Using Principal Component Analysis to support students' performance prediction and data analysis

Vinicius R. P. Borges[1], Stéfany L. Esteves[2], Patrícia De Nardi Araújo[2], Lucas C. Oliveira[2], Maristela Holanda[1]

[1]Department of Computer Science – University of Brasilia
Brasília, DF, Brazil

{viniciusrpb,mholanda}@unb.br,

[2]Department of Computer Science – Federal University of Lavras
Lavras, MG, Brazil

{lealesteves,patynardi,lucacharles}@sistemas.ufla.br

Abstract. *We propose a method based on Principal Component Analysis (PCA) for predicting students' performances and for identifying relevant patterns concerning their characteristics. The proposed method allowed us to study the predictive capability of students' performances and the effectiveness of PCA for interpreting patterns in educational data. The proposed method was validated using two public datasets describing students achievements, as well as their social and personal characteristics. Experiments were conducted by comparing the predictive performances between the datasets presenting high and reduced dimensions. The results reported that PCA retained relevant information of data and was useful for identifying implicit knowledge in students' data.*

1. Introduction

In the last decade, Educational Data Mining (EDM) has emerged with techniques and strategies to process, interpret and obtain useful and implicit knowledge on educational data [Baker 2014]. One of the most traditional tasks in EDM refers to the performance prediction of students regarding their learning, achievements and final outcomes [Shahiri and Husain 2015] [Baradwaj and Pal 2012]. Basically, it concerns on proposing models to infer a specific pattern of the data from other known data patterns. As in Knowledge Discovery in Databases (KDD) processes, prediction methodologies comprise some of the following steps: data preprocessing, data mining techniques (feature selection, classification, clustering and regression) and the interpretation of results, leading to the obtaining of implicit and relevant knowledge. In this research, we are interested in understanding the factors that affect the performances of students in the educational environment, as well as predicting their performances.

The formulation of an EDM prediction method depends on the datasets characteristics. Specifically, educational datasets can present several attributes of different types (numerical, binary, categorical, interval, among others) once some of them are generally collected by means of questionnaires. Such questionnaires contain questions in which answers are formulated in the form of single choices presenting specific values. This requires transforming the categorical attributes to binary ones that identify specific categorical values), thus increasing the data dimensionality. However, the high dimensionality

DOI: 10.5753/cbie.sbie.2018.1383

of datasets strongly affects the classification performance in prediction tasks and compromise data analysis due to the curse of dimensionality [Tan et al. 2005].

Some state-of-art methodologies in EDM for predicting students' performances choose manually the attributes of interest or employ automatic feature selection techniques to reduce the dimensionality of data [Baradwaj and Pal 2012]. Moreover, feature selection is implicit in some classification models (for example, decision trees) or are employed to identify the most relevant attributes affecting the final outcomes [Asif et al. 2017]. In addition, in order to enrich the analysis of educational data, implicit knowledge can also be obtained by performing clustering to group similar data instances according to their attributes [Dutt et al. 2015]. However, running additional clustering for data analysis invariably demands more computational processing.

The aforementioned limitations motivated us to study the applicability of a feature space transformation Principal Component Analysis (PCA) [Jolliffe 1986] for simultaneous dimensionality reduction and data analysis in educational datasets. PCA is a well-known technique that captures the most variability of data using few dimensions. It has been extensively applied for reducing the dimensionality of data, as well as to analyze data patterns in several knowledge domains, such as medicine [Polat and Güneş 2007], genetics [Duforet-Frebourg et al. 2015], social networks [Linden et al. 2017], sustainable development [Abou-Ali and Abdelfattah 2013], among others. However, to best of our knowledge, little research has been carried out to analyze patterns in educational data for prediction tasks. In order to fill this gap regarding the application of PCA in educational data mining, we formulated a method for predicting students' performances that incorporates PCA for dimensionality reduction and for data analysis. PCA is employed prior to the classification process, which considers two well-known classification models: Support Vector Machines and Naive Bayes. Two datasets related to students' final performances in secondary school are considered to validate the proposed method.

The main contribution of this paper refers to the employment of PCA for interpreting patterns in educational datasets and for dimensionality reduction in prediction tasks. For that purpose, we devised a method based on data mining, in which PCA is used prior to the classification step for dimensionality reduction, instead of the frequent use of traditional feature selection techniques. At the same time, we derive information from PCA in order to extract relevant knowledge from students' data and relations among the variables that can affect students' performances.

This paper is organized as follows: Section 2 presents previous studies on students' performance prediction in EDM. Section 3 describes the devised method which allows to obtain relevant knowledge from students' data and for reducing the dimensionality to support performance prediction tasks. Section 4 reports the experimental results, discuss the performance of proposed method and the discovered patterns in the considered datasets. Section 5 presents the final considerations.

2. Related work

Educational data mining (EDM) allows specialists in education and researchers to identify relevant knowledge in educational datasets [Romero and Ventura 2010]. Basically, a EDM methodology receives raw educational data as input and applies learning analytics, data mining and machine learning to output useful information in an efficient fashion when compared to manual and conventional data analysis. Baker et al. [Baker 2014] and Sahiri et al. [Shahiri and Husain 2015] review several methodologies employing data

mining for prediction students performances.

One of the pioneer researches on students' performance prediction, Cortez and Silva [Cortez and Silva 2008] investigated the students' performances in two Portuguese secondary schools using data mining techniques. Students' data were collected by means of questionnaires and school reports. The first step consisted of data preprocessing due to the presence of nominal attributes and some missing values. Three configurations that take into account past school grades, demographic and social attributes were tested for predicting the student's performance. Five classifiers and a regression technique were used for predicting the students' performances. Experimental results reported a high predictive confidence if the first and/or secondary school period grades are known, i.e. the student's performance is highly influenced by past performances.

Costa et al. [Costa et al. 2017] compared educational data mining techniques to predict in an early stage the students that are likely to fail in introductory programming courses. It is also investigated whether data preprocessing techniques and the peculiarities of classification models influence the students' failure prediction. The experiments comprised the application of two educational datasets collected from introductory programming courses of a Brazilian public university. The results showed that the prediction methods were able to identify prior to the beginning of the school year the students that are likely to fail at the end of the course. Moreover, some preprocessing techniques improved the accuracy on the prediction task. The best classification model as the support vector machine, which outperformed other traditional classifiers (artificial neural networks, Naive Bayes and K-Nearest neighbors).

The aforementioned state-of-art researches analyze students' data by means of clustering techniques or by interpreting the prediction results. Moreover, they perform automatic feature selection to identify the most relevant variables affecting students' performances. This has motivated us to study the well-known Principal Component Analysis (PCA), which is a popular technique for dimensionality reduction tasks and for data analysis [Jolliffe 1986]. Its successful previous applications in other knowledge domains motivated us to concentrate efforts in studying its applicability on educational datasets.

A method based on educational data mining was devised for that purpose and it is detailed in the next section.

3. Proposed method

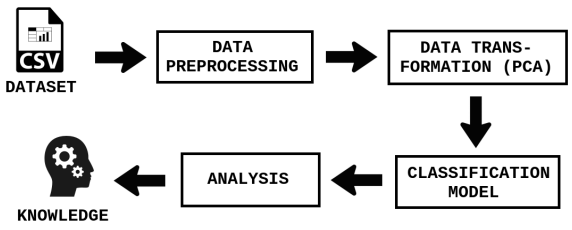


Figure 1. Flowchart of the proposed method.

The proposed method is depicted by the flowchart in Figure 1. The method receives as input the raw dataset and outputs the performance predictions with some information regarding the data transformation step.

3.1. Datasets

The datasets [1] [Cortez and Silva 2008] describe student achievements in secondary education of two Portuguese schools concerning two subjects: Portuguese and Math. The data attributes comprise student grades, their demographic, social, financial, personal and academical records. Such data was collected by means of school reports and questionnaires. The first dataset (Dataset I) contains 649 students of the Portuguese subject and it is described by 33 attributes, as reported on Table 1. The second dataset (Dataset II) is characterized by the same attributes and refers to final achievements of students in the Math subject.

The final grade (`G3` on Table 1) is taken as the class attribute, once we are interested in predicting students' performances. In their studies, Cortez and Silva reported that the class attribute `G3` presents high correlation with the attributes `G2` and `G1`. This occurs because `G3` is the final year grade (issued at the 3rd period), while `G1` and `G2` correspond to the first and second period grades.

Table 1. Data attributes: students of two secondary schools in Portugal. G3, the final grade, is the class attribute.

attribute name	description
school	student's school ("Gabriel Pereira" or "Mousinho da Silveira")
sex	student's sex (female or male)
age	student's age (from 15 to 22)
address	student's home address type ("urban" or "rural")
famsize	family size ("less or equal to 3" or "greater than 3")
Pstatus	parent's cohabitation status ("living together" or "apart")
Medu	mother's education (from 0 to 3[2])
Fedu	father's education (from 0 to 3)
Mjob	mother's job [3]
Fjob	father's job
reason	to choose school (close to 'home', school 'reputation', 'course' preference or 'other')
guardian	student's guardian {'mother', 'father' or 'other'}
traveltime	travel time to school ("<15 min.","15 to 30 min.", "30 min. to 1 hour", ">1 hour")
studytime	weekly study time ("<2 hours", "2 to 5 hours", "5 to 10 hours", ">10 hours")
failures	number of past class failures (n if $1 \leq n < 3$, else 4)
schoolsup	extra educational support ("yes" or "no")
famsup	family educational support ("yes" or "no")
paid	extra paid classes within the course subject ("yes" or "no")
activities	extra-curricular activities ("yes" or "no")
nursery	attended nursery school ("yes" or "no")
higher	wants to take higher education ("yes" or "no")
internet	Internet access at home ("yes" or "no")
romantic	with a romantic relationship ("yes" or "no")
famrel	quality of family relationships (from 1 - very bad to 5 - excellent)
freetime	free time after school (from 1 - very low to 5 - very high)
goout	going out with friends (from 1 - very low to 5 - very high)
Dalc	workday alcohol consumption (from 1 - very low to 5 - very high)
Walc	weekend alcohol consumption (from 1 - very low to 5 - very high)
health	current health status (from 1 - very bad to 5 - very good)
absences	number of school absences (from 0 to 93)
G1	first period grade (from 0 to 20)
G2	second period grade (from 0 to 20)
G3	final grade (from 0 to 20)

[1] Dataset available at `http://archive.ics.uci.edu/ml/datasets/Student+Performance`

[2] 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education

[3] 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')

3.2. Data preprocessing

The goal of this step is to prepare raw data to a suitable form enabling the application of the machine learning techniques. A preliminary analysis in data instances and attributes of Datasets I and II showed that some form of preprocessing is required, since attributes are from different types (binary, numerical and nominal). First, the data instances with missing values are removed, so that we can handle consistent data.

After that, as nominal attributes are present, we transform each one to dummy variables [Filmer and Pritchett 2001], which can be defined as binary attributes that can take the value "0" or "1" to indicate the absence or presence of a specific categorical value. For instance, the attribute `guardian` can possess the values "mother", "father" and "other". Thus, the binary attributes `guardian=''father''`, `guardian=''mother''` and `guardian=''other''` were created as a replacement to the nominal attribute `guardian`. Finally, the class label `G3` is discretized in a way that we can identify students that were approved or failed at the end of the scholar year. Such attribute is discretized as follows:

$$\texttt{new G3} = \begin{cases} Approved, \text{if} \quad \texttt{G3} \geq 10 \\ Failed, \text{otherwise} \end{cases} \tag{1}$$

3.3. Data transformation

The key idea of PCA consists of performing a linear mapping of the data in a high dimensional space to a lower-dimensional space, in which the data's variance is maximized. First, data instances are centered by subtracting the mean of each attribute to reduce the dependence on scale. The covariance matrix from the centered data (in some occasions, correlation matrix is adopted, but in this case, data should be standardized) is obtained. After that, we perform the eigendecomposition of the covariance matrix, resulting in the obtaining of eigenvectors and eigenvalues. The greatest eigenvalues are used to select the correspondent eigenvectors (principal components), defining the transformed data. Typically, the eigenvectors associated with the higher eigenvalues retain the most variability of data.

For formalization purposes, let $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$ be the dataset, in which each $\mathbf{x}_i$ refers to a data instance. An instance $\mathbf{x}_i$ described by D attributes is defined by the feature vector $[x_{i,1}, ..., x_{i,D}]$. PCA can be summarized according to the following steps:

1. Center the data by subtracting the values of each data instance $\mathbf{x}_i$ by the mean μ according to Eq. (2):
$$z_i = \mathbf{x}_i - \mu \tag{2}$$
2. Knowing that $Z = \{z_1, ..., z_N\}$, compute the covariance matrix, such as:
$$\Sigma = Z^T Z \tag{3}$$
3. Compute the eigenvalues $\lambda = \lambda_1, ..., \lambda_D$ and the eigenvectors $\mathbf{V}$ of the covariance matrix by means of a spectral decomposition [Wall et al. 2003]:
$$\Sigma = \mathbf{V} A \mathbf{V}^{-1} \tag{4}$$

in which $\mathbf{V}$ is the matrix of eigenvectors and A is a diagonal matrix with eigenvalues on the diagonal and zero elsewhere. The eigenvalues on the diagonal of A correspond the columns in $\mathbf{V}$, so that the first element of A is λ_1 and the associated eigenvector is the first column of $\mathbf{V}$, and so on.

4. Sort the eigenvalues in descending order and select the k eigenvectors associated to the k largest eigenvalues, in which k is the number of dimensions of the reduced space (low-dimensional space).
5. Construct the projection matrix from the selected k eigenvectors. It is worth noting that the eigenvectors configure the principal components, which define a linear transformation from the original attribute space to a new space in which attributes are uncorrelated:

$$PC_l = c_{l,1}x_1 + c_{1,2}x_2 + ... + c_{1,D}x_D \quad (5)$$

in which PC_l denotes the l-th principal component (PC), $\{x_1, ..., x_D\}$ are the data attributes and $\{c_{l,1}, ...c_{l,D}\}$ refer to the coefficients of PC_l.
The outputs from PCA, i.e., the eigenvalues, the principal components and their coefficients are useful for analyzing patterns in data. In many EDM tasks, the identification of the main factors affecting students' performances is extremely important. As the original representation of the data (original attributes) were transformed to principal components, we analyze the coefficients of principal components and the amount of explained variance to obtain implicit knowledge from educational data. Such coefficients express the correlation of each variable to the principal component and its signal and magnitude are taken into account for interpreting the data patterns.

In Section 4, we present some analysis concerning the selection of the principal components that defines the transformed feature space, as well as to interpret relevant patterns on them for each dataset.

3.4. Classification models

The first classification model refers to the Support Vector Machines (SVM) [Witten et al. 2016]. We chose SVM due to its previous successful applications for students' performance predictions. SVM represents a training set as a high-dimensional space in which instances are represented as points. The main principle of SVM is to find a decision boundary based on a maximal margin hyperplane that separates linearly or nonlinearly the training set into two classes. The training step attempts to compute such hyperplane by maximizing its distance to the nearest data point on each side. SVM was adjusted with a radial basis function (RBF) with the parameter width set to 2.0, since it yielded the best correct classification results among the values $\{0.01, 0.1, 0.5, 1.0, 2.0\}$. RBF has been chosen due to its capability of handling nonlinear relations between class labels and attributes, besides requiring reduced parameterization and resulting in high generalization.

The other classification model refers to the Naive Bayes, which is a based on a probabilistic approach. Naive Bayes applies the Bayes theorem, which requires the independence between features. A multivariate normal distribution is used as the probability distributions. Naive Bayes has been chosen since it is a widely employed classification model in EDM, specially in performance prediction tasks.

4. Experimental results

The experiments were conducted using the Weka 3.9.1 environment [Witten et al. 2016]. In our experiments, we compared the performances of classifiers using the original high dimensional datasets, i.e., the dataset generated after the preprocessing step, with three configurations of the dataset in its reduced form: using two dimensions, five dimensions and ten dimensions.

The strategy for evaluating a particular classification model was the hold-out validation, in which 66% of the dataset instances are used for training, while the remaining instances (34%) are used for test. The outcomes of the test instances generate a confusion matrix, which is used to derive some convenient metrics to summarize quantitatively and qualitatively the model's performance. Assuming that TP is the number of true positive instances, TN the number of true negative instances, FP the number of false positive instances and FN the number of false negative instances. From the obtained classifier's confusion matrix, we can derive the F_1-score (F_1) as shown in Eq. (6):

$$F_1 = \frac{2 \times TP}{2 \times TP + FP + FN}. \tag{6}$$

in which F_1-score can be described as the harmonic average of the classifier's precision and recall. High F_1 scores express high values for both precision and recall.

The performance predictions of students at the end of the scholar year for Datasets I and II (Portuguese and Math subjects, respectively) using the classifier SVM are described on Table 2. The obtained results shows the shortcomings of dealing with high dimensional datasets and affecting the classifier performance. Thus, the dimensionality reduction using PCA was able to concentrate the most variation of data in few dimensions while minimizing as much as possible the information loss.

Table 2. F_1 scores using the classifier SVM (a) and Naive Bayes (b) for both datasets considering the original high dimensional data (High) and the data with reduced dimensionality (PCA), which has been tests using two dimensions (2 PCs), five dimensions (5 PCs) and ten dimensions (10 PCs).

Dataset	High	2 PCs	5 PCs	10 PCs
Dataset I (Portuguese)	0.776	**0.893**	0.773	0.776
Dataset II (Math)	0.511	**0.790**	0.511	0.511

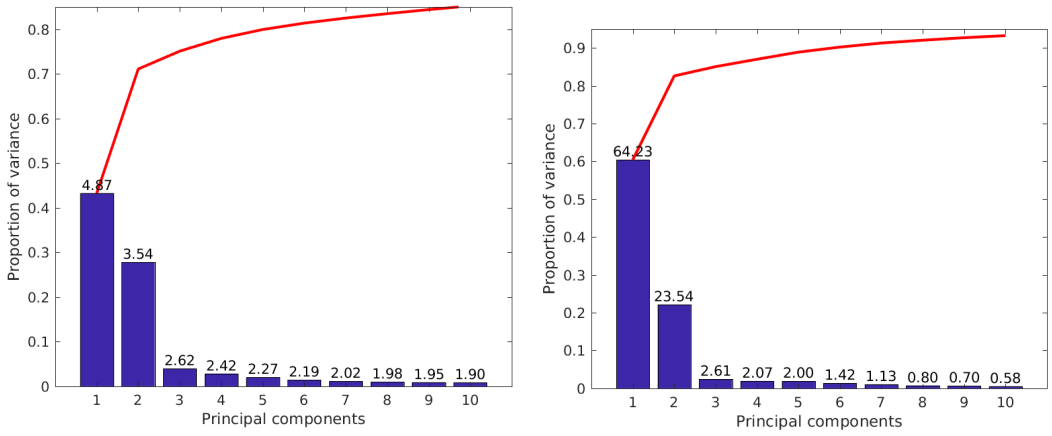
(a)

Dataset	High	2 PC's	5 PC's	10 PC's
Dataset I (Portuguese)	0.930	**0.992**	0.883	0.895
Dataset II (Math)	0.849	**0.917**	0.909	0.915

(b)

Table 2 (b) reports the students' performance predictions using the classifier Naive Bayes. We can observe that Naive Bayes outperformed SVM in all the tested configurations for both datasets. Moreover, using only two dimensions for the low-dimensional data yielded the best F_1-scores. In order to understand in more details this behavior, we provide next an individual analysis of the principal components and its explained variance.

Figure 2 plots the principal components (x-axis) which are associated to the top-10 higher eigenvalues (y-axis). Each bar is associated to a principal component and its size expresses the proportion of explained variance. The red line indicates the cumulative explained of variance through the principal components. Figure 2(a) depicts the PCA outputs when Dataset I is used as input, in which it can be seen that the top two principal components explain the most variance of data, since it concentrates an amount of more than 70% of variance. In Figure 2(b), the top two principal components explain more than 85% of the total variance of data. The plot suggests



(a) (b)

Figure 2. Top ten higher eigenvalues and their correspondent principal components: (a) Dataset I (b) Dataset II.

Table 3 presents the coefficients of each attribute in relation to the two main principal components for the Portuguese subject (Dataset I). As expected, the interpretation of principal component 1 (PC1) shows that the intermediary outcomes (`G1` and `G2`) obtained by students during the scholar year are relevant factors for students' performances, alongside the higher education of students' mothers (`Medu` = "4"). This can be explained by the large negative association to that attributes. The other variables (higher="no" and Medu = "1") present a positive association to PC1, but contrast in relation to the other three variables, which can indicate that higher students' performances are not related to the low mother's education and the student is not willing to pursue higher education. The analysis of Principal Component 2 (PC2) allows us to infer that this component expresses information regarding the alcohol consumption of students, in which low consumption, given by the positive coefficients contrasts with high consumptions of male students.

Table 3. Top-5 coefficients for the two principal components concentrating the most variance of data: (a) Dataset I; (b) Dataset II. The values between brackets express the coefficient of the associated attribute to the principal component.

PC1	PC2
G1 (-0.3)	Dalc="1" (0.357)
G2 (-0.294)	Walc="1" (0.279)
Medu = "4" (-0.259)	sex="M" (-0.265)
higher="no" (0.222)	Walc="5" (-0.219)
Medu = "1" (0.221)	Walc="4" (-0.216)

(a)

PC1	PC2
absences (-0.998)	G2 (-0.752)
age (-0.029)	G1 (-0.649)
G2 (0.024)	failures (0.058)
Walc (-0.023)	goout (0.04)
G1 (0.021)	absences (-0.034)

(b)

Table 3(b) presents the same analysis, but now considering Dataset II (Math subject). In PC1, the most relevant coefficient is associated to the number of absences of a student, which also presents a negative association to the age and weekly alcohol consumption. This is a contrast to the students' intermediary grades during the scholar year (`G1` and `G2`). PC2 concentrates more information related to the grades `G1` and `G2` and alongside to the number of absences, they contrast to the number of past failures.

The proposed method was able to improve the classification results when predicting the students' performances. It can be noted that, as more dimensions were taken into account to compose the low-dimensional transformed data, the obtained F_1-scores approximated to the F_1-scores achieved by the classifiers when using the high dimensional data. Therefore, using a two-dimensional data generated by PCA yielded to the best classification results since the associated principal components explained the most variance of data that minimized the information loss.

From a Education perspective, the proposed method was able to perform the predictions and to reveal interesting patterns that are implicit on students' characteristics. The student records describing their personal, familiar and social attributes can also be taken into account to better understand their profiles and make decisions in order to improve their learning. Particularly, in the considered datasets, the student's mother education, the alcohol consumption and the willing of enrolling in college were relevant factors that concentrate the most variance of data.

5. Conclusions

This paper described a method for students' performance prediction using principal components analysis (PCA) that can be applied simultaneously for dimensionality reduction and to extract implicit knowledge in datasets. The classifiers Support Vector Machines and Naive Bayes were used for the performances predictions. Two public educational datasets describing students' approval or failure at the end of the scholar year concerning the subjects Math and Portuguese were considered for validating the proposed method.

We performed experiments to compare the prediction performances between the high (original) and reduced dimensional datasets. Results indicated that the dataset with reduced dimensionality by PCA retained the most relevant information and relations among attributes, yielding to outperform classification performances. The analysis of the students' data resulted in the identification of some interesting personal and academical variables that might influence their performances, such as failures, period grades, mother's education and alcohol consumption. Findings like these are valuable information that educational institutes can take advantage with their own data.

Future work will concern on studying non-linear dimensionality reduction techniques for predictive tasks and Multiple Correspondence Analysis, which can also be appropriate for handling categorical attributes in educational datasets.

References

Abou-Ali, H. and Abdelfattah, Y. M. (2013). Integrated paradigm for sustainable development: A panel data study. *Economic Modelling*, 30:334–342.

Asif, R., Merceron, A., Ali, S. A., and Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113:177–194.

Baker, R. S. (2014). Educational data mining: An advance for intelligent systems in education. *IEEE Intelligent Systems*, 29(3):78–82.

Baradwaj, B. K. and Pal, S. (2012). Mining educational data to analyze students' performance. *arXiv preprint arXiv:1201.3417*.

Cortez, P. and Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. *Universidade do Minho, Portugal*.

Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., and Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 73:247–256.

Duforet-Frebourg, N., Luu, K., Laval, G., Bazin, E., and Blum, M. G. (2015). Detecting genomic signatures of natural selection with principal component analysis: application to the 1000 genomes data. *Molecular biology and evolution*, 33(4):1082–1093.

Dutt, A., Aghabozrgi, S., Ismail, M. A. B., and Mahroeian, H. (2015). Clustering algorithms applied in educational data mining. *International Journal of Information and Electronics Engineering*, 5(2):112.

Filmer, D. and Pritchett, L. H. (2001). Estimating wealth effects without expenditure data-or tears: an application to educational enrollments in states of india. *Demography*, 38(1):115–132.

Jolliffe, I. T. (1986). Principal component analysis and factor analysis. In *Principal component analysis*, pages 115–128. Springer.

Linden, R., Barbosa, L. F., and Digiampietri, L. A. (2017). "brazilian style science" - an analysis of the difference between brazilian and international computer science departments and graduate programs using social networks analysis and bibliometrics. *Social Network Analysis and Mining*, 7(1):44.

Polat, K. and Güneş, S. (2007). Detection of ecg arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine. *Applied Mathematics and Computation*, 186(1):898–906.

Romero, C. and Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6):601–618.

Shahiri, A. M. and Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72:414–422.

Tan, P.-N., Steinbach, M., and Kumar, V. (2005). *Introduction to Data Mining*. Addison Wesley, us ed edition.

Wall, M. E., Rechtsteiner, A., and Rocha, L. M. (2003). Singular value decomposition and principal component analysis. In *A practical approach to microarray data analysis*, pages 91–109. Springer.

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.