0 \tag{23}$$

where Di(1 ≤ i ≤ m) is a diagonal matrix with $\frac{1}{2|w_{ki}|}$ as the kth diagonal element and D is a diagonal matrix with $\frac{1}{2||\mathbf{w}^k\|_2}$ as the kth diagonal element. Equation (24) is then modified as follows:

$$\mathbf{w}_i = \left(\mathbf{XX}^T + \rho_1\mathbf{D}_i + \rho_2\grave{\mathbf{D}} + \rho_3\mathbf{XLX}^T\right)^{-1}\mathbf{Xy}^i \tag{24}$$

PROOF. According to Step 2 in Algorithm 1by, eq. (25)

$$\begin{aligned}\mathbf{W}^{(t+1)} = \quad & \min_{\mathbf{W}} \mathrm{Tr}\left((\mathbf{X}^T\mathbf{W} - \mathbf{Y})^T(\mathbf{X}^T\mathbf{W} - \mathbf{Y})\right) \\ & + \rho_1 \textstyle\sum_{i=1}^m \mathbf{w}_i^T\mathbf{D}_i^{(t)}\mathbf{w}_i + \rho_2 \mathrm{Tr}\ \mathbf{W}^T\grave{\mathbf{D}}^{(t)} + \rho_3\mathbf{XLX}^T\end{aligned} \tag{25}$$

$$\begin{aligned}\mathrm{Tr}\ &\left(\left(\mathbf{X}^T\mathbf{W}^{(t+1)} - \mathbf{Y}\right)^T\left(\mathbf{X}^T\mathbf{W}^{(t+1)} - \mathbf{Y}\right)\right) + \rho_1\textstyle\sum_{i=1}^m\left(\mathbf{w}_i^{(t+1)}\right)^T\mathbf{D}_i^{(t)}\mathbf{w}_i^{(t+1)} \\ &+ \rho_2\mathrm{Tr}\left(\mathbf{W}^{(t+1)}\right)^T\grave{\mathbf{D}}^t\mathbf{W}^{(t+1)} + \rho_3\mathbf{XLX}^T \\ &\le \mathrm{Tr}\left(\left(\mathbf{X}^T\mathbf{W}^{(t)} - \mathbf{Y}\right)^T\left(\mathbf{X}^T\mathbf{W}^{(t)} - \mathbf{Y}\right)\right) + \rho_1\textstyle\sum_{i=1}^m\left(\mathbf{w}_i^{(t)}\right)^T\mathbf{D}_i^{(t)}\mathbf{w}_i^{(t)} \\ &+ \rho_2\mathrm{Tr}\left(\mathbf{W}^{(t)}\right)^T\grave{\mathbf{D}}^i\mathbf{W}^{(t)} + \rho_3\mathbf{XLX}^T \\ &\Rightarrow \mathrm{Tr}\left(\left(\mathbf{X}^T\mathbf{W}^{(t+1)} - \mathbf{Y}\right)^T\left(\mathbf{X}^T\mathbf{W}^{(t+1)} - \mathbf{Y}\right)\right) \\ &+ \rho_1\textstyle\sum_{i=1}^n\sum_{j=1}^n\left(\frac{\left(\mathbf{w}_{ij}^{(t+1)}\right)^2}{2\left\|\mathbf{w}_{ij}^{(t)}\right\|} - \left\|\mathbf{w}_{ij}^{(t+1)}\right\| + \left\|\mathbf{w}_{ij}^{(t+1)}\right\|\right) \\ &+ \rho_2\textstyle\sum_{k=1}^d\left(\frac{\parallel\left(\mathbf{w}^{(t+1))^k}\right.\parallel_2^2}{2\parallel\left(\mathbf{w}^{(t))^k}\right.\parallel_2} - \left\|\left(\mathbf{w}^{(t+1)}\right)^k\right\|_2 + \left\|\left(\mathbf{w}^{(t+1)}\right)^k\right\|_2\right) \\ &+ \rho_3\mathbf{XLX}^T \le \mathrm{Tr}\left(\mathbf{X}^T\mathbf{W}^{(t)} - \mathbf{Y}\right)^T\left(\mathbf{X}^T\mathbf{W}^{(t)} - \mathbf{Y}\right)\Big) \\ &+ \rho_1\textstyle\sum_{i=1}^n\sum_{j=1}^n\left(\left\|\mathbf{w}_{ij}^{(t)}\right\| + \frac{\left(\mathbf{w}_{ij}^{(t)^2}\right)^2}{2\left\|\mathbf{w}_{ij}^{(t+t)}\right\|} - \left\|\mathbf{w}_{ij}^{(t)}\right\|\right)\end{aligned}$$

$$+\rho_2\sum_{k=1}^{d}\left(\left\|\left(\mathbf{w}^{(t)}\right)^k\right\|_2+\frac{\left\|\left(\mathbf{w}^{(t)}\right)^k\right\|_2^2}{2\|(\mathbf{w}^{(i)})^k\|_2}-\left\|\left(\mathbf{w}^{(t)}\right)^k\right\|_2\right)+\rho_3\mathbf{XLX}^T$$

$$\Rightarrow \mathrm{Tr}\left(\left(\mathbf{X}^T\mathbf{W}^{(t+1)}-\mathbf{Y}\right)^T\left(\mathbf{X}^T\mathbf{W}^{(t+1)}-\mathbf{Y}\right)\right)+\rho_1\sum_{i=1}^{d}\sum_{j=1}^{m}\left\|\mathbf{w}_{ij}^{(t+1)}\right\|$$

$$+\rho_2\sum_{K=1}^{d}\left\|\left(\mathbf{w}^{(t+1)}\right)^k\right\|_2$$

$$\leq \mathrm{Tr}\left(\left(\mathbf{X}^T\mathbf{W}^{(t)}-\mathbf{Y}\right)^T\left(\mathbf{X}^T\mathbf{W}^{(t)}-\mathbf{Y}\right)\right)+\rho_1\sum_{i=1}^{d}\sum_{j=1}^{m}\left\|\mathbf{w}_{ij}^{(t)}\right\|$$

$$+\rho_3\mathbf{XLX}^T$$

$$+\rho_2\sum_{k=1}^{d}\left\|\left(\mathbf{w}^{(t)}\right)^k\right\|_2+\rho_3\mathbf{XLX}^T$$

Algorithm for Dim_Red_KNN:
Input X,Y *Output: switch task do* *Class labels;* *Case 1* *Endsw* *Case 2* *Forecast value;* *Case 3* *Imputation value* *Endsw* *Enhancing equation (6) to get optimal solution W* *Normalizing X and Y* *Finding optimal k value for test data based on W* *Case 1* *Switch task do* *Finding class labels via majority rule* *Case 2* *End sw* *Finding prediction value via equation (9);* *Case 3* *Endsw* *Finding imputation value via equation (9)'* *Endsw*

Recursive elimination-based convolutional neural network (RE_CNN) based classification:

For creating a strong joining method for sentiment analysis, the suggested model integrates both CNN as well as RNN. It consists of 4 layers: embedding, convolutional, recursive, and classification, as illustrated in figure 2.

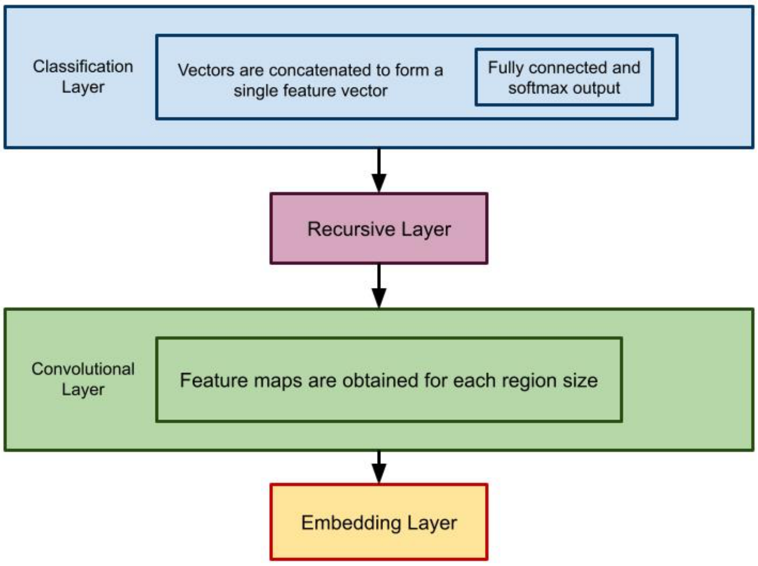


Figure 2: RE_CNN architecture

The Euclidean distances between a point xi and a point xj in a vector v of N high-dimensional points x1, x2, xn are translated into a conditional probability Pj|i, which gives the similarity between point xi and point xj. As a result, the conditional probability pj|i increases for data points that are close together, whereas pj|i decreases for data points that are far apart. The conditional probability Pj|i can be expressed mathematically as eq. (26).

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|_2 / 2\sigma_i{}^2\right)}{\sum_{k\neq i} \exp\left(-\|x_i - x_j\|^2 / 2\sigma_i{}^2\right)} \tag{26}$$

where σi is the Gaussian variance centred at the location xi. Because the method is focused on modeling pairwise similarities. Similarly, the conditional probability representing the similarities of the map points yi to yj and indicates by qj|i is given by $q_{ji} = \frac{\exp\left(-\|y_i - y_j\|_2\right)}{\sum_{k\neq i} \exp\left(-\|y_i - y_j\|^2\right)}$ in low-dimensional equivalents yi and yj of higher dimensionality xi and xj.

Because this method only involves modelling pairwise similarities, the conditional probability qi|i is similarly set to zero (qi|i = 0). In t-SNE, this is done repeatedly for a given cost function C using the gradient descent method, so that eq. (27)

$$C = \sum_i KL(P_i \parallel Q_i) = \sum_i \sum_j p_{ji} \log \frac{p_{iji}}{q_{iji}} \tag{27}$$

$KL(P_i \parallel Q_i)$ is Kullback-Leibler divergence function of $P_i \parallel Q_i$ [45]. Kullback-Leibler divergence $KL(P_i \parallel Q_i)$ between them is given as eq. (28)

$$KL(P_i \parallel Q_i) = -\sum_{x \in X} P_i(x) \log\left(\frac{Q_i(x)}{P_i(x)}\right) KL(P_i \parallel Q_i) = \sum_{x \in X} P_i(x) \log\left(\frac{P_i(x)}{Q_i(x)}\right) \quad (28)$$

Expectation of the logarithmic difference between probabilities Pi and Qi is determined by Equation (29) above. In P_i and Q_i, it can produce any continuous, random variable x as

$$KL(P_i \parallel Q_i) = \int_{-\infty}^{\infty} p_{ji}(x) \log\left(\frac{p_{ji}(x)}{q_{ji}(x)}\right) dx$$

$$\Phi(\mathbf{w}, \xi) = \frac{1}{2} \parallel \mathbf{w} \parallel^2 + C\sum_{i=1}^{n} \xi_i \quad (29)$$

Subject to: $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0$

Where y_i is corresponding to target, $y_t = \{\pm 1\}, i = 1, \dots, m$.

Optimization issue is solved in a dual issue by eq. (30):

$$W(\boldsymbol{\alpha}) = \sum_{i=1}^{\mathbb{W}} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{\mathrm{m}} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) \quad (30)$$

Subject to: 1) $0 \leq \alpha_i \leq C, i = 1, \dots, m$

2) $\sum_{i=1}^{m} \alpha_i y_i = 0$

where α_i are the Lagrange coefficients.

The probability densities of P_i and Q_i are pj|i and qj|i. When P_i and Q_i are calculated over continuous sets X and P, Kullback–Leibler divergence function can be rewritten as eq (31)

$$KL(P_i \parallel Q_i) = \int_X \log\left(\frac{dP_i}{dQ_i}\right) dP_i \quad (31)$$

where $\frac{dP_i}{dQ_i}$ in (7) is known as Radon–Nikodym derivative of Pi with respect to Qi . Utilizing the chain rule,

$$KL(P_i \parallel Q_i) = \int_X \log\left(\frac{dP_i}{dQ_i}\right) \frac{dP_i}{dQ_i} dQ_i \quad (32)$$

Algorithm for RE_CNN:

$\boldsymbol{F}$: = set of ranked features = $\emptyset$
$\boldsymbol{S}$: = score of a criterion function for a certain
$\boldsymbol{R}$: = set of remaining features = $\{1,2,3, \dots, n\}$
Number of remaining features, $k := n$,
Step 1: Train a SVM, evaluate score

$$\boldsymbol{S}_1 = \boldsymbol{S}\{(\boldsymbol{x}_i(\boldsymbol{R}), \boldsymbol{y}_i)\}$$

Step 2: Evaluate w_i and rank features based on values of w_i, $\mathfrak{I} = \{\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_k\}$.
Step 3: Set $k = k - 1$ and $j = 1$.
Step 4: create a new feature set R_l by eliminating feature f_j from $\boldsymbol{R}$.
Step 5: Train a SVM on samples with remaining features and evaluate score $S_2 = S\{(x_i(R_1), y_i)\}$
Step 6: If $S_1 < S_2$,

$$\boldsymbol{F} = \boldsymbol{F} \cup \boldsymbol{f}_j, \ = 1$$

go to step 7
else if $j = k$,

$$\boldsymbol{F} = \boldsymbol{F} \cup \boldsymbol{f}_1, \boldsymbol{R} = \boldsymbol{R} - \boldsymbol{f}_1$$

go to step 7
else

$$j = j + 1$$

go to step 4.
Step 7: Repeat steps $1 - 6$ until $\boldsymbol{R} = \emptyset$
start $i \leftarrow 1$
$\sigma =$ variance of Gaussian
For every pair of points x_i and x_j in ∇ do
If $x_i = x_j$ then

$$P_{\text{in}} = 0$$

For every counterpart pair of points y_j and y_j in low-dimension do
If $y_i = y_j$ then
Compute $q_N = \frac{\exp(-\|y_- - y_1\|_2)}{\sum \exp(-|y_2 - y_1/|_2)}$
End if
End for
Write $/SON_4$ file ← class, image_name, $y_i, 3/$
return (JSON file)
If corresponding y_i, and y_j are NOT in $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)$ then
Oulliers[[] += image_name
Else
Continue
Return (Oullers[]])
Call train_vgBl60
Call t-SNE0

Input data is passed through several levels of processing, such as normalization, feature deletion, and classification, as shown in Figure 3.

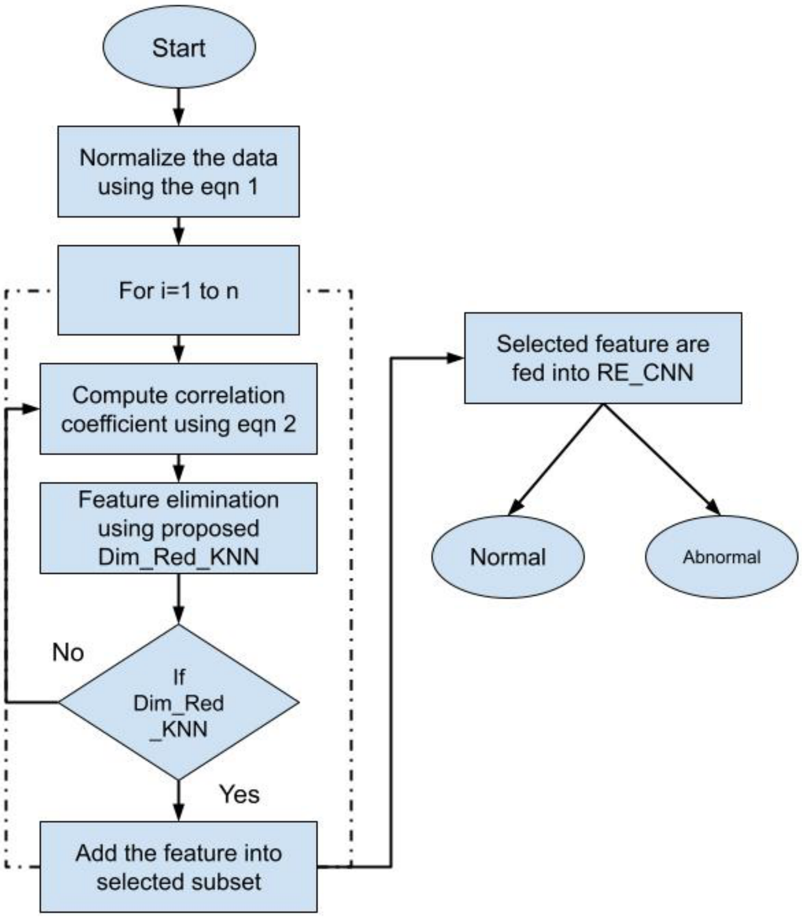


Figure 3: Proposed feature elimination and classification.

Performance analysis:

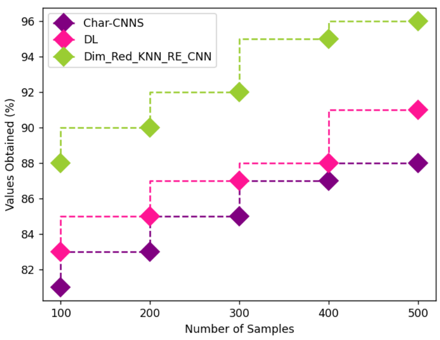
Models of Dim_Red_KNN_RE_CNN were tested on the testing dataset after training using three metrics to help researchers choose the best model. We first generated responses using the proposed technique using 2770 postings in the testing dataset. The postings in the testing set served as model "prompts," meaning the Dim_Red_KNN_RE_CNN models generated responses in response to those prompts.

Evaluation	**Techniques**	**Accuracy**	**Precision**	**Recall**	**RMSE**	**Computational time**
Student feedback classification	**Char-CNNS**	88	85	81	88	77
	DL	91	89	85	82	72
	Dim_Red_KNN_RE_CNN	96	90	88	78	69

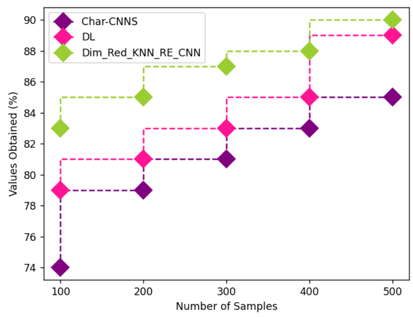
Course Recommendation	**Char-CNNS**	88	88	83	82	77
	DL	92	90	89	78	69
	Dim_Red_KNN_RE_CNN	96	92	93	68	65

Table 1: Evaluation analysis between proposed and existing techniques

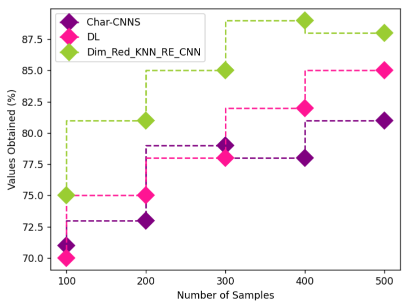
The comparative analysis has been shown in table-1 based on the evaluation of student feedback classification and course recommendation analysis in terms of accuracy, precision, recall, RMSE, and computational time. Here the proposed techniques compared Char-CNNS and DL with the proposed technique in higher vocational education analysis based on course recommendation and student feedback.



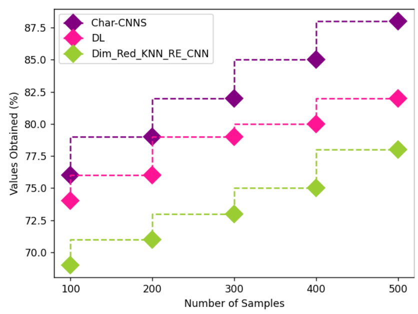
(a) Accuracy



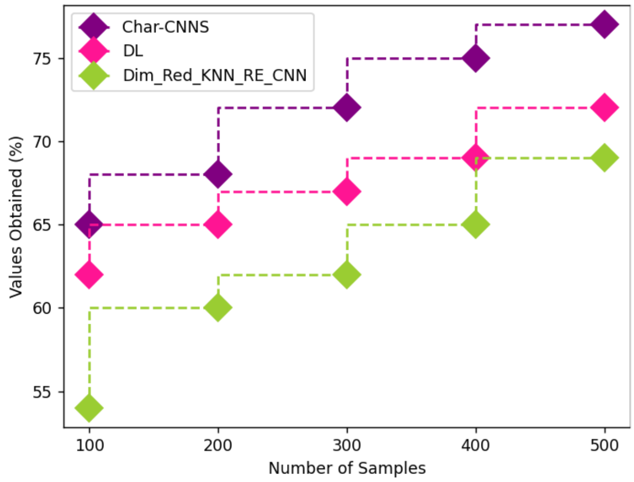
(b) Precision



(c) Recall

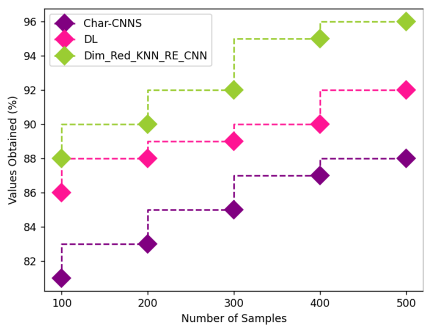


(d) RMSE

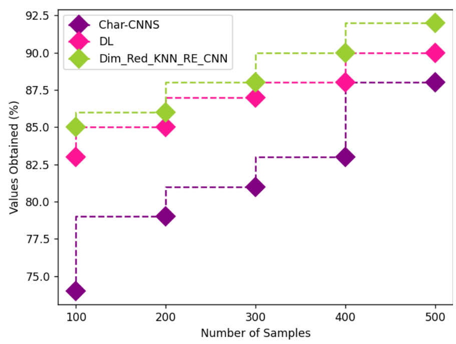


(e) Computational Time

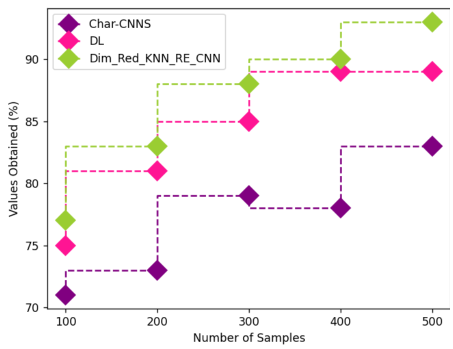
Figure 4: Evaluation analysis based on Student feedback classification in terms of (a) accuracy, (b) precision, (c) recall, (d) RMSE, (e) Computational time



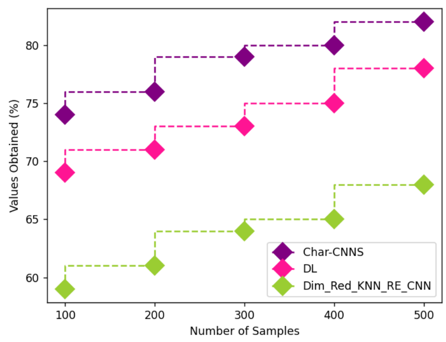
(a) Accuracy



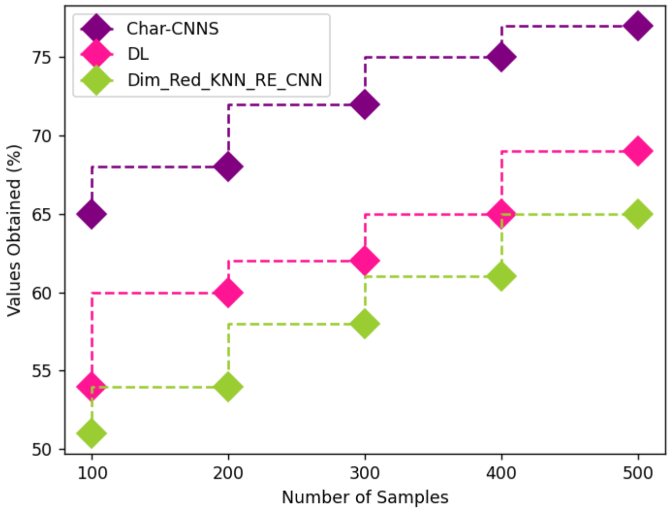
(b) Precision



(c) Recall



(d) RMSE



(e) Computational Time

Figure 5: Evaluation analysis based on course recommendation in terms of (a) accuracy, (b) precision, (c) recall, (d) RMSE, (e) Computational time

The above Figures 4 and 5 show the proposed technique evaluation in Higher Vocational Education in English teaching. The evaluation has been carried out based on course selection, recommendation, and students' feedback in terms of accuracy, precision, recall, RMSE, and computational time. The selection recommendation for the course has been evaluated based on the students' feedback analysis. For student feedback-based evaluation, the proposed technique obtained an accuracy of 96%, precision of 90%, recall of 88%, RMSE of 78%, and computational time of 69%. From this student's feedback evaluation, course recommendation has been given and its evaluation obtained by the proposed technique in terms of accuracy as 96%, precision of 92%, recall of 93%, RMSE of 68%, and computational time of 65%. The proposed technique obtained optimal results in Higher Vocational Education based on English teaching in course selection from this evaluation.

5 Conclusion

This research proposes novel English teaching techniques based on artificial intelligence in course selection based on students' feedback. The dataset has been collected based on the students' feedback on courses for Higher Vocational Education in English teaching. This dataset has been processed to remove invalid data, missing values, and noise. The processed data features have been dimensionality reduction integrated with K-means Neural network. And the extracted features have been classified with higher accuracy using recursive elimination-based convolutional neural network. Based on this feedback data classification, recommendations for courses in Higher Vocational Education in English teaching were suggested. Experimental analysis shows various student feedback dataset validations and trainings with 96% accuracy, 92% precision, 93% recall, 68% RMSE, and 65% computational time.

References

[Bernardo et al., 2021] Bernardo, A. B., Cordel, M. O., Lucas, R. I. G., Teves, J. M. M., Yap, S. A., & Chua, U. C.: "Using machine learning approaches to explore non-cognitive variables influencing reading proficiency in English among Filipino learners", *Education Sciences,* 11(10) (2021), 628.

[Bucea-Manea-Tonis et al., 2020] Bucea-Manea-Tonis, R.; Bucea-Manea-Tonis, R.; Simion, V.E.; Ilic, D.; Braicu, C.; Manea, N.: "Sustainability in Higher Education: The Relationship between Work-Life Balance and XR E-Learning Facilities", *Sustainability* 12(2020), 5872.

[Chen and Jeong, 2007] Chen, X. W., & Jeong, J. C.: "Enhanced recursive feature elimination", *Sixth International Conference on Machine Learning and Applications (ICMLA 2007),* (2007), (pp. 429-435).

[Dhawan and Batra, 2021] Dhawan, S.; Batra, G: "Artificial Intelligence in Higher Education: Promises, Perils, and Perspective", *OJAS: New Delhi, India*, (2021), pp. 11–22. 35.

[Dong and Tsai, 2021] Dong, H., & Tsai, S.B.: "An empirical study on application of machine learning and neural network in English learning", *Mathematical Problems in Engineering*, (2021).

[Du, 2021] Du, H.: "An english network teaching method supported by artificial intelligence technology and WBIETS system", *Scientific Programming,* (2021).

[Farahabadi et al., 2021] Badakhshan Farahabadi, F., Fathi Vajargah, K., & Farnoosh, R.: "Dimension Reduction Big Data Using Recognition of Data Features Based on Copula Function and Principal Component Analysis", *Advances in Mathematical Physics*, 1(2021).

[Ilic et al., 2020] Ilic, M.; Popovic Ševic, N.; Ristic, B.: "Integrating Human Resources and Customer Relationship Management for Student Satisfaction in Higher Education: Gaining a Competitive Advantage", *International May Conference on Strategic Management (IMCSM), Bor, Serbia*, 25–27 , (2020), pp. 501–508.

[Mosavi et al., 2020] Mosavi, A.; Faghan, Y.; Ghamisi, P.; Duan, P.; Ardabili, S.F.; Salwana, E.; Band, S.S.: "Comprehensive review of deep reinforcement learning methods and applications in economics", *Mathematics,* 8(2020), 1640.

[National Body, 2021] National Body for Accreditation and Quality Assurance in Higher Education (NAT) Available online: https://www.nat.rs/ (2021).

[Perez and Tah, 2020] Perez, H., & Tah, J. H.: "Improving the accuracy of convolutional neural networks by identifying and removing outlier images in datasets using t-SNE", *Mathematics*, 8(5) (2020), 662.

[Radic et al., 2020] Radic, G.; Ristic, B.; Andelic, B.; Kuleto, V.; Ilic, M.: "E-learning Experiences of higher education institutions in the Republic of Serbia during COVID-19, content analysis and case study ITS Belgrade. Sustainable Economic Development and Advancing Education Excellence in the era of Global Pandemic", *Proceedings of the 36th International Business Information Management, Granada, Spain*, 4–5 (2020), ISBN 978-0-9998551-5-7.

[Reddy et al., 2020] Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T.: "Analysis of dimensionality reduction techniques on big data", *IEEE Access*, 8(2020), 54776-54788.

[Sadr et al., 2019] Sadr, H., Pedram, M. M., & Teshnehlab, M.: "A robust sentiment analysis method based on sequential combination of convolutional and recursive neural networks", *Neural Processing Letters*, 50(3), (2019), 2745-2761.

[Sarveniazi, 2014] Sarveniazi, A.: "An actual survey of dimensionality reduction", *American Journal of Computational Mathematics*, (2014).

[Taylor et al., 2021] Taylor, D.L.; Yeung, M.; Bashet, A.Z.: "Personalised and Adaptive Learning", *Innovative Learning Environments in STEM Higher Education. SpringerBriefs in Statistics; Ryoo, J., Winkelmann, K., Eds.; Springer: Cham, Switzerland*, (2021).

[Venkateswaran, 2022] N. Venkateswaran, "Hybridized Wrapper Filter Using Deep Neural Network for Intrusion Detection", *Computer Systems Science & Engineering* DOI:10.32604/csse.2022.021217

[Xiao and Yi, 2020] Xiao, M., & Yi, H.: "A practical and efficient multi-assessment system for vocational teaching based on machine learning", *The International Journal of Electrical Engineering & Education*, 0020720920940573.

[Zhang et al., 2017] Zhang, S., Li, X., Zong, M., Zhu, X., & Cheng, D.: "Learning k for knn classification", *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(3)(2014), 1-19.