Utilização do sucesso acadêmico para prever o abandono escolar de estudantes do ensino superior: um caso de estudo[1]

António Carlos Corte-Real de Sousa[2]
Carlos Alberto Bragança de Oliveira[2]
José Luís Cabral Moura Borges[2]

Resumo

O abandono escolar é um problema complexo que afeta a maioria dos programas de graduação pós-secundária, em todo o mundo. O curso de engenharia industrial do Instituto ISVOUGA, localizado em Santa Maria da Feira, Portugal, não é exceção. Este estudo usou um conjunto de dados contendo informações gerais dos estudantes e suas notas para as unidades curriculares já avaliadas. A partir deste conjunto de dados, foram selecionados dezessete preditores potenciais: cinco intrínsecos (gênero, estado civil, situação profissional, idade e regime de dedicação aos estudos – integral ou parcial) e doze extrínsecos (as notas em todas as doze unidades curriculares ministradas durante os dois primeiros semestres do curso). O objetivo principal desta investigação foi prever a probabilidade de um estudante abandonar o curso com base nos referidos preditores. Foi usada uma regressão logística binária para classificar os estudantes como tendo uma probabilidade alta ou baixa de não se reinscreverem no curso. Para validar se a metodologia utilizada é apropriada para o estudo em causa, a precisão obtida com o modelo de regressão logística foi comparada, por via de uma validação cruzada com cinco partições, com a precisão obtida pela utilização de três métodos muito utilizados em *data mining*: *One R*, *K Nearest Neighbors* e *Naive Bayes*. O modelo de regressão logística identificou quatro variáveis significativas na previsão do abandono escolar (as classificações nas unidades curriculares de ciência dos materiais, eletricidade, cálculo 1 e química). Os dois preditores mais influentes do abandono dos estudantes são não conseguir aprovação nas unidades curriculares menos exigentes: ciência dos materiais e eletricidade. Ao contrário do que seria de supor antes desta investigação, descobrimos que a não aprovação em unidades curriculares mais exigentes, como física ou estatística, não tem influência significativa no abandono escolar.

1- Agradecemos à Prof. Teresa Leão, diretora do Instituto Superior de Entre Douro e Vouga (ISVOUGA), por nos ter facultado o acesso à base de dados do Instituto e pelas ideias e conhecimento que partilhou relativas ao problema do abandono escolar no ISVOUGA. Este trabalho é financiado por Fundos FEDER através do Programa Operacional Competitividade e Internacionalização (COMPETE 2020) no âmbito do projeto POCI-01-0145-FEDER-006961 e por Fundos Nacionais através da Fundação para a Ciência e a Tecnologia (FCT) através do projeto UID/EEA/50014/2013.
2- Universidade do Porto, Porto, Portugal. Contato: a.sousa@doc.isvouga.pt; ORCID: http://orcid.org/0000-0002-6493-6161; braganca@fe.up.pt, ORCID: http://orcid.org/orcid.org/0000-0002-9505-8170; jlborges@fe.up.pt; ORCID: http://orcid.org/0000-0001-9946-5614.

DOI: http://dx.doi.org/10.1590/S1678-4634201844180590

Palavras chaves

Abandono escolar – Retenção – Regressão logística – Data mining.

Using Academic Performance to Predict College Students Dropout: a case study

Abstract

Student dropout is a complex problem that affects most post-secondary undergraduate programs, all over the world. The Industrial Engineering program of the ISVOUGA Institute, located in Sta. Maria da Feira, Portugal, is no exception. This research used a dataset containing students' general information and the students' marks for the already assessed courses. From this dataset, 17 potential predictors have been selected: five intrinsic predictors (gender, marital status, professional status, full/part time student, and age) and 12 extrinsic ones (the marks in all the 12 courses taught during the first two semesters of the program). The main goal of this research was to predict the likelihood of a student to dropout, based on the referred predictors. A binary logistic regression was used to classify students as having a high or low probability not to re enroll the program. To validate the appropriateness of the used methodology, the accuracy of the logistic model was compared, by means of a 5-fold cross-validation, to the accuracy of three classification methods commonly used in Data Mining: One R, K Nearest Neighbors, and Naive Bayes. Four variables were significant to the logistic model (the marks in Materials Science, Electricity, Calculus 1, and Chemistry). The two most influential predictors for student dropout are failing to pass in the less challenging courses of Materials Science and Electricity. Contrary to what we would think prior to this research, we found that failing in more challenging courses such as Physics or Statistics does not have a significant influence on student dropout.

Keywords

Student dropout – Retention – Logistic regression – Data mining.

Introdução

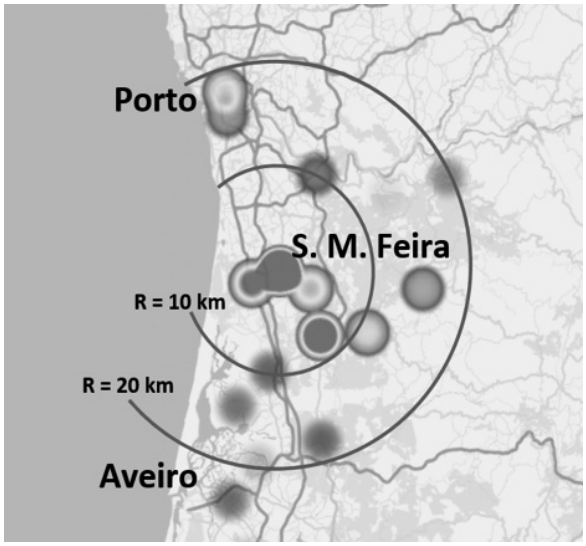
A educação é um desafio para a maioria dos países. Em 2016, Portugal ainda estava abaixo do nível de escolaridade da UE22 para os estudantes com idade na faixa 25 a 34 anos, tendo estes, uma taxa de conclusão de educação pós-secundária de 35%. A correspondente taxa média da UE22 era de 41%. No que se refere a uma maior faixa de

idades (25 a 64 anos), em 2016 só 24% da população adulta portuguesa atingiu o ensino pós-secundário, valor abaixo da média da OCDE de 37% (OCDE, 2017).

O Instituto Superior de Entre o Douro e Vouga (ISVOUGA) é uma escola privada de graduação pós-secundária, com 350 estudantes em 2015. Está localizado em Santa Maria da Feira, uma pequena cidade com população de cerca de 12.500 habitantes (INE, 2011). Santa Maria da Feira está localizada em Portugal, 25 km a sul do Porto e 30 km a norte de Aveiro. O Porto é a segunda maior cidade portuguesa com uma população de 250.000 habitantes (INE, 2011), com quarenta instituições de ensino pós-secundário, públicas ou privadas, uma das quais é a pública Universidade do Porto, com 30.000 estudantes com um *ranking* no intervalo 301 a 400 da Shanghai Ranking Consultancy (ARWU, 2016). Aveiro é uma cidade de tamanho médio com uma população de 65.000 habitantes (INE, 2016), com cinco instituições de ensino pós-secundário, públicas ou privadas, uma das quais é a pública Universidade de Aveiro com 12.000 estudantes e um *ranking* no intervalo 401 a 500 da Shanghai Ranking Consultancy (ARWU, 2016). A região em que o ISVOUGA está localizado pertence a uma região mais ampla, denominada Entre o Douro e Vouga.

Os estudantes do ISVOUGA vêm de povoações localizadas num raio de 25 km em redor de Santa Maria da Feira e destes, mais de 90% vivem a menos de 10 km do ISVOUGA. Em média, entre 2010 e 2014, na região de Entre o Douro e Vouga, apenas 2,3% da população residente com idade entre 18 e 22 anos frequentou cursos de graduação pós-secundária, enquanto a média nacional correspondente foi de 31,8% (INE, 2016). Esse pequeno valor significa que a maioria dos jovens da região abandona a escola no final do ensino secundário, sem terem frequentado nenhum curso pós-secundário. A idade média dos calouros do ISVOUGA foi de 25,1 anos em 2015. Este valor elevado da média de idade dos calouros sugere que a maioria dos estudantes interrompeu os estudos no passado, voltando para a escola vários anos depois.

Figura 1– Proveniência dos estudantes do ISVOUGA



Fonte: arquivo dos autores.

Frequentar um curso pós-secundário é hoje uma expectativa comum para as pessoas de diferentes contextos socioculturais (MÍNGUEZ; SAN JULIÁN, 2013; PÁRAMO FERNÁNDEZ et al., 2017). No entanto, pode ser uma tarefa exigente para a maioria dos estudantes, pois envolve uma variedade de fatores intrínsecos (relacionados com os estudantes) e extrínsecos (relacionados com a escola, histórico social e histórico econômico). O abandono escolar é um problema complexo que afeta a maioria dos cursos de graduação pós-secundária.

As características da região de Entre o Douro e Vouga criam complexos desafios à direção do ISVOUGA, por um lado, identificar e compreender os problemas externos que podem desencorajar a inscrição de novos estudantes no instituto e, por outro lado, as questões internas que podem levar ao abandono escolar dos estudantes atuais. Nas instituições públicas, as elevadas taxas de abandono escolar resultam no desperdício de dinheiro dos contribuintes e numa população com baixo nível de escolaridade, o que leva, consequentemente, a menores oportunidades de emprego para cargos que exigem alta qualificação (PAURA; ARHIPOVA, 2014; LITALIEN; GUAY, 2015). Numa instituição privada como o ISVOUGA, que não é apoiada por fundos públicos, as altas taxas de abandono não implicam desperdício de dinheiro dos contribuintes, mas de recursos pessoais dos estudantes (tempo e dinheiro). Para evitar o abandono voluntário (isto é, quando os alunos decidem não se inscrever novamente), o ISVOUGA é particularmente cuidadoso na definição de políticas que possam ajudar a monitorizar e apoiar os calouros da escola.

A explicação e a previsão do desempenho acadêmico dos estudantes tem sido amplamente estudada, principalmente após os anos 1980. Vários estudos focam-se em conceitos teóricos relacionados a esses assuntos. Por exemplo, Pascarella e Terenzini (1980), Astin (1984, 1993), Kuh (2003) e outros estudaram as relações entre características acadêmicas, sociais, pessoais e emocionais e a adaptação e taxa de retenção dos estudantes. Essas investigações são úteis para a avaliação institucional e para uma compreensão social do assunto, mas não são facilmente transponíveis para ações de correção que melhorem a integração e o envolvimento dos estudantes. Tinto (1982), propõe vários tipos de razões que influenciam a probabilidade de abandono do ensino superior: condições financeiras dos estudantes; gênero; integração acadêmica e social; interesses, competências, valores e compromissos com os objetivos do ensino superior e com a instituição específica; competências estudantis (acadêmicas, sociais ou de outra forma) e/ou capacidades intelectuais; interesse, compromisso e motivação dos estudantes para finalizar um curso.

De acordo com o modelo de Tinto, um maior grau de integração está diretamente relacionado ao maior comprometimento com a instituição de ensino e com o objetivo de conclusão dos estudos. Stratton, O'Toole, e Wetzel (2007) concluem que fatores como a inscrição a tempo integral ou a tempo parcial não são os fatores mais importantes na decisão de abandonar os estudos.

Após a identificação dos estudantes em risco de abandonar os estudos, Neild, Balfanz e Herzog (2007) bem como Litalien and Guay (2015) sugerem várias estratégias

que podem ajudar a manter esses alunos no caminho de concluir o curso. Sugerem que a utilização de recursos motivacionais e de apoio psicológico através de tutores/conselheiros, professores ou até de outros estudantes, podem influenciar a diminuição das taxas de retenção e das intenções de abandono escolar. Adicionalmente, a orientação por colegas estudantes bem-sucedidos, pode moldar hábitos de estudo nos calouros, resultando num aumento nas taxas de aprovação, integração social e no envolvimento com a comunidade universitária (MORALES; AMBROSE-ROMAN; PEREZ-MALDONADO, 2016). Usando uma perspectiva semelhante à da orientação por colegas estudantes, alguns estudos identificam que a instrução suplementar é um método complementar na ajuda aos alunos para melhorarem seu desempenho em unidades curriculares consideradas difíceis (MALM; BRYNGFORS; MÖRNER, 2012; MARTÍNEZ-LÓPEZ et al., 2014; MALM; BRYNGFORS; MÖRNER, 2015). A instrução suplementar é uma ação que consiste no desenvolvimento de atividades colaborativas, sob a orientação de um estudante sênior. Alguns autores também sugerem melhorias em programas introdutórios para os novos estudantes e mudanças nos métodos de avaliação (HOVDHAUGEN, 2011).

O curso de engenharia industrial do ISVOUGA é um curso de primeiro ciclo com seis semestres (nível de bacharelado) de acordo com Bologna Working Group on Qualifications Frameworks (2005). Tem como base a engenharia mecânica e é um curso com 180 créditos ECTS (Sistema Europeu de Transferência de Créditos). Os estudantes podem candidatar-se a este curso se tiverem concluído com sucesso o ensino secundário. No entanto, estudantes com mais de 23 anos podem candidatar-se sem terem completado o ensino secundário. Neste último caso, os estudantes devem enviar o seu currículo ao conselho científico do Instituto, efetuar um exame de avaliação e uma entrevista com um comitê de avaliação. O curso de engenharia industrial é um dos cinco cursos de licenciatura oferecidos, presentemente, pelo ISVOUGA. Este curso oferece os conhecimentos e competências fundamentais requeridos a um engenheiro industrial, em áreas como matemática, física, desenho industrial, ciência dos materiais, mecânica dos sólidos e fluidos, estatística e análise de dados. Para obter uma avaliação objetiva do nível intrínseco de dificuldade de cada unidade curricular do primeiro e segundo semestres do curso, foi calculada uma classificação para o nível de dificuldade de cada unidade curricular. Esta classificação tem dois componentes: as notas médias padronizadas de cada unidade curricular e o número médio de anos até obter aprovação nessa unidade curricular. A classificação final é obtida pela média ponderada das duas componentes, considerando um peso de 40% para as notas médias padronizadas e um peso de 60% para o número de anos até obter aprovação. A estrutura do primeiro ano do curso de engenharia industrial e as pontuações correspondentes são apresentadas na Tabela 1.

Tabela 1– Unidades curriculares do 1º ano (1º e 2º semestre)

Unidade Curricular	Semestre	Classificação média [0-20]	Número médio de tentativas até obter aprovação	Nível de dificuldade da unidade curricular * 1 – Mais Fácil, 4 – Mais difícil
Álgebra linear e geometria descritiva	1	12.3	1.7	1
Ciência dos materiais	1	12.1	2.5	2
Física	1	11.2	3.8	4
Folhas de cálculo em engenharia	1	13.4	3.0	3
Desenho técnico 1	1	13.8	1.1	1
Programação VB	1	13.2	2.7	3
Cálculo 1	2	11.7	4.8	4
Química	2	11.2	4.5	4
Eletricidade	2	12.9	1.6	2
Estatística	2	11.9	3.2	3
Investigação operacional	2	12.1	2.1	2
Desenho técnico 2	2	12.7	1.4	1

* Baseado no número de ordem (*rank*) dos valores das medias ponderadas das médias padronizadas (ponderação de 40%) e do número de anos até obter aprovação (ponderação de 60%)
Fonte: base de dados ISVOUGA.

Entre 2009 e 2013, no curso de engenharia industrial, a nota média global para os alunos que concluíram o curso foi de 12,98 (numa escala de 20) e necessitaram, em média, de 7,38 semestres para completar com sucesso todas as unidades curriculares lecionadas ao longo dos seis semestres. No período de 2007 a 2013, uma média de 37 calouros inscreveram-se anualmente no curso. As unidades curriculares do primeiro ano do curso (1º e 2º semestres) mais exigentes e historicamente consideradas difíceis são: cálculo 1, física e química. A taxa média de abandono foi ligeiramente superior a 30% e, dos calouros que se inscreveram no curso, a maioria dos abandonos (66%) ocorreu no final do 2º semestre.

O objetivo desta investigação, foi identificar os fatores que, ao nível escolar (desempenho acadêmico), aumentam a probabilidade de um abandono voluntário do curso de engenharia industrial.

Neste trabalho, foi considerado que um estudante abandonou o curso quando não obteve o respectivo diploma, obteve pelo menos seis notas finais em unidades curriculares e também verifica uma das duas seguintes condições: não renovou ou anulou a sua inscrição ou não participou de nenhuma avaliação para qualquer das unidades curriculares por um período de tempo superior a um ano.

Focamo-nos na análise do abandono escolar neste curso devido a sua estrutura geral estável, que sofreu apenas pequenas alterações desde seu início em 1996, e porque os autores desta investigação são engenheiros industriais, fato que lhes confere uma

importante visão sobre as características da maioria das unidades curriculares e das particularidades do curso.

Como caso de estudo, esta investigação focou-se no desempenho acadêmico dos estudantes nos dois primeiros semestres do curso (as classificações categóricas utilizadas foram *passar / não passar* nas doze unidades curriculares ministrados durante os dois primeiros semestres do curso) e nas características intrínsecas do aluno (gênero, estado civil, situação profissional, estudante a tempo integral/parcial e idade no início do curso). Esta investigação não teve como objetivo analisar fatores sociais, econômicos e psicológicos que possam também influenciar o abandono escolar (por exemplo, características sociodemográficas do estudante, motivações para o estudo, integração social e acadêmica na escola, condições de vida etc.). Estes fatores demonstraram estar relacionados com altas taxas de abandono escolar (JORDAN; LARA; McPARTLAND, 1994; PIERRAKEAS et al., 2004; PAURA; ARHIPOVA, 2014).

Dos dezessete preditores potenciais selecionados a partir da base de dados com as informações dos estudantes, só preditores considerados como fatores extrínsecos (as classificações dos estudantes) seriam suscetíveis a ações práticas que, caso fossem incluídos como significativos no modelo, pudessem diminuir a probabilidade de abandono escolar. As características intrínsecas não podem ser alteradas, mas foram incluídas como possíveis preditores para avaliar se têm influência nos resultados. O modelo logístico não selecionou nenhum dos preditores intrínsecos como fator significativo promotor de abandono escolar.

O resultado desta investigação foi a identificação de uma lista de unidades curriculares do primeiro ano do curso de engenharia industrial que têm uma grande influência no abandono escolar dos estudantes e a validação de uma metodologia que, com base nas notas dos estudantes nessas unidades curriculares, permite uma antecipação probabilística de estudantes com forte probabilidade de vir a abandonar o curso. Estes resultados forneceram à direção do ISVOUGA uma informação importante para abordar o problema do abandono escolar.

Dados e métodos

Dados

Como caso de estudo, a população alvo desta investigação é o conjunto de estudantes que frequentam o curso de engenharia industrial do ISVOUGA. Para criar um modelo de previsão do abandono dos estudantes durante o primeiro ano, tivemos acesso à base de dados acadêmicos do Instituto. Foram usados como dados de amostra os registros dos estudantes que frequentaram este curso de 2007 até 2013 (192 registros de estudantes, sendo que 154 registros foram usados para treino e 38 registros para validação), depois de terem sido limpos[3].

3- Processo de detecção, diagnóstico e edição de dados defeituosos.

A base de dados do Instituto contém informações pessoais dos estudantes (curso selecionado, nome, gênero, estado civil, situação profissional, tempo integral/parcial, idade no início do curso), bem como as datas e notas finais nas unidades curriculares para as quais os estudantes foram avaliados até outubro de 2013. Primeiramente, removemos todas as informações pessoais não relevantes para o objetivo deste trabalho, tornando os dados anônimos. Os dados foram seguidamente pré-processados através da limpeza de registros errados / incompletos, mantendo apenas registros com ao menos seis notas finais em qualquer das unidades curriculares dos seis semestres do curso. Como resultado deste pré-processamento, apenas 192 dos 310 registros iniciais foram mantidos.

Dos 192 registros, 87 (45,3%) foram classificados como Y = 1 (estudante que abandonou o curso) e 105 (54,7%) foram classificados como Y = 0 (estudante que não abandonou o curso). Dezessete variáveis foram consideradas preditoras / regressoras potenciais, conforme Tabela 2. Os cinco primeiros preditores (x_1 a x_5) representam fatores intrínsecos (relacionados com o estudante) e os doze restantes (x_6 a x_{17}) representam fatores extrínsecos (relacionados a instituições).

Tabela 2– Variáveis consideradas preditores potenciais da variável dependente

	Variável	Descrição	Tipo
	X_1	Gênero	Categórica
	X_2	Estado civil	Categórica
	X_3	Situação profissional	Categórica
	X_4	Estudante a tempo integral ou parcial	Categórica
	X_5	Idade no início do curso	Numérica
Classificações nas unidades curriculares do 1º ano (escala 0 – 20)	X_6	Álgebra linear e geometria descritiva	Numérica
	X_7	Ciência dos materiais	Numérica
	X_8	Física	Numérica
	X_9	Folhas de cálculo em engenharia	Numérica
	X_{10}	Desenho técnico 1	Numérica
	X_{11}	Programação VB	Numérica
	X_{12}	Cálculo 1	Numérica
	X_{13}	Química	Numérica
	X_{14}	Eletricidade	Numérica
	X_{15}	Estatística	Numérica
	X_{16}	Investigação operacional	Numérica
	X_{17}	Desenho técnico 2	Numérica

Fonte: dados da pesquisa.

A multicolinearidade entre as variáveis foi avaliada utilizando o Fator de Inflação de Variância (VIF). Como regra prática, utiliza-se geralmente um valor de VIF=5 como o limiar a partir do qual há indicação de multicolinearidade (MENARD, 2001). Todas as

treze variáveis numéricas obtiveram um valor VIF menor do que 3.1, indicando um nível moderado de multicolinearidade entre as variáveis, que permitiu que todas as variáveis tenham sido usadas como preditores potenciais.

Métodos

Como a variável de saída, Y, é categórica podendo tomar dois resultados (1 – estudante que abandonou o curso, 0 – estudante que não abandonou o curso), a aproximação para P(Y|X) é um problema de classificação. Realizamos uma regressão logística binária para classificar os alunos como 0 ou 1. Neste contexto, uma regressão logística binária devolve a probabilidade condicional de um aluno pertencer à classe 1, dado que apresenta como valores os regressores X (ver Tabela 2). A expressão correspondente é dada pela equação (1).

Equação (1) $$p(class = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta X)}}$$

A regressão logística usa o método de máxima verossimilhança para estimar os valores de coeficientes para β_1, β_2,..., β_n. Quando o regressor X_j com um coeficiente β_j, aumenta em uma unidade, controlando as demais variáveis, as probabilidades, $p/(1\text{-}p)$, aumentar por uma quantidade multiplicativa de $e^{\beta j}$, em que p é a probabilidade associada à classe 1.

As variáveis indicadas na Tabela 2 são os regressores que foram inicialmente utilizados no modelo logístico. Depois de executar o modelo com estas variáveis, repetimos o processo de regressão logística usando um conjunto de novas variáveis, denominadas x'_6 a x'_{17}, definidas para substituírem as variáveis originais, x_6 a x_{17}. Este novo conjunto de variáveis (x'_6 a x'_{17}) são as variáveis x_6 a x_{17} transformadas para uma escala categórica, usando a seguinte transformação: se $x_i \geq 10$, então $x'_i = 0$ (o estudante foi aprovado na unidade curricular), se $x_i < 10$, então $x'_i = 1$ (o estudante não foi aprovado na unidade curricular). Na classificação dos estudantes, o modelo com as variáveis categóricas produziu um nível global de precisão semelhante ao modelo que com variáveis numéricas. Como a interpretação dos resultados do modelo que utiliza as variáveis categóricas x'_6 a x'_{17} para representar as classificações dos estudantes nas unidades curriculares do primeiro ano do curso é mais fácil do que a interpretação do modelo que utiliza as classificações numéricas x_6 a x_{17}, prosseguimos esta investigação utilizando o modelo com variáveis categóricas. A utilização de uma amostra com uma dimensão de 192 estudantes, valor maior que o recomendado de pelo menos cem observações para modelos de regressão logística (LONG, 1997), proporcionou um elevado nível de confiança na interpretação dos coeficientes de regressão que obtivemos.

Complementarmente, para validar a adequação da metodologia utilizada nesta investigação, foram usados três métodos de classificação comumente usados em *data mining* (HAND, 2007) para comparar o nível geral de precisão obtido por estes com o nível de precisão obtido através da regressão logística binária. Estes métodos foram: a) *One R*;

b) *K Nearest Neighbors* (KNN); c) *Naive Bayes*. O *One R* é uma metodologia muito simples que induz regras de classificação baseadas apenas no valor de um único preditor. O KNN (COVER; HEART, 1967) é um algoritmo que classifica os dados com base numa distância que, nesta investigação, foi a métrica *value difference metric* (WILSON; MARTINEZ, 1997). O classificador *Naive Bayes* baseia-se na aplicação do Teorema de Bayes com pressupostos de independência condicional entre todos os preditores (RUSSEL; NORVIG, 1995). Uma metodologia de validação cruzada (REFAEILZADEH; TANG; LIU, 2009) foi utilizada para comparar o desempenho relativo dos quatro métodos. Na validação cruzada, usamos cinco partições, pois o tamanho da amostra era demasiado pequeno para usar um número maior. Mantivemos as proporções de Y = 1 e Y = 0 em cada subconjunto equilibrada com os valores correspondentes da amostra global, isto é, 45,3% de Y = 1 (estudante que abandonou o curso) e 54,7% de Y = 0 (estudante que não abandonou o curso). A amostra de 192 de dados foi dividida em dois grupos: um de 154 registos (quatro das cinco partições), usado para treinar o modelo, e um de 38 registos (a restante das cinco partições), usado para validar o modelo. O modelo logístico final foi construído usando todos os 192 registos da amostra e as estatísticas de qualidade e precisão foram calculadas utilizando este modelo final.

Resultados

Nesta seção, apresentamos os resultados do modelo de regressão logística. Como estatísticas para avaliação da qualidade do modelo apresentamos:

a) o logaritmo da função de verossimilhança associada ao modelo reduzido (modelo que dá probabilidade p_0 independentemente dos valores dos preditores independentes);
b) o logaritmo da função de verossimilhança associada ao modelo completo (o modelo que inclui os preditores independentes);
c) a razão de verossimilhanças;
d) o pseudo R^2 de McFadden.

Fizemos uma regressão passo a passo (*stepwise*) para selecionar os preditores. Para cada preditor, é apresentado o valor de p ($Pr > \chi^2$) correspondente ao nível de significância. Se este valor for menor que o limite de significância (foi utilizado o valor 0.10), a contribuição do preditor para o ajuste do modelo é considerada significativa. A qualidade global do modelo é apresentada por meio da matriz de confusão.

A comparação entre o modelo logístico binário e os três métodos de classificação alternativos é apresentada por meio de três tabelas de resumo dos métodos de validação cruzados com utilização de cinco partições:

a) as taxas globais de acerto para todos os métodos de classificação, para cada partição de validação cruzada e os valores médios correspondentes;

b) as taxas de especificidade para todos os métodos de classificação, para cada partição de validação cruzada e os valores médios correspondentes;
c) as taxas de sensibilidade para todos os métodos de classificação, para cada partição de validação cruzada e os valores médios correspondentes.

A análise estatística foi realizada usando o software XLSTAT-Pro, Version 05.33993, 2016, Addinsoft, Inc., USA.

O modelo logístico binário para o abandono escolar

As estatísticas de qualidade de ajuste apresentadas na Tabela 3 mostram que o modelo é uma boa representação da relação entre o abandono escolar e o sucesso/falha dos estudantes nas doze unidades curriculares do primeiro ano do curso. De fato, $R^2_{McF} = 0.502$ revela que o modelo se ajusta bem aos dados, e a razão de verossimilhança (χ^2) de134.57[4] significa que o modelo completo (o modelo que inclui os preditores independentes) é significativamente melhor do que o modelo independente (o que dá probabilidade p_0 quaisquer que sejam os valores dos preditores).

Tabela 3– Estatísticas de qualidade de ajuste para o modelo de regressão logística

Estatística	Modelo Independente	Modelo Completo	Razão de Verossimilhança (χ^2)	Graus de Liberdade	Pr > χ^2
-2 Log Likelihood	267.860	133.290	134.57	4	<0.0001
R^2 (McFadden)	0.000	0.502	-	-	-

Fonte: dados da pesquisa.

Apenas quatro dos dezessete potenciais preditores foram incluídos no modelo. O nível de significância dos quatro preditores incluídos no modelo é apresentado na Tabela 4.

Tabela 4– Nível de significância dos preditores incluídos no modelo

Preditor	Preditor/Acrônimo	Tipo	Pr > LR
X'_7	Ciência dos materiais (Mat_Sc)	Categórica (0-pass / 1-fail)	<0.001
X'_{14}	Eletricidade (Elect)	Categórica (0-pass / 1-fail)	<0.003
X'_{12}	Cálculo 1 (Cal_1)	Categórica (0-pass / 1-fail)	<0.059
X'_{13}	Química (Che)	Categórica (0-pass / 1-fail)	<0.078

Fonte: dados da pesquisa.

A Equação 2 mostra a expressão para o cálculo da probabilidade de um aluno estar na classe 1 (um aluno que abandonou), condicional aos preditores *X'*.

4- $P(\chi^2_{GL=4})>134.57)<0.0001$

Equação (2)

$$P(Class=1 \mid X') = \frac{1}{1+e^{-(-2.964+2.058*Mat\_Sc+1.430*Elect+0.927*Cal\_1+1.116*Che)}}$$

Quando todos os quatro preditores *X*' (Mat_Sc, Elect, Cal_1, and Che marks) são iguais a zero (passar na unidade curricular), a probabilidade de abandono é:

Equação (3) $P(Class=1 \mid X'=0)=0.049$

Mudando, alternadamente, cada um dos preditores de 0 para 1 (de passar na unidade curricular para não passar na unidade curricular), mantendo os valores dos três restantes preditores como zero, as probabilidades de abandono passam a ser as seguintes:

Equação (4) $P(Class=1 \mid Mat\_Sc=1,Elect=0,Cal\_1=0,Che=0)=0.288$

Equação (5) $P(Class=1 \mid Elect=1,Mat\_Sc=0,Cal\_1=0,Che=0)=0.177$

Equação (6) $P(Class=1 \mid Cal\_1=1,Mat\_Sc=0,Elect=0,Che=0)=0.115$

Equação (7) $P(Class=1 \mid Che=1,Mat\_Sc=0,Elect=0,Cal\_1=0)=0.136$

Quando todos os quatro preditores (Mat_Sc, Elect, Cal_1, and Che marks) são iguais a um (não passar na unidade curricular), a probabilidade de abandono é

Equação (8) $P(Class=1 \mid X'=1)=0.929$

A coerência do modelo e os resultados de precisão foram avaliados usando validação cruzada com cinco partições. A Tabela 5 apresenta a precisão (a proporção do número total de previsões corretas), a sensibilidade (a proporção de estudantes que abandonaram o curso e que foram classificados corretamente) e a especificidade (a proporção de estudantes que não abandonaram o curso e que foram classificados corretamente) obtidos pelo modelo logístico usando como dados de entrada os 38 registos da partição utilizada para validar o modelo que tinha sido previamente treinado com os 154 registos das restantes quatro partições. Estes resultados mostram que o modelo de regressão logística binária obtido tem uma boa precisão geral na classificação dos estudantes que abandonaram o curso.

Tabela 5– Resultados relativos à precisão preditiva do modelo de regressão logística binária, utilizando como dados de entrada os registos da amostra de validação: sumário da validação cruzada

	Partição 1	Partição 2	Partição 3	Partição 4	Partição 5	MÉDIA
Precisão	82%	85%	84%	90%	84%	**85%**
Especificidade	90%	86%	86%	90%	76%	**86%**
Sensibilidade	71%	83%	82%	89%	94%	**84%**

Fonte: dados da pesquisa.

Resultados obtidos com outros métodos de classificação

Os resultados dos modelos de classificação *One R, KNN* e *Naive Bayes* foram obtidos executando cada modelo usando como dados de entrada os 38 registos da partição utilizada para validar o modelo.

a) Método *One R*

Esta regra de classificação baseia-se no valor de um único preditor. Na avaliação cruzada com cinco partições, dois dos quatro preditores que foram selecionados como estatisticamente significativos pelo modelo logístico binário (ver Tabela 4) foram os escolhidos em diferentes partições como o único classificador explicativo para o abandono escolar: as classificações em eletricidade (0 – obteve aprovação na unidade curricular / 1 – não obteve aprovação na unidade curricular) foram selecionadas em três das cinco partições e as classificações em ciência dos materiais (0 – obteve aprovação na unidade curricular / 1 – não obteve aprovação na unidade curricular) foram selecionadas em duas das cinco partições. A Tabela 6 apresenta os resultados de precisão, sensibilidade e especificidade obtidos utilizando a metodologia *One R*.

Tabela 6– Precisão preditiva do modelo *One R*: sumário da validação cruzada

	Partição 1	Partição 2	Partição 3	Partição 4	Partição 5	MÉDIA
Precisão	68%	63%	58%	61%	61%	**62%**
Especificidade	90%	95%	86%	90%	86%	**89%**
Sensibilidade	41%	28%	24%	28%	29%	**30%**

Fonte: dados da pesquisa.

b) K Nearest Neighbors, com K=7

Esta metodologia classifica os dados de entrada com base numa distância calculada usando os quatro preditores que foram usados no modelo logístico binário (ver Tabela 4). A Tabela 7 apresenta os resultados de precisão, sensibilidade e especificidade obtidos executando a metodologia *K Nearest Neighbors*, usando como dados de entrada os 38 registos da partição utilizada para validar o modelo. Foram testados vários valores para K no intervalo entre 3 e 10. O valor K = 7 foi selecionado, pois produziu melhores resultados.

Tabela 7– Precisão preditiva do modelo *K Nearest Neighbors*: sumário da validação cruzada

	Partição 1	Partição 2	Partição 3	Partição 4	Partição 5	MÉDIA
Precisão	82%	84%	84%	87%	82%	**84%**
Especificidade	90%	85%	86%	85%	76%	**84%**
Sensibilidade	71%	83%	83%	89%	88%	**83%**

Fonte: dados da pesquisa.

c) *Naive Bayes*

Esta metodologia classifica os dados de entrada com base no Teorema de Bayes com fortes pressupostos de independência. A metodologia emprega os quatro preditores do modelo logístico binário (ver Tabela 4). A Tabela 8 apresenta resultados de precisão, sensibilidade e especificidade obtidos executando a metodologia *Naive Bayes*, usando como dados de entrada os 38 registos da partição utilizada para validar o modelo.

Tabela 8– Precisão preditiva do modelo *Naive Bayes*: sumário da validação cruzada

	Partição 1	Partição 2	Partição 3	Partição 4	Partição 5	MÉDIA
Precisão	82%	89%	84%	92%	84%	86%
Especificidade	90%	95%	90%	91%	81%	90%
Sensibilidade	71%	83%	76%	94%	88%	82%

Fonte: dados da pesquisa.

Discussão e conclusões

Como caso de estudo, a análise que foi realizada nesta investigação foi destinada a ser transposta para ações práticas que possam diminuir a probabilidade de abandono voluntário de um curso específico do ISVOUGA. Os resultados obtidos não são facilmente comparáveis com os resultados de outras investigações na mesma área, pois a maioria dos estudos (pelo menos aqueles que conhecemos) basicamente tentam explicar o abandono escolar usando como variáveis explicativas o corpo docente, o corpo estudantil e os contextos institucional, social e familiar (ARAQUE; ROLDÁN; SALGUERO, 2009).

Nesta investigação, examinamos o efeito de fatores intrínsecos (relacionados com o estudante) e extrínsecos (relacionados com a instituição) no abandono dos estudantes matriculados no curso de engenharia industrial do ISVOUGA. Os dados de 192 estudantes, com informações gerais (curso selecionado, nome, gênero, estado civil, situação profissional, estudante a tempo integral/parcial, idade no início do curso) e as classificações dos alunos foram utilizados para produzir um conjunto de preditores para o abandono escolar. Dezessete variáveis foram consideradas preditores potenciais (ver Tabela 2).

Realizamos uma regressão logística binária que classificou os alunos como 0 (aluno que não abandonou o programa) ou 1 (aluno que abandonou o programa), usando a probabilidade de o estudante pertencer a cada classe e um limite de probabilidade de 0,5 para separar as duas classes. Foram usados ainda três métodos de classificação comumente usados em *data mining* para comparar o nível geral de precisão com o nível de precisão obtido usando a regressão logística binária.

A regressão logística binária é um método estatístico clássico usado para classificar/prever a probabilidade de um item pertencer a uma de duas classes, e é também tulizado em *data mining*. Outros métodos de classificação de *data mining* podem ser usados em lugar da regressão logística binária (nós utilizamos três deles: *One R*, *K Nearest Neighbors*, e *Naive Bayes*). Nesta investigação, o modelo de regressão logística binária produziu um

nível de precisão geral muito bom (ver Tabela 5), comparável com o nível de precisão do método *Naive Bayes*, e superior ao nível de precisão dos métodos *One R* ou *K Nearest Neighbors*. Alguns autores questionam as técnicas de *data mining* por necessitarem de um longo processo de treino, pela incapacidade de identificar a importância relativa dos preditores de entrada e por dificuldades interpretativas (SABZEVARI; SOLEYMANI; NOORBAKHSH, 2007). O modelo logístico binário produziu previsões de alta qualidade, usando um número reduzido de preditores que foram selecionados com base na sua significância estatística. Este fato facilita a compreensão da contribuição de cada preditor para o ajuste do modelo. Além disso, a equação correspondente ao modelo obtido, equação 2, permite a avaliar o nível de influência de cada preditor nos resultados do modelo.

Quatro das dezessete variáveis disponíveis foram consideradas significativas para o modelo e incluídas como preditores. Estas quatro variáveis correspondem às classificações, expressas numa escala categórica (0 – obteve aprovação na unidade curricular / 1 – não obteve aprovação na unidade curricular), obtidas pelos estudantes em quatro unidades curriculares do primeiro ano do curso de engenharia industrial: ciência dos materiais, eletricidade, cálculo 1 e química. Os dois primeiros preditores incluídos no modelo de regressão logística (ciência dos materiais e eletricidade) são substancialmente mais influentes no resultado do modelo do que cálculo 1 e química. É interessante notar que os dois preditores mais influentes no abandono escolar são as classificações nas unidades curriculares menos exigentes, isto é, ciência dos materiais (nível de dificuldade 2) e eletricidade (nível de dificuldade 2), conforme nível de dificuldade/exigência de todas as unidades curriculares na Tabela 1 (coluna 5). Não obter aprovação em unidades curriculares mais exigentes como cálculo 1 (nível de dificuldade 4) ou química (nível de dificuldade 4) tem menos influência no abandono escolar do que não obter aprovação nas unidades curriculares ciência dos materiais ou eletricidade. Nenhuma das variáveis X_1 – gênero, X_2 – estado civil, X_3 – situação profissional, X_4 – estudante em tempo integral/parcial ou X_5 – idade no início do curso foram incluídas no modelo, por não serem estatisticamente significativas. Quando na equação 2 todos os quatro preditores são iguais a 1 (não obter aprovação na unidade curricular), a probabilidade de abandono é 0.929, sendo de 0.049 quando todos os quatro preditores são iguais a 0 (obter aprovação na unidade curricular), ver equações (3) e (8). O modelo logístico mostrou uma precisão global (a proporção do número total de previsões corretas) de 85% ao usar como entrada os registos correspondentes às amostras de validação (ver Tabela 5).

Ao contrário do que pensávamos ser expectável antes do desenvolvimento desta investigação, descobrimos que ficar retido (não obter aprovação) em unidades curriculares mais exigentes, como física (dificuldade 4 numa escala 1-4, ver Tabela 1), folhas de cálculo em engenharia, programação VB e estatísticas (dificuldade 3 numa escala 1-4, ver Tabela 1) não tem influência significativa no abandono escolar. Surpreendentemente, não obter aprovação a unidades curriculares menos exigentes como ciência dos materiais (dificuldade 2 numa escala 1-4, ver Tabela 1) e eletricidade (dificuldade 2 numa escala 1-4, ver Tabela 1) tem uma influência significativa no abandono escolar. O motivo dessa ambiguidade não é claramente determinável pela análise dos resultados desta investigação, embora nos pareça razoável formular a hipótese de que, quando os estudantes do primeiro

ano recolhem informações sobre o curso de engenharia industrial, antes de sua inscrição, identificam várias unidades curriculares do curso com um alto nível de exigência. Este conhecimento prévio faz com que os calouros estejam intrinsecamente preparados para não obter aprovação em algumas dessas unidades curriculares. Pelo contrário, não obter aprovação em unidades curriculares classificadas como sendo de baixo nível de exigência causa desmotivação e, em alguns casos, leva ao abando escolar.

Os resultados desta investigação serão utilizados pelo ISVOUGA para ajustes dos programas de ciência dos materiais e eletricidade. Além disso, estas unidades curriculares poderão ser reposicionadas dentro dos seis semestres do curso de engenharia industrial, de tal forma que possam melhorar as taxas de sucesso, mantendo os alunos motivados para completar o curso. Além disso, deverá ser experimentada a orientação por pares, com estudantes de sucesso a influenciar os hábitos de estudo dos calouros, para aumentar as taxas de aprovação e diminuir as taxas de abandono geral.

O mesmo tipo de análise utilizada nesta investigação sobre o curso de engenharia industrial da ISVOUGA pode ser útil para qualquer outra instituição, a fim de identificar as unidades curriculares que não estão de acordo com as expectativas dos estudantes em termos de nível de esforço para conclui-las com sucesso. Os resultados desta investigação parecem indicar que não ser bem-sucedido nas unidades curriculares menos exigentes de um curso é mais desmotivador para os alunos do que não ser bem-sucedido em unidades curriculares muito exigentes. Em algumas situações, será possível ajustar o programa das unidades curriculares, sem comprometer o objetivo geral do curso, para que possam ficar mais próximas das expectativas dos estudantes. Quando esses ajustes de programa não são uma possibilidade efetiva, essas unidades curriculares devem, se possível, ser reposicionadas ao longo do curso, de tal forma que, quando os alunos as frequentam, estejam melhor preparados para o seu nível intrínseco de dificuldade e seja menor o esforço necessário para poderem ser bem-sucedidos.

Referências

ARAQUE, Francisco; ROLDÁN, Concepcion; SALGUERO, Alberto. Factors influencing university drop out rates. **Computers and Education**, Amsterdã, v. 53, n. 3, p. 563-574, 2009.

ARWU. **Academic Ranking of World Universities 2016.** Disponível em: <http://www.shanghairanking.com/ARWU2016.html>. Acesso em: 13 out. 2016.

ASTIN, Alexander. Student involvement: a developmental theory for higher education. **Journal of College Student Development**, Baltimore, v. 40, n. 5, p. 518-529, 1984.

ASTIN, Alexander. **What matters in college:** four critical years revisited. San Francisco: Jossey-Bass, 1993.

BOLOGNA WORKING GROUP. **A framework for qualifications of the European higher education area.** Copenhagen: Ministry of Science, Technology and Innovation, 2005.

COVER, Thomas; HEART, P. Nearest neighbor pattern classification. **Information Theory**, Piscataway, v. 13, n. 1, p. 21-27, 1967. DOI: 10.1109/TIT.1967.1053964.

HAND, David J. Principles of data mining. **Drug Safety**, Berlim, v. 30, n. 7, p. 621-622, 2007.

HOVDHAUGEN, Elisabeth. Do structured study programmes lead to lower rates of dropout and student transfer from university? **Irish Educational Studies**, London, v. 30, n. 2, p. 237-251, 2011. Disponível em: <https://doi.org/10.1080/03323315.2011.569143>. Acesso em: 13 out. 2106.

INE. Instituto Nacional de Estatística. **Censos 2011 norte**. v. 60. [S.l.] INE, 2011.

INE. Instituto Nacional de Estatística. **Portal do Instituto Nacional de Estatística.** [S.l.] INE, 2016. Disponível em: <https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0003920&contexto=bd&selTab=tab2>. Acesso em: 30 out. 2016.

JORDAN, Will J.; LARA, Julia; McPARTLAND, James M. **Exploring the complexity of early dropout causal structures.** Baltimore: Center for Research on Effective Schooling for Disadvantaged Students, 1994.

KUH, George D. What we're learning about student engagement from NSSE: benchmarks for effective educational practices. Change. **The Magazine of Higher Learning**, London, v. 35, n. 2, p. 24-32, 2003. DOI: https://doi.org/10.1080/00091380309604090.

LITALIEN, David; GUAY, Frédéric. Dropout intentions in PhD studies: a comprehensive model based on interpersonal relationships and motivational resources. **Contemporary Educational Psychology**, Amsterdã, v. 41, p. 218-231, Apr. 2015. DOI: https://doi.org/10.1016/j.cedpsych.2015.03.004.

LONG, J. Scott. Regression models for categorical and limited dependent variables. In: PAPER SERIES on quantitative applications in the social sciences. Thousand Oaks: Sage, 1997. p. 52-54. (Series n. 07, 7).

MALM, Joakim; BRYNGFORS, Leif; MÖRNER, Lisse-Lotte. Supplemental instruction for improving first year results in engineering studies. **Studies in Higher Education**, London, v. 37, n. 6, p. 655-666, 2012. DOI: https://doi.org/10.1080/03075079.2010.535610.

MALM, Joakim; BRYNGFORS, Leif; MÖRNER, Lise-Lotte. The potential of supplemental instruction in engineering education - helping new students to adjust to and succeed in university studies. **European Journal of Engineering Education**, London, v. 40, n. 4, p. 347-365, 2015. DOI: https://doi.org/10.1080/03043797.2014.967179.

MARTÍNEZ-LÓPEZ, Zelita et al. Apoyo social en universitarios españoles de primer año: propiedades psicométricas del social support questionnaire-short form y el social provisions scale. **Revista Latinoamericana de Psicologia**, Barcelona, v. 46, n. 2, p. 102-110, 2014. DOI: https://doi.org/10.1016/S0120-0534(14)70013-5.

MENARD, Scott. Applied logistic regression analysis. In: PAPER SERIES on quantitative applications in the social sciences. 2. ed. Thousand Oaks: Sage, 2001. p. 75-79. (Series n. 07-106).

MÍNGUEZ, Almudena Moreno; SAN JULIÁN, Elena Rordríguez. **Informe juventud en España 2012.** Madrid: [s.n., 2012].

MORALES, Erik E.; AMBROSE-ROMAN, Sarah; PEREZ-MALDONADO, Rosa. Transmitting success: comprehensive peer mentoring for At-Risk students in developmental math. **Innovative Higher Education**, Madrid, v. 41, n. 2, p. 121-135, 2016.

NEILD, Ruth Curran; BALFANZ, Robert; HERZOG, Liza. An early warning system. **Educational Leadership**, Alexandria, v. 65, n. 2, p. 28-33, 2007.

OECD. Portugal. In: EDUCATION at a glance 2017: OECD indicators. Paris: OECD, 2017. p. 42-50.

PÁRAMO FERNÁNDEZ, Maria Fernanda et al. Predictors of students' adjustment during transition to university in Spain. **Psicothema**, Bethesda, v. 29, n. 1, p. 67-72, fev. 2017. DOI: 10.7334/psicothema2016.40.

PASCARELLA, Ernest T.; TERENZINI, Patrick T. Predicting freshman persistence voluntary dropout decisions from a theoretical model. **The Journal of Higher Education**, New York, v. 51, n. 1, p. 60-75, 1980. DOI: 10.2307/1981125.

PAURA, Liga; ARHIPOVA, Irina. Cause analysis of students' dropout rate in higher education study program. In: WORLD CONFERENCE ON BUSINESS, ECONOMICS AND MANAGEMENT, 2., 2014, Amsterdã. **Anais…** v. 109. Amsterdã: [s. n.], 2014. p. 1282-1286. DOI: https://doi.org/10.1016/j.sbspro.2013.12.625.

PIERRAKEAS, Christos et al. A comparative study of dropout rates and causes for two different distance education courses. **International Review of Research in Open and Distance Learning**, Athabasca, v. 5, n. 2, p. 1-14, 2004.

REFAEILZADEH, Payam; TANG, Lei; LIU, Huan. Cross-validation. In: ENCYCLOPEDIA of database systems. [S. l.]: Springer, 2009. p. 532-538.

RUSSEL, Stuart; NORVIG, Peter. Artificial intelligence: a modern approach. New Jersey: Prentice Hall, 1995.
SABZEVARI, Hassan; SOLEYMANI, Mehdi; NOORBAKHSH, Eman. A comparison between statistical and data mining methods for credit scoring in case of limited available data. In: CRC CREDIT SCORING CONFERENCE, 3., 2007, Edinburgh. **Proceedings of the…** Edinburgh: [s. n.], 2007. p. 1-8.

STRATTON, Leslie Stundt; O'TOOLE, Dennis M.; WETZEL, James N. Are the factors affecting dropout behavior related to initial enrollment intensity for college undergraduates? **Research in Higher Education**, Berlim, v. 48, n. 4, p. 453-485, Feb. 2007.

TINTO, Vincent. Limits of theory and practice in student attrition. **The Journal of Higher Education**, New York, v. 53, n. 6, p. 687-700, 1982. DOI: 10.2307/1981525.

WILSON, D. Randall; MARTINEZ, Tony R. Improved heterogeneous distance functions. **Journal of Artificial Intelligence Research**, v. 6, p. 1-34, 1997.

Recebido em:02.06.2017
Revisões em: 25.10.2017
Aprovado em: 21.01.2018

António Carlos Corte-Real de Sousa é professor adjunto no Instituto Superior de Entre Douro e Vouga (ISVOUGA), professor auxiliar convidado no Departamento de Engenharia e Gestão Industrial da Faculdade de Engenharia da Universidade do Porto (FEUP) e membro integrado do Centro de Engenharia e Gestão Industrial (CEGI) do INESC TEC.

Carlos Alberto Bragança de Oliveira é professor auxiliar no Departamento de Engenharia e Gestão Industrial da Faculdade de Engenharia da Universidade do Porto (FEUP), membro do Conselho de Departamento de Engenharia e Gestão Industrial e membro integrado do Centro de Engenharia e Gestão Industrial (CEGI) do INESC TEC.

José Luís Cabral Moura Borges é professor associado no Departamento de Engenharia e Gestão Industrial da Faculdade de Engenharia da Universidade do Porto (FEUP), membro do Conselho de Departamento de Engenharia e Gestão Industrial, membro integrado do Centro de Engenharia e Gestão Industrial (CEGI) do INESC TEC e colaborador do Instituto de Interface: Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial.