Evaluating Student Performance in E-learning Systems: A Two-step Robust Bayesian Multiclass Procedure

Pacifico, Antonio and Giraldi, Luca and Cedrola, Elena

University of Macerata, LUISS Guido Carli University, University of Bruxelles

February 2023

Evaluating Student Performance in E-learning Systems: A Two-step Robust Bayesian Multiclass Procedure

Antonio Pacifico*
Luca Giraldi†
Elena Cedrola‡

Abstract

This paper addresses a computational method to evaluate student performance through convolutional neural network. Image recognition and processing are fundamentals and current trends in deep learning systems, mainly with the outbreak in coronavirus infection. A two-step system is developed combining a first-step robust Bayesian model averaging for selecting potential candidate predictors in multiple model classes with a frequentist second-step procedure for estimating the parameters of a multinomial logistic regression. Methodologically, parametric conjugate informative priors are used to deal with model uncertainty and overfitting, and Markov Chains algorithms are designed to construct exact posterior distributions. An empirical example to the use of e-learning systems on student performance analysis describes the model's functioning and estimation performance. Potential prevention policies and strategies to address key technology factors affecting e-learning tools are also discussed.

Keywords: E-learning systems; Student Performance; Bayesian Inference; Policy Issues; Logistic Regression; Variable Selection Procedure.

*Corresponding author: Antonio Pacifico, Department of Economics and Law, University of Macerata (Marche), Italy. Email: antonio.pacifico@unimc.it ORCID: https://orcid.org/0000-0003-0163-4956

†Corresponding author: Luca Girarldi, Department of Economics and Law, University of Macerata (Marche), Italy. Email: l.giraldi@unimc.it ORCID: https://orcid.org/0000-0002-3446-8993

‡Corresponding author: Elena Cedrola, Head of the Department of Economics and Law (Department of Economics and Law), University of Macerata (Marche), Italy. Email: elena.cedrola@unimc.it ORCID: https://orcid.org/0000-0002-3382-5055

1 Introduction

The Covid-19 pandemic, ongoing for more than three years, has significantly impacted education worldwide. In the last two years, to cope with the impact of Covid, the use of online technology for teaching has become increasingly common, allowing students to take courses and exams from locations outside the physical classroom (see, for instance, O'Reilly and Creagh (2015)). In distance learning, people can learn the things they need, at any time and any place, so much so that distance learning has proven to be more effective than traditional education in many cases.

The widespread adoption of online learning courses by public and private institutions is a further incentive for developing learning platform programs with reduced course and training costs. However, the online learning environment is different, with higher distraction factors than the classroom, as the student is surrounded by his/her home environment. Therefore, the influence of online education on learning efficiency is a very important issue worth exploring, as student engagement has always been one of the key topics in education. Some studies have indicated that students engagement can be improved through appropriate instructional interventions, good study design and instant feedback (see, for instance, Naim et al. (2021)).

To measure student engagement, face detection is the ideal starting point. When a human perceives a stimulus, facial expressions reflect emotional and cognitive states (see, e.g., Picard (1997)). Facial expression is one of the most powerful signals for humans to convey emotions or intentions, so Facial Expression Recognition (FER) has wide applications in human-computer interaction and affective computing and has attracted much research interest in recent years (see, e.g., Chen et al. (2019)). This increase in the use of facial recognition is due to the rapid growth of user Big Data and an increase in machine learning and deep learning performance (see, for instance, Penczynski (2019)). Researchers have gradually started to use deep learning methods, big data, and convolutional neural networks, which have achieved much accuracy in computer vision, and the models generally perform well (see, e.g., Khan et al. (2019) and Naim and Alahmari (2020)).

Numerous methods have been developed for automatic RES that can be summarised into two main classes: traditional and deep learning methods. More recent research is shifting its focus from the traditional method to the application of Deep Convolution Neural Networks (CNN), which outperform the former and enable analysis without explicit knowledge of the underlying process model (see, e.g., Heinrich et al. (2021)). Unlike traditional approaches, deep learning methods, such as the Convolution Neural Network (CNN), use an end-to-end mode to train an extremely deep network structure with millions of parameters that adaptively learn the useful features from huge data without the need for hand-crafted features.

Different machine learning techniques have already been used for facial expression recognition in computer vision, and the facial expressions are directly linked to the perceived engagement of online learners (see, e.g., Pons and Masip (2017)). Yang et al. (2017) fused two CNNs trained separately using the grey images and Local Binary Patterns (LBP) images. Hasani and Mahoor (2017) combined CNN with Conditional Random

Fields (CRF) to capture the spatial relation within facial images and the temporal relation between the image frames.

In economics, unsupervised and supervised Machine Learning (ML) techniques are used for linear dimensionality reduction of large parameter spaces. The former accounts for unlabeled data extracting their generative features and exploratory purposes (see, e.g., Blei et al. (2003) (latent Dirichlet allocation model); Gopalan et al. (2015) (hierarchical Poisson factorization); Athey and Imbens (2015, 2016) (diffuse prior distributions in large model classes); and Shashua and Levin (2001), Tucker (1966), and Tipping and Bishop (1999) (Principal Component Analysis and related approaches)). Because of shrinking and then grouping unlabeled data, these algorithms tend to be computationally complex providing unaccurated and biased estimates. Supervised ML methods classify and group factors through labeled datasets predicting outcomes accurately (see, e.g., Varian (2014) and Mullainathan and Spiess (2017) (Least Absolute Shrinkage and Selection Operator), Hoerl and Kennard (1970) (Ridge regression), Breiman (2001) (Random Forest), and Awad and Khanna (2015) (Support Vector Machines for classification problems)). However, the data compression involved in these algorithms does not have any reference to the outcomes, and then they are unable to deal with some open related questions in variable selection problems such as model uncertainty when a single model is selected `a priori` to be the true one (see, e.g., Miller (1984), Madigan and Raftery (1994), Breiman (1992, 1995), Breiman and Spector (1992), Raftery et al. (1995)), overfitting when multiple models are selected provinding a somewhat better fit to the data than simpler ones (see, e.g., Madigan et al. (1995), Raftery et al. (1997), Mullainathan and Spiess (2017), and Pacifico (2020)), and endogeneity issues because of omitted factors and unobserved heterogeneity (see, for instance, Gelfand and Dey (1994) and Pacifico (2020)).

The methodology developed in this article aims to overtake the aforementioned issues by developing a Two-step Robust Bayesian Multiclass (TRBM) procedure that combines a first-step Bayesian strategy for selecting the only (potential) predictors affecting the outcomes with a frequentist second-step procedure for estimating the parameters of a multinomial logistic regression. It is robust because of using a set of informative priors for every possible outcomes rather than a single one.

The first step builds on and improve the Pacifico (2020)'s analysis, who develops a Robust Open Bayesian (ROB) procedure – through two stages – implementing Bayesian Model Selection (BMS) and Bayesian Model Averaging (BMA) to shrink large model and parameter space when studying the dynamics of the economy in either time-invariant moderate data or time-varying high dimensional multivariate data. The implementation cosists of adapting the strategy to a multiclass decomposition problem, predicting the probability of different possible outcomes of a categorically distributed dependent variable. In ML classification algorithms, multiclass or multinomial decomposition refers to the problem of classifying instances into more than two classes. Indeed, the `best` final model solution (or combination of predictors) obtained in the shrinking process needs to predict the probability of every potential outcome of the variable of interest, where `best` stands for the model solution providing the most accurate predictive performance over all candidate models. More precisely, the first stage entails finding a pool of predictors on a set of cross-sectional data with highly strong explanatory powers on the potential outcomes. A generalization of the Conjugate Informative Proper priors developed in Pacifico (2020) – named Multiclass CIP (MCIP) priors – are used to design parameters' distributions and estimating them via Monte Carlo Markov Chain algorithms. TIn this way, the variable selection procedure simultaneously moves through multiclass model and parameter space up to obtain a reduced set containing the `best` submodel solutions that mainly explain and fit the data. Then, a further shrinkage is conducted to obtain the smallest final subset of `best` submodels containing the only `significant` solutions, where `significant` stands for multiclass models having statistically significant predictive capability on the potential outcomes. The submodel with higher log Bayes Factor (lBF) will be the final solution on which making inference.

Given the final `best` sample, the second step involves estimating a multinomial logistic regression in order to evaluate how the mean values of the potential predictors affect the outcomes (predicted probabilities). In ML and facial recognition e-learning platform, natural images are formed by the interaction of multiple factors related to different characteristics of individuals and scenario (see, e.g., Chellappa et al. (1995), Moses et al. (1996), and Turk and Pentland (1991)). Thus, a generalized logistic regression version is entailed facing to multiclass problems and evaluating the average effect of changes in predictors on the change in the probability of different outcomes (known as marginal effects).

The contributions addressed in this study are fourfold. First, variable selection problems are dealt with selecting the `best` combination of predictors through a hierarchical robust Bayesian method. Second, MCIP priors and MCMC-based Posterior Model Probabilities (PMPs) are used to include cross-sectional individuals' information from the whole system, acting as a strong model selection in multiple model classes. In this context, PMPs denote the probability of each candidate model performing the data. Third, structural model uncertainty, because of functional forms of misspecification, is also avoided let the framework be hierarchical and robust. Fourth, better policy strategy can be perfomed by three main features: `(i)` the hierarchical robust Bayesian approach; `(ii)` the use of informative mixture priors; and `(iii)` the inclusion of an 'ad-hoc' model and variable selection.

The empirical example focuses on monitoring and evaluating students' level of attention and attendance throughout training courses. A Massive Open Online Courses (MOOCs)-based e-learning platform – acting for variable of interest – is used to assess student performance attending classes through facial recognition and image processing. The analysis refers to a Survey conducted on 56 students and aims to highlight the benefit of e-learning systems based on digital web technologies in proving access to a high-quality institution.

The remainder of this paper is organized as follows. Section 2 introduces the econometric model and the estimating procedure. Section 3 displays prior specification strategy and posterior distributions dealing with Bayesian inference. Section 4 describes the empirical analysis. The final section contains some concluding remarks.

2 Two-step Robust Bayesian Multiclass Procedure

2.1 Bayesian Inference: First Step

According to Pacifico (2020), the baseline model is:

$$Y_i = \theta_0 + \sum_{j=1}^{k} \theta_j x_{ji} + \epsilon_i \quad \text{or} \quad f(Y_i|\theta_{ji}, \sigma) = N(X\theta, \sigma^2 I_n) \quad , \tag{1}$$

where Y_i is a $n \cdot 1$ vector denoting the outcomes of a categorically distributed dependent variable with more than two levels, with $i = 1, 2, \ldots, n$, θ_0 denotes a constant term, $X = (x_{1i}, x_{2i}, \ldots, x_{ki})'$ is a $n \cdot k$ matrix including observable continuous and/or discrete covariates, with $j = 1, 2, \ldots, k$ denoting the predictors (or covariates), $\theta_j = (\theta_1, \theta_2, \ldots, \theta_k)$ is a $k \cdot 1$ vector of unknown regression coefficients for k potential covariates, and $\epsilon_i \sim N(0, \sigma^2)$ is a $n \cdot 1$ vector of disturbances, with σ to be an unknown positive scalar. Here, the error component is assumed independent and identically distributed (*i.i.d.*) and homoskedastic.

The main thrust of Bayesian dimensionality reduction approaches is to find the `best` submodel solution among large model classes better explaining and then fitting the data. The involved variable selection procedure aims to estimate 2^m distinct regression models, with m denoting all possible model solutions (or combination of predictors), and average over them (BMA) by excluding all the covariates not improving prediction and thus over-confident inferences and decisions about quantities of interest (BMS). Three drawbacks of these approaches are in order: `(i)` the use of either non-informative (or diffuse) and 'common' informative priors estimating the unknown regression coefficients θ_j and variance σ^2; `(ii)` the addition of penalty terms or restrictions on data-supported models when there is no relationship between potential predictors; and `(iii)` the shrinkage of model and parameter let the outcomes be normally distributed. The first two problems are overtook in the Pacifico (2020)'s ROB procedure and the last is addressed by developing MCIP priors for every multiclass model with categorically distributed dependent variable(s). In this way, one will be sure to account for the only relevant factors improving the relationship between student performance and e-learning platform discarding non-relevant variables within the system.

The variable selection problems are addressed by using two (generalized) auxiliary indicator variables. The first corresponds to a vector χ_{ji}, containing every possible 2^m subset choices, with $\chi_{ji} = 0$ if θ_j is small (absence of k-th covariate in the model given i) and $\chi_{ji} = 1$ if θ_j is sufficiently large (presence of k-th covariate in the model given i). The second is a vector β_{ji}, corresponding to the regression parameter θ_j when it is sufficiently large (presence of predictors x_{ji} in the procedure for every i); conversely, the predictor x_{ji} will be dropped by the system.

In this study, the ROB procedure involves in jointly shrinking the model and the parameter space by matching all candidate multiclass models in order to deal with individual contributions (overfitting) and model uncertainty. Model misspecification problems are also avoided by pooling the cross-sectional individual information from the whole system. The shrinking process consists of evaluating the probability of every possible outcome to perform the data given a set of explanatory variables, named Posterior Model Probability. Let $\mathcal{M}$ and $\mathcal{J}$ be the natural model and parameter space, respectively, containing all candidate multiclass model solutions for n possible outcomes (or levels), the full model set is:

$$\mathcal{F} = \Big\{ M_{ji} : M_{ji} \subset \mathcal{F}, M_{ji} \in \mathcal{M}, j \in \mathcal{K},\ \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \ldots + \theta_k x_{ki} + \epsilon_i \Big\} \quad , \tag{2}$$

where M_{ji} denotes a countable collection of candidate multiclass models containing the vector of the unknown parameters θ for n possible outcomes. Thus, the PMP can be defined as:

$$f(M_{ji}|Y_i = n) = \frac{f(M_{ji}) \cdot f(Y_i = n|M_{ji})}{\sum_{M_{ji} \in \mathcal{M}} f(M_{ji}) \cdot f(Y_i = n|M_{ji})} \quad , \tag{3}$$

where $f(Y_i = n|M_{ji}) = \int f(Y_i = n|M_{ji}, \theta_j) \cdot f(\theta_j|M_{ji}) d\theta_j$ is the marginal likelihood, with $f(\theta_j|M_{ji})$ denoting the conditional prior distribution of θ_j. In this context, predicting the PMP associated to any level (j) for m model solutions, one outcome is chosen as a `pivot` and then the other $k - 1$ outcomes are separately regressed against the `pivot` outcome.

Nevertheless, when the data (corresponding to all possible 2^m model solutions) are highly larger than the time frame, the calculation of the integral in $f(Y_i = n|M_{ji})$ is unfeasible and then MCMC algorithms and implementations need to be accounted for. More precisely, the idea is to generate recursively the observations from the joint posterior distribution $f(M_{ji}, \theta_j|Y_i = n)$ of (M_{ji}, θ_j) for estimating $f(M_j|Y_i = n)$ and $f(\theta_j|M_{ji}, Y_i = n)$.

In the first stage of ROB procedure, a pool of `best` submodels is obtained:

$$\mathcal{S} = \Big\{ M_{\tilde{j}i} : M_{\tilde{j}i} \subset \mathcal{S}, \mathcal{S} \in \mathcal{F}, \tilde{j} \in \mathcal{J},\ \sum_{\tilde{j}=1}^{\tilde{k}} \pi\Big(M_{ji}|Y_i = n, \theta_{\tilde{j}}\Big) \geq \tau \Big\} \quad , \tag{4}$$

where $M_{\tilde{j}i}$ contains $x_{\tilde{j}i}$ covariates, with $\tilde{j} = 1, 2, \ldots, \tilde{k}$, $M_{\tilde{j}i} \ll M_{ji}$, $\tilde{j} \ll j$, with $\{1 \leq \tilde{j} < j\}$, and τ is a threshold chosen arbitrarily for an enough posterior consistency ensuring that the PMP concentrates on the true model class. In this study, $\tau = 0.5\%$ in order to jointly manage all cross-sectional equations within the system for every potential outcome m.

The second stage entails further reducing the model space $\mathcal{S}$ obtaining a smaller subset of `best` submodel classes:

$$\mathcal{E} = \Big\{ M_{\xi i} : M_{\xi i} \subset \mathcal{E}, \mathcal{E} \in \mathcal{S},\ \sum_{\xi=1}^{\varkappa} \pi\Big(M_{\xi i}|Y_i = n, \theta_{\xi}\Big) \geq \tau \Big\} \quad , \tag{5}$$

where $M_{\xi i}$ contains the only `significant` solutions contained in the reduced class set $\mathcal{E}$, with $M_{\xi i} \ll M_{\tilde{j}i}$. The final regression model has the form:

$$Y_i = \sum_{\xi=1}^{\varkappa} \theta_\xi x_{\xi i} + \eta_i \quad , \tag{6}$$

where $x_{1i}, x_{2i}, \ldots, x_{\varkappa i}$ is a subset of $x_{1i}, x_{2i}, \ldots, x_{ki}$, with ξ denoting a subparameter index sufficiently smaller than $\tilde{j}$ by construction, θ_ξ denotes the unknown parameters belonging to $M_{\xi i}$, which contains the only `significant` submodel solutions, and η_i is the *i.i.d.* error term.

Finally, the exact and final solution will correspond to one of the submodels $M_{\xi i}$ with higher log natural Bayes Factor (lBF):

$$lBF_{\xi,\tilde{j}} = log\left\{\frac{\pi(M_{\xi i}|Y_i = n)}{\pi(M_{\tilde{j}i}|Y_i = n)}\right\} \quad . \tag{7}$$

In this study, the scale of evidence used for interpreting the lBF in (7) is a generalised version of Kass and Raftery (1995):

$$\begin{cases} 0.00 \leq lB_{\xi,\tilde{j}} \leq 5.00 & \text{no evidence for submodel } M_{\xi i} \\ 5.00 < lB_{\xi,\tilde{j}} \leq 10.00 & \text{moderate evidence for submodel } M_{\xi i} \\ 10.00 < lB_{\xi,\tilde{j}} \leq 15.00 & \text{strong evidence for submodel } M_{\xi i} \\ lB_{\xi,\tilde{j}} > 15.00 & \text{very strong evidence for submodel } M_{\xi i} \end{cases} \tag{8}$$

2.2 Multinomial Logistic Regression: Second Step

Let the final regression in (6) and the last outcome ($\bar{k}$) be chosen as pivot, the multinomial logit model is described as a set of independent binary regressions:

$$ln\Big(\frac{\pi(Y_i = 1)}{\pi(Y_i = n)}\Big) = \theta_1 x_{1i} \ ; \ ln\Big(\frac{\pi(Y_i = 2)}{\pi(Y_i = n)}\Big) = \theta_2 x_{2i} \ ; \ \ldots \ ; \ ln\Big(\frac{\pi(Y_i = n-1)}{\pi(Y_i = n)}\Big) = \theta_\varkappa x_{\varkappa i} \quad . \tag{9}$$

According to the Two-step Robust Bayesian Multiclass (TRBM) procedure, the lBF in (7) can be re-written as Additive Log Ratios (ALRs) in function of candidate multiclass models $M_{\xi i}$:

$$ALR(M_{\xi i}) = \left(ln\left\{\frac{\pi(M_{11}|Y_i = 1)}{\pi(M_{\tilde{j}n}|Y_i = n)}\right\} \cdot ln\left\{\frac{\pi(M_{22}|Y_i = 2)}{\pi(M_{\tilde{j}n}|Y_i = n)}\right\} \cdot \ldots \cdot ln\left\{\frac{\pi(M_{\varkappa n-1}|Y_i = n-1)}{\pi(M_{\tilde{j}n}|Y_i = n)}\right\}\right) \quad . \tag{10}$$

Given the predictors $x_{1i}, x_{2i}, \ldots, x_{\varkappa i}$, the probabilities for every i potential outcomes of Y_i given $M_{\xi i}$ final submodel solutions are:

$$\pi_k(M_{\xi i}) \quad := \quad \pi(Y_i = n|x_{1i}, x_{2i}, \ldots, x_{\varkappa i}) = \frac{exp\{\theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \ldots + \theta_\varkappa x_{\varkappa i}\}}{1 + exp\{\theta_0 + \sum_{\xi=1}^{\varkappa} \theta_1 x_{1i} + \theta_2 x_{2i} + \ldots + \theta_\varkappa x_{\varkappa i}\}} =$$

$$= \frac{1}{1 + exp\{\theta_0 + \sum_{\xi=1}^{\varkappa-1} \theta_1 x_{1i} + \theta_2 x_{2i} + \ldots + \theta_\varkappa x_{\varkappa i}\}} \quad . \tag{11}$$

Equation (11) implies that $\sum_{\xi=1}^{\varkappa} \pi_j(M_{\xi i}) = 1$ and there are $(\varkappa - 1) \cdot (k + 1)$ coefficients observed and estimated through the auxiliary parameters χ_{ji} and β_{ji}, respectively. Taking the ratio between the first and the second term to the right-hand-side (RHS) of (11), the results gives:

$$\frac{\pi_i(M_{\xi i})}{\pi_n(M_{\xi i})} = exp\{\theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \ldots + \theta_\varkappa x_{\varkappa i}\} \quad , \tag{12}$$

or

$$\pi_i(M_{\xi i}) = exp\{\theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \ldots + \theta_\varkappa x_{\varkappa i}\} \cdot \pi_n(M_{\xi i}) \quad . \tag{13}$$

Thus, applying the logarithm to both sides, the Odds Ratio (OR) are written as:

$$ln\left(\frac{\pi_i(M_{\xi i})}{\pi_n(M_{\xi i})}\right) = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \ldots + \theta_\varkappa x_{\varkappa i} \quad , \tag{14}$$

where the left-hand-side (LHS) denotes the logarithm of the ratio of two probabilities and the RHS refers to a linear combination of the predictors. A log-odds or a logistic regression for $\pi(Y_i = n)$ entails the probabilities on the LHS to be complementary. In a context of variable selection procedure with multinomial logistic regression, equation (14) represents a set of $(n - 1)$ independent logistic regressions over m potential model solutions for the probability of $Y_i = i$ versus the probability of the reference $Y_i = n$.

Here, three main findings are in order. `(i)` The probabilities of $Y_i = i$ and $Y_i = n$ when $x_{1i} = x_{2i} = \ldots = x_{\varkappa i} = 0$ set to $exp\{\theta_0\}$, corresponding to the ratio between $\pi_i(\mathbf{0})/\pi_n(\mathbf{0})$. Let $exp\{\theta_0\} > 1$ (or $\theta_0 > 0$), $\pi(Y_i = i)$ will be more likely than $\pi(Y_i = n)$; conversely when $exp\{\theta_0\} < 1$. `(ii)` All variables within the system will be – by construction – potentially *significant* with highly strong predictive accuracy. `(iii)` The variable selection problem in (1) can be modelled hierarchically, with observable outcomes (Y_i) created conditionally on the regression parameters (θ_j), which themselves are assigned a distribution in terms of further common parameters named hyperparameters. Because of common parameters tend to change meaning from one model solution to another, every prior distribution has to change in a corresponding fashion. This hierarchical thinking would solve the trade-off between inaccurate fit and overfitting, playing an important role in developing computational strategies.

3 Prior Specification Strategy and MCMC Distributions

The variable selection procedure entails estimating the regression parameters (θ_{ji}) and the auxiliary indicators (χ_{ji} and β_{ji}) as posterior means (the probability that a variable is *in* the model). All observal variables in (1) are hierarchically modelled via Multiclass Conjugate Informative Proper (MCIP) priors:

$$\pi(\beta, \sigma^2, \chi|Y) = \pi(\beta|\sigma^2, \chi) \cdot \pi(\sigma^2|\chi, \beta) \cdot \pi(\chi) \quad . \tag{15}$$

The main thrust of the MCIP priors is analytically margining out of β and σ^2 from $\pi(\beta, \sigma^2, \chi|Y)$. Thus, the equation (15) can be rewritten as:

$$\pi\Big(\beta|\chi\Big) = N\Big(\mu_\chi, \Sigma_\chi\Big) \quad , \tag{16}$$

$$\pi(\chi) = w_{|\chi|} \cdot \begin{pmatrix} j \\ |\chi| \end{pmatrix}^{-1} \quad , \tag{17}$$

$$\pi(\sigma^2|\chi) = IG\Big(\frac{\omega}{2}, \frac{\nu}{2}\Big) \quad , \tag{18}$$

with $N(\cdot)$ and $IG(\cdot)$ standing for Normal and independent Inverse-Gamma distributions, respectively, μ_χ is a hyperparameter for the auxiliary regression coefficients β_{ji}, $\Sigma_\chi = \Gamma_\chi \otimes V_k$ denotes the $[(k+1)\cdot(k+1)]$ covariance matrix, $w_{|\chi|}$ refers to the model prior choice related to the sum of the PMPs (or Prior Inclusion Probabilities) with respect to the model size $|\chi|$, through which the β_{ji}'s will require a non-0 estimate or the θ_{ji}'s should be included in the model, and ω and ν are hyperparameters to be chosen in order to decrease with the size of the final selected subset $M_{\xi,i}$. In this way, one would weight more according to model size and – setting $w_{|\chi|}$ large for smaller $|\chi|$ – assign more weight to parsimonious models. Moreover, the use of independent $IG(\cdot)$ distribution allows cross-equation independence of the coefficient distributions.

The covariance matrix in (16) is obtained from two components: $V_k = (\sigma^2 \cdot I_k)$, where σ^2 denotes the residual variance for the χ-th model, and the diagonal matrix $\Gamma_\chi = diag(I_k, a^2_{\chi 1i}\Sigma'_{x'_{1i}x_{1i}}, a^2_{\chi 2i}\Sigma'_{x'_{2i}x_{2i}}, \dots, a^2_{\chi ki}\Sigma'_{x'_{ki}x_{ki}})$, where $a^2_{\chi_{ji}}$ is a hyperparameter equal to 0 whether $\chi_{ji} = 0$ and equal to 1 otherwise, and $\Sigma_{x_{ji}x_{ji}} = \frac{1}{2}(x'_{ji}x_{ji})^{-1}$. Here, the k-th diagonal element of Γ_χ is appropriately set to be small or large according to whether $\chi_{ji} = 0$ or $\chi_{ji} = 1$, respectively.

All hyperparameters are known. More precisely, collecting them in a vector ψ, where $\psi = \Big(\mu_\chi, w_{|\chi|}, \omega, \nu\Big)$, they are treated as fixed and are either obtained from the data to tune the prior to the specific applications (such as $\mu_\chi, w_{|\theta|}, \omega$) or selected a priori to produce relatively loose priors (such as ν). Here, μ_χ and $w_{|\chi|}$ are restricted to a benchmark prior[1] $max\Big(kn, |\chi|\Big)$ according to the non-0 components of regression parameter

[1] It would be a generalization of the standard objective prior of Zellner (1986).

θ_{ji}.

Let $|\chi|$ be the model size, an ergodic Markov Chain (MC) to obtain posterior distributions is:

$$\beta^{(0)}, \sigma^{(0)}, \chi^{(0)}, \beta^{(1)}, \sigma^{(1)}, \chi^{(1)}, \beta^{(2)}, \sigma^{(2)}, \chi^{(2)}, \ldots, \quad \xrightarrow{d} \quad \pi(\beta, \sigma^2, \chi|Y) \quad , \tag{19}$$

where $\beta^{(0)}$, $\sigma^{(0)}$, and $\chi^{(0)}$ are automatically assigned to the model selection procedure in absence of any relationship between potential predictors, with $\sigma^{(0)}$ denoting the full variance of Y. The MC sequence in (19) converges in distribution to the full posterior $\pi(\beta, \sigma^2, \chi|Y)$ and corresponds to an auxiliary Gibbs sequence. According to the variable selection procedure, in large problems (e.g., when k is more than 15), this latter would provide useful and faster information performing more with respect to model size.

Combining the likelihood from (1) with the priors (16)-(18), it yields to the joint posterior:

$$\pi(\beta, \sigma, \chi|Y) \propto \sigma^{-(n+m_\chi+\omega+1)} \cdot |\Sigma_\chi|^{-1/2} \cdot exp\Big\{ -\frac{1}{2\sigma^2} \cdot |\bar{Y} - \bar{\mathcal{X}}_\chi \beta|^2 \Big\} \cdot exp\Big\{ -\frac{\nu}{2\sigma^2} \Big\} \cdot \pi(\chi) \quad , \tag{20}$$

where $\bar{Y} = [Y \quad 0]'$ and $\bar{\mathcal{X}}_\chi = \Big[\mathcal{X}_\chi \quad (\Sigma_\chi)^{-1/2}\Big]'$ are $(2 \cdot 1)$ vectors, with $\mathcal{X}_\chi$ being a $(n \cdot m_\chi)$ matrix whose columns correspond to the components of β_{ji} and m_χ denoting the size of the χ-th subset.

The generation of the components in (19) in conjuction with $\pi(\chi)$ in (17) is obtained through Bernoulli draws. The required sequence of Bernoulli probabilities can be computed swiftly and efficiently by exploiting the appropriate updating scheme for $\pi(\chi)$:

$$\frac{\pi\Big(\chi_{ji} = 1, \chi_{(ki)}|Y\Big)}{\pi\Big(\chi_{ji} = 0, \chi_{(ki)}|Y\Big)} = \frac{\pi\Big(\chi_{ji} = 1, \chi_{(ki)}\Big)}{\pi\Big(\chi_{ji} = 0, \chi_{(ki)}\Big)} \quad . \tag{21}$$

At each step of the iterative simulation from (19), one of the values of $\pi(\chi)$ in (21) will be available from the previous component simulation. The exact relative probability of two values χ_0 and χ_1 is obtained as $\Big[\pi(\chi_0)/\pi(\chi_1)\Big]$ allowing more accurate identification of submodel solutions with higher PIPs. These relative probabilities can be easily computed since $\pi(\chi)$ has to be calculated for every visited χ_{ji} value in the execution of the MCMC algorithm.

With these specifications, the posterior distributions can be defined as:

$$\pi\Big(\beta|\chi\Big) = N\Big(\bar{\mu}_\chi, \tilde{\Sigma}_\chi\Big) \quad , \tag{22}$$

$$\pi(\chi) = \tilde{w}_{|\chi|} \cdot \binom{\varkappa}{|\chi|}^{-1} \quad , \tag{23}$$

$$\pi(\hat{\sigma}^2|\chi) = IG\Big(\frac{\bar{\omega}}{2}, \frac{\bar{\nu}}{2}\Big) \quad , \tag{24}$$

Here, some considerations are in order. In equation (22), $\bar{\mu}_\chi$ and $\tilde{\Sigma}_\chi$ denote the arbitrary scale parameter and the posterior predictive covariance matrix for σ^2 according to any level j in homoskedastic case, respectively. In this analysis, $\bar{\mu}_\chi \cong 1.0$ and $\tilde{\Sigma}_\chi$ is obtained estimating the components Γ_χ and V_k according to the sample size $|\chi|$ as described in (8).

In equation (23), $\tilde{w}_{|\chi|}$ refers to the model posterior choice according to the sum of the PMPs obtained in the second stage with respect to model size $|\chi|$, with $\tilde{w}_{|\chi|} = max(kn, |\chi|)$ accounting for the non-0 components in $\mathcal{E}$.

In equation (24), $\bar{\omega} = \omega_0 \cdot \varkappa$ and $\bar{\nu} = \nu_0 \cdot \varkappa$, with ω_0 and ν_0 denoting the arbitrary degrees of freedom (sufficiently small) and the arbitrary scale parameter, respectively. In this analysis, $\omega_0 \cong 0.1$ and $\nu_0 \cong 0.001$.

4 Empirical Analysis: Evidence from a 56-person Survey

4.1 MOOC-based E-learning Platform

The web-based platform for managing MOOC-based e-learning courses (hereafter, unless otherwise specified, it is simply referred as e-learning platform) measures student effectiveness and satisfaction in online learning through face coding by analysing student involvement in learning as an influencing variable (see, for instance, Hu and Hui (2012)). Through a webcam, the system can:

- carry out an identity recognition and detect the presence of strangers in the action range of the webcam;
- keep track of attention level by monitoring head and gaze orientation;
- record the level of satisfaction;
- detect actions to circumvent the analysis systems or access to other pages during the lesson;
- map modified data.

In short, the platform's operation follows the framework shown in Figure 1.

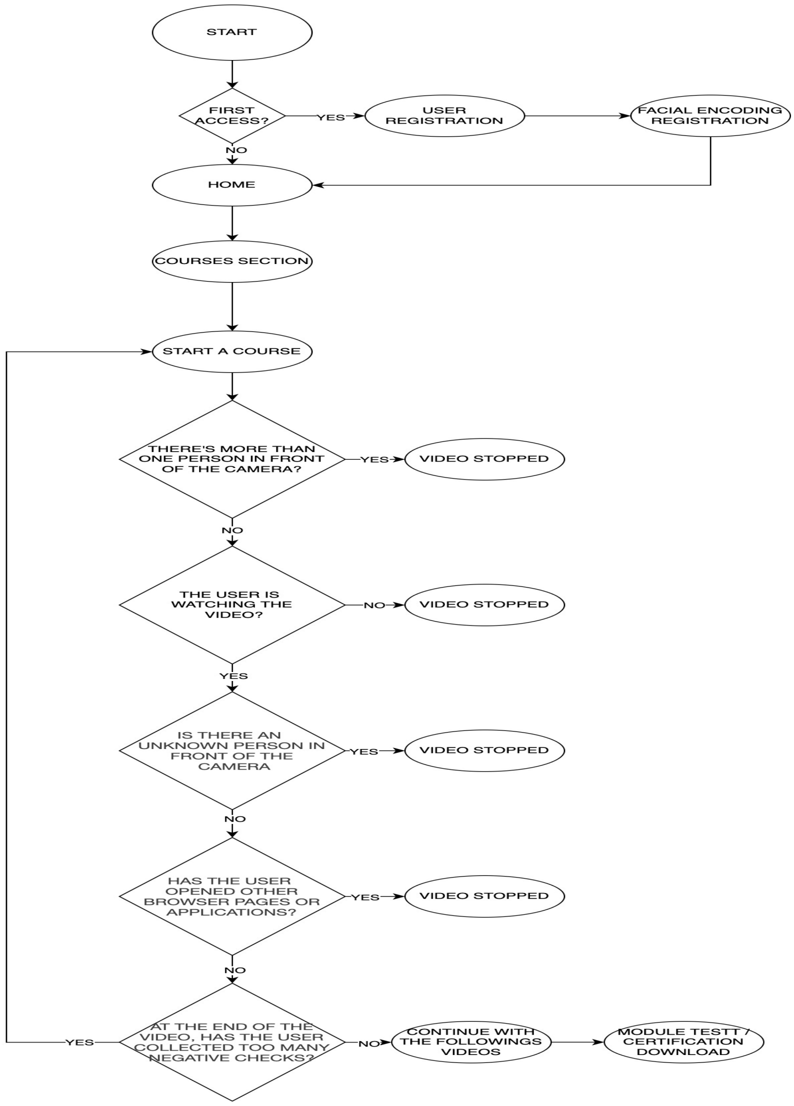

Figure 1: Platform Operation Framework

In the first login phase, the system encodes the student's facial features in a series of numerical metadata, which will be systematically compared in order to validate subsequent logins or not. It should be noted that a privacy acknowledgement is required, in which the user is informed of the encoding of biometric traits and the processing of his/her data, and in order to proceed with taking the photo, the user must first give his/her consent to privacy. After the acceptance, the photo is taken with the webcam (Figure 2); there is a minimum threshold of acceptability (generally 0.8 or 80%), below which the system automatically discards the image and asks for it to be taken again.

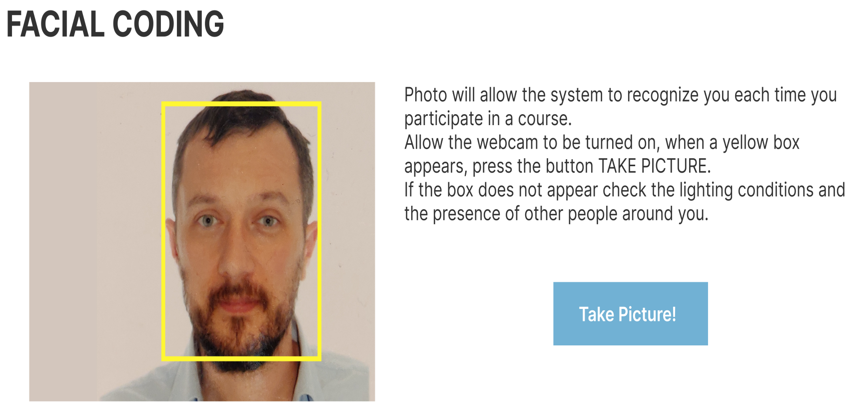

Figure 2: Facial coding – First login

««««««««««««««««««««

Once the face-coding phase is completed, the user is directed to the platform, where he/she can search for courses (Figure 3).

When opening a video course, the student must again consent for the webcam to be activated. In this way, the student is monitored throughout the course. The student's involvement may be affected by the presence of other people in the room or by distracting objects: the platform can detect the presence of other people in the room or suspicious objects (smartphones), interrupting the playback of the video. Similarly, if the student identified in the login phase exchanges with someone else, the system detects an anomaly in the biometric data (they do not match) and automatically interrupts video playback.

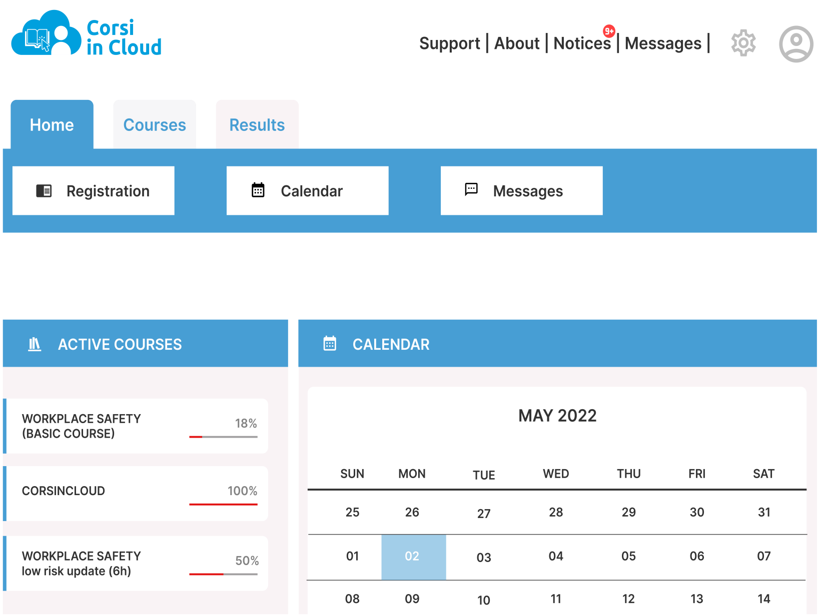

Figure 3: Platform homepage

The control has been empirically set at 5 seconds but, depending on the various needs encountered, a custom time can be set. Therefore, an image is acquired from the camera placed frontally for each shot of set periods. The captured image is used to detect: `(i)` arousal, defined as the level of involvement in learning; `(ii)` valence, evaluated as the positivity or negativity of the learning experience; `(iii)` level of attention through gaze analysis, according to the third-party Dilb library (see, e.g., King (2009)); and `(iv)` students' feel emotions through a CCN deployed in Python, integrating the Keras and Tensorflow frameworks to allow satisfaction detection (see, for instance, Talipu et al. (2019)). Concerning the attention level, if the System detects anomalies, the video stops suggesting to the student a pause or a better calibration of the webcam. As regards the emotive analysis, it is based on the Facial Coding System di Ekman (see, for instance, Rosenberg and Ekman (2020)).

Finally, the platform checks that the lesson page does not lose focus, i.e., is not overshadowed by other pages or applications. Again, if the check is considered false, the platform blocks the playback of the video. When the user clicks the "next" button after watching the video, the platform verifies that the user has completed a minimum number of attention checks before moving on to the next video or if the video has recently seen was the last video in the course, the last page of the course. The minimum control threshold is not fixed; instead, it is changed based on the length of the video. If the minimum number of checks is not met, the user has attempted to go on to the next video without first viewing the current video or tampering with the platform. The user is obliged to watch the video again in this situation. The software may also manage end-of-course or intermediate tests. In this situation, the user must pass these exams to go to the next video or the course's final page. If the user fails to pass the test, he may be compelled to watch the video or take the test again. Once the user completes the course, they may download their certificate. The user may also track their progress on their profile, where they can see their outcomes and download certificates for courses they have accomplished. Administrators and teachers, i.e. users who publish and maintain courses, can examine an in-depth part, including platform data under the platform's content management area. Teachers may access statistics for each course and submit videos using this material, which shows the amount of attention, authentication, and average satisfaction of students (Figure 4). The attention pie chart depicts the percentage of time the system determines the user is focused on the course, while the authentication pie chart depicts the percentage of times face recognition provided user identification confirmation. The satisfaction pie chart summarises the feelings experienced by users (they are clustered in 3 classes: satisfied, neutral, and dissatisfied).

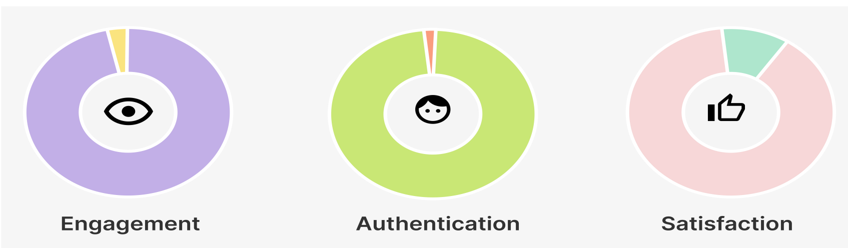

Figure 4: Pie of attention, authentication and satisfaction of the student

This data is invaluable for teachers because it allows them to understand the critical aspects of the course and its contents, tailoring the training course to each student, thus adapting it to their needs and cognitive-emotional state (see, e.g., Kratzwald et al. (2018)).

4.2 Data Description and Results

The dataset has been constructed conducting a Survey on 56 students and collecting about 30 (potential) covariates descripted through specific questions on e-learning platform such as use, utility, desirability, satisfaction, experience, and so forth. Most of the data are qualitative factors – expect for gender and age – and then trasnformed in discrete variables (either nominal or ordinal). The variable of interest is a dummy variable = 1 whether the user has benefited from either MOOC or webinar, and = 0 whether the user has benefited from webinar only.

According to a descriptive statistical analysis, three main considerations are in order: `(i)` the most users using MOOC-based e-learning platform are males; `(ii)` there is no correlation between age and the use of a platform[2]; and `(iii)` on average, the user attending either both platform or webinar only is 38 years old.

According to the TRBM procedure, in the first stage, 16 `best` covariates better fit the data explaining the 97.36% of the residual variance, with Posterior Inclusion Probabilities (PIPs) $\geq \tau$ in (4). Thus, there are 2^{16} potential submodel solutions ($M_{\tilde{j}i} \subset \mathcal{S}$). Because of the curse of dimensionality, the model and then the parameter space are further shrinked performing the second stage involved in the TRBM procedure. In Table 1, the final 8 `best` submodel classes ($M_{\xi i} \subset \mathcal{E}$) are showed, with PIPs $\geq \tau$ in (5) and lBF equals to 14.83 obtained through the $ALR(M_{\xi i})$ in (10). All variables are described in Table 3 on Appendix A.

Here, some considerations are in order. `(i)` Model uncertainty and overfitting involved in the variable selection procedure are avoided: indeed, the Conditional Posterior Sign (CPS) assumes values close to 1 or 0, indicating that a covariate has a positive or negative effect on the outcome, respectively. `(ii)` Satisfaction, usefulness, and application show the highest Posterior Inclusion Probabilities (PIPs) and positive effects on the outcomes. These findings would confirm the practical importance of an e-learning platform for the stundent. `(iii)` The discrete variable `age` shows a sufficient weight in explaining the variable of interest but with negative sign. It highlights that a younger student would be more likely to learn a new e-learning platform. `(iv)` The discrete variable `sex` tends to matter less than `age`, but keeping negative sign. Thus, let the construction of the categorical variable, a male user would be more likely to use an e-learning platform. `(v)` Finally, all 8 predictors will correspond to the ones to be accounted for the final solution, with PIPs $\geq \tau$. Thus, the model to be estimated is:

[2] The analysis has been conducted through the Pearson's test obtaining a p-value = 0.4516 and then the null of 'no correlation' cannot be rejected.

$$mooc\_web_i = \theta_0 + \theta_1 age_{1i} + \theta_2 sex_{2i} + \theta_3 utility_{3i} + \theta_4 use_{4i} + \theta_5 concentr_{5i} + \\ + \theta_6 attitude_{6i} + \theta_7 course_{7i} + \theta_8 satisf_{8i} + \epsilon_i \quad . \tag{25}$$

Table 1: `Best` Candidate Predictors – TRBM (`second stage`)

`Idx.`	`Predictor`	`Label`	`PIP (%)`	`CPS`
1	age	`age`	41.37	0.001
2	gender	`sex`	27.51	0.003
3	usefulness	`utility`	53.15	0.994
4	application	`use`	51.11	0.985
5	concentration	`concentr`	43.36	0.897
6	implication	`attitude`	36.63	0.983
7	assignment	`course`	31.84	0.896
8	satisfaction	`satisf`	63.75	0.995
-	benefits	`mooc_web`	-	-

The Table is so split: the first column denotes the predictor number; the second and the third column describe the predictors and the corresponding labels; the fourth column refers to the PIPs (in %); and the fifth column displays the CPS. The last row refers to the outcome of interest.

From a modelling perspective, according to (14), the Odds Ratio (OR) of the multinomial logistic regression in (26) are computed and displayed in Table 2. The results confirm the ones obtained in the variable selection procedure either in terms of estimates or sign. Indeed, all predictors are significant and the CPSs are observed. In addition, accouting for the two discrete variables within the system, an increasing of `age` and `sex` by one unit, the odds of $mooc\_web_i = 1$ decreases by $(0.30 - 1) \cdot 100 = -70\%$ and $(0.40 - 1) \cdot 100 = -60\%$, respectively. Or, the odds of $mooc\_web_i = 1$ are 0.30 and 0.40 timer lower when `age` and `sex` increase by one unit, respectively, keeping all other predictors constant. The relative OR for the other predictors can be computed similarly. Finally, the predictive probability running the (9) for any submodel solution $M_{\xi i}$ and holding all predictor values to their means, is $Pr(mooc\_web_i = 1) = 83\%$. This latter highlights that the multinomial logistic regression estimated through a TRBM procedure in high dimensional parameter space performs well and then there are benefits of e-learning systems – based on digital web technologies – in proving access to a high-quality institution.

These results are useful in guiding both policy makers and education providers. For policy makers, the opportunities arising from combining online training modules with traditional training are evident. This implies regulatory evolution that facilitates the introduction of hybrid modes in both secondary and undergraduate education. Regulatory evolution must necessarily be accompanied by an investment policy that enables colleges and universities to operate professionally and in a manner appropriate to evolving technological standards.

For individual educational institutions, be they secondary schools and universities, it will be important not only to upgrade the IT infrastructure and equipment, but also to train the administrative and teaching

Table 2: Multinomial Logistic Regression - Odds Ratio

Idx.	Label	Estimate	OR
0	intercept	0.334 (-)	1.396
1	age	−1.831 (**)	0.303
2	sex	−1.206 (*)	0.401
3	utility	2.452 (***)	0.207
4	use	2.234 (***)	0.253
5	concentr	1.914 (**)	0.897
6	attitude	0.495 (*)	0.609
7	course	0.223 (*)	0.559
8	satisf	3.544 (***)	0.125

The Table is so split: the first column denotes the predictor number; the second column describes the predictors according to their labels; and the last two columns display the estimates and odds ratio, respectively. The significant codes displayed in the third column stand for: (***) significance at 1%; (**) significance at 5%; (*) significance at 10%; and (-) no significance.

staff, making it possible to create up-to-date, high-level courses, as well as to make it possible for publics who cannot afford residential or full-time attendance to attend. Online modalities could then be useful in strengthening actions to support the reduction of school dropout, a very pronounced phenomenon in some countries. Lastly, especially for the European school and university sphere, the provision of online courses will go a long way toward promoting the internationalization and mobility of students not only from Europe, but also and especially from Asia and Africa. In this sense, online could help increase the competitiveness of education systems and their attractiveness. It will especially benefit those countries that already express tourist attractiveness and in particular those universities located in cities less traversed by tourist flows.

5 Concluding remarks

This paper involves a two-step robust Bayesian variable selection procedure in a multinomial logistic regression to evaluate student performance through convolutional neural network. The proposed methodology is obtained by combining a first-step Bayesian strategy for selecting potential predictors affecting the outcomes with a frequentist second-step procedure for estimating the parameters of the multiclass logit model. Multiclass conjugate informative proper mixture priors are addressed to design parameters' distributions, and MCMC algorithms are used to construct their exact posterior distributions and then perform accurate policy issues when investigating key technology factors potentially affecting e-learning tools.

An empirical application is developed by monitoring and evaluating students' level of attention and attendance throughout training courses, where the MOOC-based e-learning platform is used to evaluate student performance attending classes. The analysis refers to a Survey conducted on 56 students and aims to highlight the benefit of e-learning systems based on digital web technologies in proving access to a high-quality institution.

From a modelling perspective, the multinomial logistic regression estimated through robust Bayesian inference in a high dimensional context performs well highlight benefits of e-learning systems based on digital web technologies in proving access to a high-quality institution. From a policy perspective, the findings arising from combining online training modules with traditional training highlight that regulatory evolution has to necessarily be accompanied by an investment policy that enables colleges and universities to operate professionally and in an appropriate manner.

A Data Collection

Table 3 displays the predictors involved in the analysis according to the second stage of the TRBM procedure. They refer to the subset $\mathcal{E} \in \mathcal{S}$ (equation (5)) describing all potential `best` submodel solutions.

Table 3: Data Source – Submodel Class $\mathcal{E}$

Idx.	Predictor	Description
1	age	Discrete variable describing the age of the student
2	gender	Categorical variable = 1 whether the student is female, = 0 otherwise
3	usefulness	Utility using the e-learning platform (from 1 to 7)
4	application	Is the e-learning platform easy to use? (From 1 to 7)
5	concentration	Does the e-learning platform help focusing on a given assignment? (From 1 to 6)
6	implication	Would the student suggest the e-learning platform? (From 1 to 5)
7	assignment	Does the e-learning platform help focusing on a coursework? (From 1 to 7)
8	satisfaction	Satisfaction using the e-learning platform (from 1 to 6)
-	benefits	= 1 whether the user has benefited from either MOOC or webinar and = 0 otherwise

All variables refers to a Survey conducted on 56 students and aims to highlight the benefit of e-learning systems based on digital web technologies in proving access to a high-quality institution.

Compliance with Ethical Standards

Funding: No fund was used to conduct this study.

Conflict of Interest: The author declares no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

Athey, S. and Imbens, G. (2015). A measure of robustness to misspecification. *American Economic Review*, 105(5):476–480.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.

Awad, M. and Khanna, R. (2015). Support vector machines for classification. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Springer Nature, Switzerland (AG):39–66.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Breiman, L. (1992). The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error. *Journal of the American Statistical Association*, 87(419):738–754.

Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, 37(4):373–384.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(4):5–32.

Breiman, L. and Spector, P. (1992). Submodel selection and evaluation in regression. the x-random case. *International Statistical Review*, 60(3):291–319.

Chellappa, R., Wilson, C. L., and Sirohey, S. (1995). Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740.

Chen, J., Lv, Y., Xu, R., and Xu, C. (2019). Automatic social signal analysis: Facial expression recognition using difference convolution neural network. *Journal of Parallel and Distributed Computing*, 131:97–102.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B*, 56(3):501–514.

Gopalan, P., Hofman, J. M., and Blei, D. M. (2015). Scalable recommendation with hierarchical poisson factorization. *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pages 326–335.

Hasani, B. and Mahoor, M. H. (2017). Spatio-temporal facial expression recognition using convolutional neural networks and conditional random fields. pages 790–795.

Heinrich, K., Zschech, P., Janiesch, C., and Bonin, M. (2021). Process data properties matter: Introducing gated convolutional neural networks (gcnn) and key-value-predict attention networks (kvp) for next event prediction with deep learning. *Decision Support Systems*, 143:113–494.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Applications to nonorthogonal problem. *Technometrics*, 12(1):69–82.

Hu, P. J.-H. and Hui, W. (2012). Examining the role of learning engagement in technology-mediated learning and its effects on learning effectiveness and satisfaction. *Decision support systems*, 53(4):782–792.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of American Statistical Association*, 90(430):773–795.

Khan, N., Naim, A., Hussain, M. R., Naveed, Q. N., Ahmad, N., and Qamar, S. (2019). The 51 v's of big data: survey, technologies, characteristics, opportunities, issues and challenges. pages 19–24.

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 10:1755–1758.

Kratzwald, B., Ilić, S., Kraus, M., Feuerriegel, S., and Prendinger, H. (2018). Deep learning for affective computing: Text-based emotion recognition in decision support. *Decision Support Systems*, 115:24–35.

Madigan, D. and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using occam's window. *Journal of American Statistical Association*, 89:1535–1546.

Madigan, D., York, J., and Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 63(2):215–232.

Miller, A. J. (1984). Selection of subsets of regression variables. *Journal of the Royal Statistical Society: Series A*, 147(3):389–425.

Moses, Y., Edelman, S., and Ullman, S. (1996). Generalization to novel images in upright and inverted faces. *Perception*, 25:443–461.

Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.

Naim, A. and Alahmari, F. (2020). Reference model of e-learning and quality to establish interoperability in higher education systems. *International Journal of Emerging Technologies in Learning (iJET)*, 15(2):15–28.

Naim, A., Sattar, R. A., Al Ahmary, N., and Razwi, M. T. (2021). Implementation of quality matters standards on blended courses: A case study. *FINANCE INDIA Indian Institute of Finance*, 35(3).

O'Reilly, G. and Creagh, J. (2015). Does the shift to cloud delivery of courses compromise quality control.

Pacifico, A. (2020). Robust open bayesian analysis: Overfitting, model uncertainty, and endogeneity issues in multiple regression models. *Econometric Reviews*, 40(2):148–176. DOI: `https://doi.org/10.1080/07474938.2020.1770996`.

Penczynski, S. P. (2019). Using machine learning for communication classification. *Experimental Economics*, 22(4):1002–1029.

Picard, R. (1997). Affective computing cambridge. *MA: MIT Press [Google Scholar]*.

Pons, G. and Masip, D. (2017). Supervised committee of convolutional neural networks in automated facial expression analysis. *IEEE Transactions on Affective Computing*, 9(3):343–350.

Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of American Statistical Association*, 92(437):179–191.

Raftery, A. E., Madigan, D., and Volinsky, C. T. (1995). Accounting for model uncertainty in survival analysis improves predictive performance. *Bayesian Statistics*, 6:323–349.

Rosenberg, E. L. and Ekman, P. (2020). What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (facs).

Shashua, A. and Levin, A. (2001). Linear image coding for regression and classification using the tensor-rank principle. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 1:179–191. DOI: `https://doi.org/10.1109/CVPR.2001.990454`.

Talipu, A., Generosi, A., Mengoni, M., and Giraldi, L. (2019). Evaluation of deep convolutional neural network architectures for emotion recognition in the wild. pages 25–27.

Tipping, M. E. and Bishop, C. N. (1999). Mixtures of probabilistic principal component analysers. *Neural Computation*, 11(2):443–482.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311.

Turk, M. A. and Pentland, A. P. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–27.

Yang, B., Cao, J., Ni, R., and Zhang, Y. (2017). Facial expression recognition using weighted mixture deep neural network based on double-channel facial images. *IEEE access*, 6:4630–4640.

Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, 81(394):446–451.