

PERCEPÇÕES E PRÁTICAS EDUCATIVAS NO DESEMPENHO ACADÊMICO: UMA ABORDAGEM MACHINE LEARNING

Fernanda Silva dos SantosEmail1
fernandada8@hotmail.com
Universidade Católica de Brasília

Carlos Enrique Carrasco Gutierrez
carlosenrique@ucb.br
Universidade Católica de Brasília

Thiago Christiano Silva
thiago.christiano@ucb.br
Universidade Católica de Brasília

RESUMO

Este trabalho tem como objetivo estudar as percepções dos estudantes em relação à aplicação de práticas educativas não ortodoxas sobre o desempenho acadêmico no Ensino Superior no Brasil. Usamos a base de dados do ENADE no triênio entre 2016 e 2018 e modelos econométricos que incorporam efeitos fixos determinados pelas ferramentas de aprendizado de máquina (Machine Learning). Os resultados mostram que as práticas de ensino do uso de tecnologias da informação e comunicação (TICs) como projetor multimídia, laboratório de informática, ambiente virtual de aprendizagem, tiveram um efeito positivo no desempenho do estudante. Da mesma forma, as práticas de incentivar a participação dos alunos em eventos internos e/ou externos à instituição também contribuiu para o bom desempenho.

Palavras-Chave: desempenho; estudante; práticas educativas; ENADE.

1. INTRODUÇÃO

As Instituições de Ensino Superior (IES) têm sido motivadas a refletir acerca das mudanças do processo da educação, reconhecer seu papel social e enfrentar seus desafios, entre os quais o de romper com as estruturas cristalizadas e os modelos de ensino tradicional. Devido aos avanços tecnológicos, mudanças na cultura, demandas da qualificação no mercado de trabalho, entre outros fatores, algumas IES vêm reestruturando seus métodos de ensino-aprendizagem trazendo inovações em suas práticas educativas a fim de formar o aluno segundo as exigências do mercado laboral.

As práticas educativas são experiências que são passadas através de técnicas, ou de compartilhamento, entre os docentes mais experientes para os mais novos que podem favorecer a aprendizagem de forma mais eficiente. Uma prática educativa é um acontecimento social e universal indispensável à existência de todas as sociedades. Segundo Souza et al. (2014), um dos grandes desafios para a Instituição de Ensino Superior é estimular, capacitar o corpo docente e proporcionar infraestrutura para o emprego dos variados métodos de ensino-aprendizagem.

Chickering e Gamson (1987, 1991) apresentaram a aplicação dos sete princípios de boas práticas no Ensino de Graduação. Esses princípios, que foram baseados em décadas de pesquisas, investigam a experiência educacional dos cursos de graduação e foram desenvolvidos com base

na preocupação dos professores com o declínio do desempenho acadêmico, mau ensino e o desinteresse dos alunos (Hutchins, 2003). Essas práticas foram usadas em diferentes estudos (Batts, Colaric e McFaden 2006) mostrando ser relevantes para a avaliação dos aspectos tecnológicos do processo educacional.

Os setes princípios de boas práticas são classificados como: (a) incentiva o contato entre professores e alunos, (b) incentiva a reciprocidade e a cooperação entre os alunos, (c) incentiva a aprendizagem ativa, (d) dá feedback imediato, (e) enfatiza o tempo na tarefa, (f) comunica altas expectativas e (g) respeita diversos talentos e formas de aprendizagem. Esses princípios têm sido comumente usados para avaliar e definir o ensino eficaz em ambientes tradicionais de sala de aula como apresentados em Batts, Colaric e McFaden (2006) e Chickering e Ehrmann (1996).

Embora esses princípios sejam apresentados como estratégias para a melhoria do Ensino no IES, muitos deles ainda não foram considerados e nem implementados no Brasil. Pelo contrário, há um intenso debate em relação aos métodos de ensino empregados em sala de aula, defendendo pela manutenção do ensino tradicional ortodoxo em detrimento do uso de novas práticas educacionais. No entanto, pouco se sabe sobre as percepções dos alunos sobre como essas ferramentas de ensino exclusivas influenciam sua experiência geral.

Este estudo tem como objetivo estudar a percepção dos estudantes em relação aos efeitos dessas práticas educativas no desempenho dos estudantes do IES no Brasil. Tendo como referência os sete princípios de boas práticas no Ensino de Graduação levantamos perguntas relacionadas a algumas práticas de ensino a fim de verificar seus efeitos no desempenho educacional. Nosso projeto de pesquisa explorará dez questões fundamentais respondidas pelos alunos, são elas:

- O curso propiciou experiências de aprendizagem inovadoras?
- O curso favoreceu a articulação do conhecimento teórico com atividades práticas?
- Os professores apresentaram disponibilidade para atender os estudantes fora do horário das aulas?
- O curso ofereceu condições para os estudantes participarem de eventos internos e/ou externos à instituição?
- Foram oferecidas oportunidades para os estudantes participarem de projetos de iniciação científica e de atividades que estimularam a investigação acadêmica?
- Foram oferecidas oportunidades para os estudantes participarem de programas, projetos ou atividades de extensão universitária?
- A instituição promoveu atividades de cultura, de lazer e de interação social?
- O curso disponibilizou monitores ou tutores para auxiliar os estudantes?
- Os professores utilizaram tecnologias da informação e comunicação (TICs) como estratégia de ensino (projeto multimídia, laboratório de informática, ambiente virtual de aprendizagem)?
- No curso você teve oportunidade de aprender a trabalhar em equipe?

Buscamos trazer o conhecimento empírico que demonstre se há ou não um efeito dessas práticas no desempenho do estudante. Este estudo utiliza uma abordagem temática indutiva baseada no autorrelato do aluno sobre percepções e preocupações individuais. Estabelece como parâmetro a prática educativa dentro das unidades de ensino superior. Para tal, utilizamos os microdados do Exame Nacional de Desempenho dos Estudantes (ENADE) para os anos 2016, 2017 e 2018 contemplando a totalidade de cursos do IES apresentados na Tabela 7 (anexo). Como procedimento metodológico utilizamos os modelos de regressão para estimar os efeitos dos

atributos no desempenho educacional e usamos a ferramenta de aprendizado de máquina (Machine Learning) para a seleção das variáveis.

A contribuição deste trabalho é múltipla. Primeiro, trazemos evidências empíricas dessas práticas no debate educacional sobre as melhores práticas no IES. Segundo nossa análise não se limita apenas ao uso dos modelos de regressão. A abordagem traz uma intercepção entre duas áreas; Machine Learning em Ciências de Computação e Econometria em Economia. Com a utilização do Machine Learning realizamos seleção automática de atributos que possam afetar a variável de interesse e isso diminui a subjetividade do analista na escolha das variáveis que possam ser mais ou menos relevante no estudo. O Machine Learning não substitui o analista, mas complementa a análise do analista, ou seja, a metodologia aponta as variáveis que o analista deixou passar despercebido, evitando qualquer escolha ad hoc das variáveis. O uso do Machine Learning é um diferencial no procedimento metodológico empregado neste estudo em relação a outros estudos parecidos ao considerar um conjunto maior de informação na análise gerando resultados mais robustos trazendo contribuições importantes para a literatura.

Finalmente, os resultados empíricos demonstram que as práticas que contribuem positivamente no desempenho do estudante são aquelas relacionadas com o uso de tecnologias da informação e comunicação (TICs) como estratégia de ensino (projektor multimídia, laboratório de informática, ambiente virtual de aprendizagem) e aquelas relacionadas as práticas de incentivar a participação dos alunos em eventos internos e/ou externos à instituição.

2. REVISÃO DE LITERATURA

As práticas utilizadas têm que ser renovadas constantemente, levando em consideram as mudanças que ocorrem no mundo. Gaižiūnienė (2018) ressalta que a prática educativa em Lituânia revela o processo de inovação e o processo decisório de inovação através de interação entre professor e aluno no ensino superior. Baseia-se no conceito de um moderno currículo, que estabelece o que deve ser constantemente renovado para se aplicar a um específico contexto em mudança.

Há vários trabalhos acadêmicos em diversos campos, como ciências contábeis, medicina, etc. Cornachione et al. (2010) que realizaram uma pesquisa com uma amostra de 826 alunos do curso de Ciências Contábeis. Malerva e Escorza (2018) mostraram uma análise feita com estudantes do 2º e 3º ano do curso de medicina. Alencar e Fleith (2004) avaliaram o instrumento denominado Inventário de Práticas Docentes, que foi montado pelas autoras com base em 37 itens com três versões focando perspectivas diferentes: uma respondida pelo professor, a segunda a ser respondida pelo professor na perspectiva de seus alunos e a terceira sendo respondida pelos estudantes avaliando o referido professor. McCabe e Meuter (2011) utilizaram os Sete Princípios de Boas Práticas no Ensino de Graduação para avaliar as percepções dos alunos sobre o ambiente de aprendizagem. Os autores descobriram que os alunos não consideram as ferramentas altamente eficazes para aprimorar a experiência de aprendizado, entretanto eles gostam de usar nos cursos ferramentas de gerenciamento.

Em Washington, D.C. a pesquisa realizada pelos Bonwell e Eison (1991) apontam que o processo de aprendizagem precisa ser realizado, colocando o aluno como elemento ativo, de forma engajada, compreendendo que todo ciclo de atividades será mais bem aproveitado quando esse indivíduo está envolvido na tarefa em vez de simplesmente absorver passivamente os conhecimentos. Percebendo que a tecnologia se tornou algo fundamental na sociedade, Huffman (2011) realizou um estudo para analisar os benefícios trazidos com a utilização da tecnologia no meio acadêmico. No entanto, viram que o grande desafio que os professores tinham é como

fazer o aluno utilizar a tecnologia de uma forma que auxilia a aprendizagem em sala de aula. Os resultados mostraram que o conhecimento na utilização da tecnologia estava associado a frequência do uso do computador e que o sucesso acadêmico estava relacionado com o uso da tecnologia voltado para a aprendizagem especificamente a tecnologia considerada resultados.

Cornachione et al. (2010) realizaram uma pesquisa observando atributos comuns na literatura e no desempenho acadêmico de alunos da graduação em ciências contábeis de quatro universidades brasileiras. A base utilizada, com perguntas realizadas por eles, apresentava um conjunto de respostas de 826 alunos. Constataram que 68% dos alunos consideraram seu desempenho acadêmico superior e atribuíram isso ao seu próprio esforço, enquanto menos de 10% o relacionaram a causas externas. Daqueles que avaliaram seu desempenho acadêmico como inferior quase 24% relacionaram esse fracasso a causas externas (família, provas, colegas e professores). Concluíram que o desempenho acadêmico superior é mais atribuído a causas internas que o desempenho acadêmico inferior, com uma interessante característica entre os gêneros. Por outro lado, tivemos Santana e Araújo (2010) que fizeram um estudo com os docentes dos cursos de Ciências Contábeis e constataram que a falta de especialização é uma das maiores deficiências na educação contábil no Brasil. Levantaram aspectos do perfil dos professores dos cursos de Ciências Contábeis segundo o modelo proposto por Freire (1996). Esse perfil foi obtido a partir da percepção dos coordenadores dos cursos de Ciências Contábeis das universidades federais que participaram do Exame Nacional de Desempenho dos Estudantes (ENADE) 2006. O estudo analisou o reflexo do perfil desses professores no conceito obtido no ENADE 2006 pelas instituições nas quais ensinam. O resultado estatístico encontrado por meio da relação bivariada permitiu verificar que não há relação entre o conceito do ENADE 2006 e o perfil dos professores das universidades pesquisadas. Portanto, a análise constitui em buscar o desempenho dos estudantes no ENADE observados apenas os dados relacionados aos cursos de Ciências Contábeis das universidades federais. Analisaremos o resultado estatístico da relação entre práticas educativas e o desempenho do aluno, levando em consideração todas as oportunidades e incrementos que a universidade proporciona para os discentes e docentes.

A ideia de Alencar e Fleith (2004) avaliou o instrumento denominado Inventário de Práticas Docentes, que foi montado pelas autoras com base em 37 itens com três versões focando perspectivas diferentes: uma respondida pelo professor, a segunda a ser respondida pelo professor na perspectiva de seus alunos e a terceira sendo respondida pelos estudantes avaliando o referido professor. Este instrumento foi aplicado em 1068 estudantes de universidades públicas e privadas brasileiras. Foi efetuada uma análise fatorial e 4 fatores foram gerados: Incentivo a Novas Ideias; Clima para Expressão de Ideias; Avaliação e Metodologia de Ensino; e Interesse pela Aprendizagem do Aluno. Os coeficientes obtidos para os fatores foram superiores a 0,72. Os resultados indicam que o inventário discrimina distintas dimensões do comportamento docente que são relevantes para o desenvolvimento da criatividade, constituindo-se em um instrumento útil para fins de pesquisa e diagnóstico de práticas docentes. Logo, esse trabalho reuniu 37 itens para gerar 4 fatores: incentivo a novas ideias; clima para expressão de ideias; avaliação e metodologia de ensino; e interesse pela aprendizagem do aluno.

No México, o estudo realizado por Malerva e Escorza (2018) mostraram uma análise feita com estudantes do 2º e 3º ano do curso de medicina. No questionário realizado por eles, levantaram dados de 135 alunos, usando como fatores: idade, sexo, nível de inteligência etc., aqueles relacionados ao seu ambiente familiar ou social e aqueles relacionadas com a escola e suas características. Mostrando um resultado quando separados por faixas, os resultados adquiridos indicam que a variável que melhor explica o alto desempenho acadêmico é a gestão de recursos

de informação. No caso da faixa intermediária são: a estratégia de preparação e o ano do curso de graduação. Não foram encontradas variáveis para explicar aqueles com baixo desempenho.

3. METODOLOGIA

3.1 Dados

Os dados são resultados do ENADE e integram o Sistema Nacional de Avaliação da Educação Superior, que é utilizado para avaliar o rendimento dos alunos do último ano dos cursos de graduação em relação aos conteúdos programáticos previstos nas diretrizes curriculares dos cursos; o desenvolvimento de competências e habilidades necessárias ao aprofundamento da formação geral e profissional; e o nível de atualização dos estudantes com relação à realidade brasileira e mundial. Os resultados do ENADE, aliados às respostas do Questionário do Estudante, são insumos para o cálculo dos Indicadores de Qualidade da Educação Superior usados para medir o conceito ENADE nas instituições de ensino superior, as notas são dadas entre 1 a 5.

Utilizamos a base de dados do Exame Nacional de Desempenho do Estudante (ENADE) para os anos de 2016 a 2018. Essas variáveis contemplam um conjunto de características com objetivo de comportar o perfil educacional das IES sobre o desenvolvimento do aluno, apurando o impacto de medidas educacionais utilizada pela instituição para contribuir no desempenho do aluno. A cada ano o ENADE se dedica a um Ciclo Avaliativo Trienal. Entre os anos de 2016 a 2018, foram avaliados os estudantes das áreas apresentadas na tabela 4. Consideramos no estudo apenas os cursos para modalidade presencial.

Incorporamos no estudo 56 variáveis independentes de um total de 64 disponíveis no ENADE para explicar o desempenho educacional. Usamos como indicador de desempenho a nota geral do aluno. A amostra tem 853.981 respostas dos alunos que realizaram a prova. A descrição completa das variáveis é apresentada na Tabela 8 no anexo.

A Tabela 1 apresenta estatísticas descritivas resumidas das variáveis que utilizamos. Em média, observamos uma variação na nota geral de 48,725%. Manusearemos essas mudanças da variável dependente para apurar se a relação com as variáveis independentes relacionadas a práticas educativas.

Tabela 1: Estatística descritiva das variáveis quantitativas.

Estatística Descritiva	N	Média	Desvio Padrão	Min	Q1	Q3	Max
idade	853,981	27,519	7,182	17	23	30	89
nota_geral	853,981	48,782	17,841	0	36,50	61,70	99,4
qt_pessoas	847,026	2,461	1,655	0	1	3	8
renda_fam	847,026	4.702,98	5.424,8	702,75	2.108,25	4.919,25	28.110
quant_livro_leu	847,025	2,923	2,981	0	1	3	10
horas_estud	847,025	5,337	4,482	0	2	5,500	15

Fonte: autores.

Nota: A descrição das variáveis, nota_geral, qt_pessoas, renda_fam, quant_livro_leu, e horas_estud está na Tabela 8 (anexo).

3.2 Modelos de Regressão Regularizados

Para esta pesquisa utilizamos o modelo de regressão regularizado (ou modelos de penalização) conhecidos também como regressão *Elastic Net* (Rede Elástica). Tal regressão combina de forma ideal a regularização L_2 -norm (Ridge) e L_1 -norm (Lasso). Portanto, somos capazes de evitar qualquer *overfitting*¹ (sobreajuste) no nosso modelo empírico. Além disso, usamos uma combinação convexa de L_1 -norm, que tende a encolher a maioria das regressões não relevantes para zero e manter o não zero mais importante, e L_2 -norm, que tende a produzir não zero e coeficientes aproximados para todos os regressores similares. Usando ambos os esquemas de regularização, podemos desfrutar das características positivas de ambos os esquemas.

Para selecionar os atributos mais importantes, utilizamos o resíduo do modelo (2) da regressão da nota da prova, como variável dependente, e diferentes variáveis da base ENADE 2016 a 2018, tais como, características educacionais dos professores, alunos e universidade como variáveis independentes, como segue:

$$y_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + error_i \quad (2)$$

no qual x_{it} é um vetor composto por mudanças passadas na base ENADE e características dos estudantes (sexo, renda, bolsa auxílio, tempo de estudo etc.). O termo $error_i$ é o erro padrão. De acordo com o procedimento da rede elástica, selecionamos β que minimiza a seguinte função de perda $L(\beta)$:

$$L(\beta) = \sum_{i=1}^N (y_i - \sum_{j=1}^p \beta_j x_{it}^j)^2 + \lambda \left[(1 - \alpha) \frac{\|\beta\|_2^2}{2} + \alpha \|\beta\|_1^2 \right] \quad (3)$$

Em que $i \in \{1, \dots, N\}$ indexam anualmente e i indexam os estudantes. O termo x_{it}^j indexa o j th regressor da nota i no tempo t . Os operadores $\|\cdot\|_1$ e $\|\cdot\|_2$ indicam L_1 - e L_2 - normas tomadas sobre a entrada vetorial.

A primeira expressão em (3) denota o erro tradicional de ajuste de dados (resíduos), enquanto a segunda representa o termo de regularização. O parâmetro λ modula a importância do termo tradicional e do termo de regularização. O termo α controla a mistura convexa de L_1 e L_2 de regularização. A regularização funciona penalizando os grandes coeficientes β . Portanto, ela reduz os coeficientes estimados e a função de ajuste geral torna-se mais suave na distribuição dos dados.

3.3 Especificação Econométrica

O modelo de econométrico considerado neste estudo leva em conta as variáveis contínuas apresentadas na Tabela 1 e os efeitos fixos a serem selecionados pelo procedimento *Elastic Net* do *Machine Learning*. O modelo de regressão considerado é:

$$Nota_i^j = \beta_0 + \gamma_j X_i^j + \beta_1 renda_fam_i + \beta_2 idade_i + \beta_3 hr_estud_i + \beta_3 qt_livros_i + \beta_3 qt_pess_i + D_i \theta' + \varepsilon_i \quad (1)$$

Em que:

$Nota_i^j$: nota geral na prova do Enade do indivíduo i para a prática educativa j .

¹ Quando um modelo estatístico se ajusta muito bem ao conjunto de dados anteriormente observado, mas se mostra ineficaz para prever novos resultados.

$renda_fam_i$: é a renda total de sua família pelo indivíduo i ;

$idade_i$: idade do aluno;

hr_estud_i : horas de estudo;

qt_livros_i : quantidade de livros lido;

qt_pess_i : quantidade de pessoas que moram com o aluno;

D_i : efeitos fixos do conjunto de variáveis escolhidas pelo *Machine Learning*;

X_i^j : prática educativa j em estudo {curso_experi, curso_teorico, disp_prof, curso_eventos, proj_ini_cientif, prog_extensao, inst_social, curso_moni, prof_estra, curso_equipe}.

γ_j : parâmetro de interesse;

θ : vetor de parâmetros relacionado as variáveis encontradas pelo *Maching learning*.

4. RESULTADOS

4.1. Machine learning

Nesta seção, analisamos o poder preditivo de nossos atributos na explicação das percepções dos alunos às mudanças no desempenho estudantil. Usamos métodos de aprendizado de máquina orientados a dados para identificar as variáveis mais relevantes.

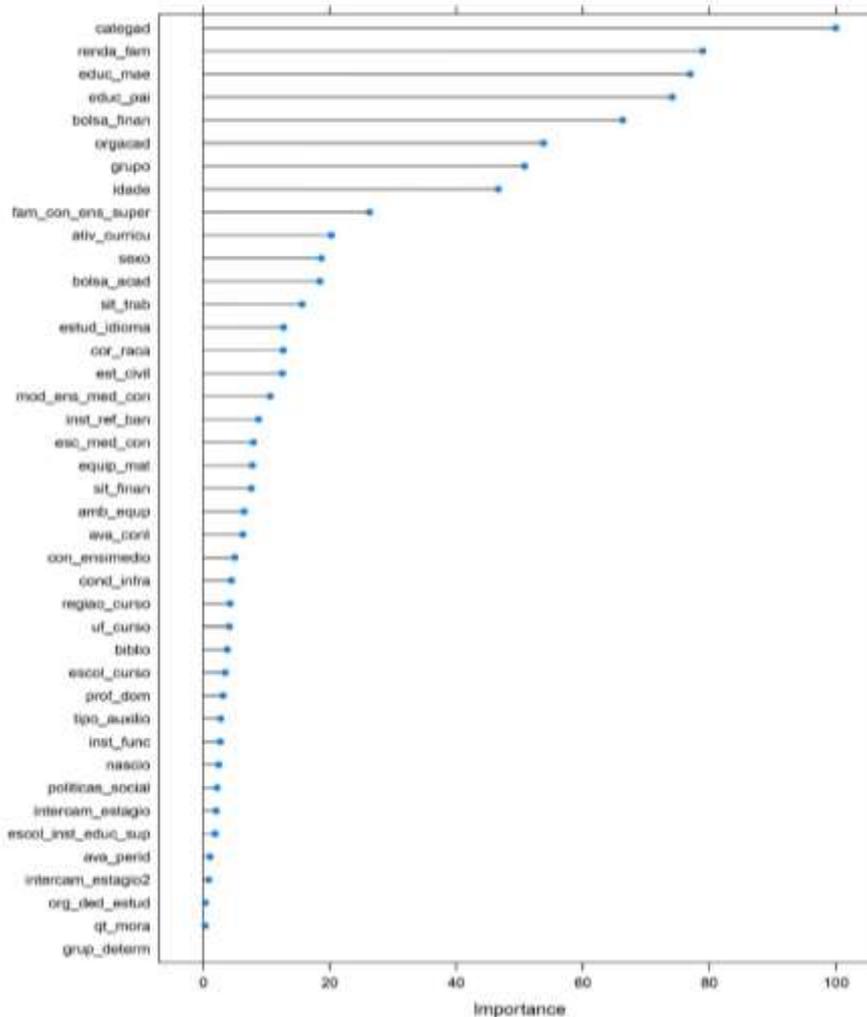
Na regressão *elastic net*, os valores retoma entre 0 e 1. Ajustamos α e λ usando um procedimento de *crossvalidation* alinhado com $k = 10$ dobras e 100 repetições independentes para robustez estatística [29, 49]. Neste procedimento, utilizamos $k - 1 = 9$ dobras para treinamento e a dobra restante para testes. Este procedimento *overfitting* (sobreajuste) em ciclos de k vezes, de modo que cada dobra aparece exatamente uma vez para teste. Esta metodologia permite-nos afinar os parâmetros de regularização, evitando ao mesmo tempo o *overfitting* (sobreajuste) do modelo. Nós otimizamos um espaço de busca sobre a grade $\{0, 0.05, 0.10, \dots, 1\}$ e λ sobre $\{0, 0.1, 0.2, \dots, 5\}$. Como prática padrão, nós pré-processamos todos os regressores aplicando uma padronização Z-score sobre todos os pontos de dados usando valores predeterminados extraídos apenas dos dados de treinamento (de forma a evitar vazamento de dados do conjunto de teste).

A Figura 1 mostra os resultados da regularização *Elastic Net*. Vemos as variáveis mais importantes são a categoria que corresponde ao código da categoria administrativa da IES, $renda_fam$ que corresponde a renda da família, $educ_mãe$ que mostra o nível de escolaridade da mãe, $educ_pai$ mostra o nível de escolaridade da pai e $bolsa_finan$ que corresponde ao tipo de bolsa de estudos ou financiamento do curso recebido para custear todas ou a maior parte das mensalidades, e assim por diante. Além disso, podemos observar a ordem de importância das variáveis na nota geral. É interessante pontuar que as variáveis tradicionais como $sexo$ e cor_raca usados em muitos estudos de regressão ficaram bem abaixo na determinação da variável dependente (nota geral). Isto reforça mais uma vez a análise previa da metodologia *Machine Learning* em nosso estudo.

4.2 Análise econométrico com as variáveis selecionadas

Além das variáveis apresentadas no modelo de regressão (1), consideramos as 15 variáveis mais influentes (de um total de 55 variáveis) escolhidas pelo procedimento *Elastic Net*. As variáveis escolhidas são apresentadas no vetor: $D = [categad, educ_mae, educ_pai, bolsa_finan, orgacad, grupo, fam_con_ens_super, ativ_curricu, bolsa_acad, sit_trab, estud_idioma, est_civil, mod_ens_med_con, inst_ref_ban, esc_med_con]$. Observamos que a variável $categad$, que se refere à categoria administrativa da IES, como pública Federal, Pública Estadual, Pública Municipal, privada com fins lucrativos etc. Essa variável tem um peso

importante na explicação do desempenho estudantil identificada pelo procedimento *Machine Learning*.



Nota: Resultados do *feature selection* usando um procedimento de *Elastic Net* com e regularização. Os coeficientes são normalizados em termos do atributo mais importante “grupo”.

Fonte: autores.

Figura 1: Resultados da seleção de características utilizando um procedimento de rede elástica com regularização L2 e L1.

A variável *educ_mae* e *educ_pai* se referem ao grau de instrução da mãe do pai respectivamente, mostrando-se importante na explicação do desempenho estudantil assim como mostrado em muitos estudos. A variável *bolsa_finan* se refere ao tipo de bolsa de estudos ou financiamento do curso que o aluno recebeu para custear todas ou a maior parte das mensalidades. A variável *orgacad* se refere ao tipo de organização acadêmica da IES se for Faculdade, Centro Federal de Educação Tecnológica, Universidade etc. A variável *grupo* se refere à área de enquadramento do curso e assim por diante. Portanto, vemos que todas essas variáveis são importantes na explicação do desempenho estudantil e que muitas vezes estas variáveis não são consideradas nos estudos científicos. Ao consideramos estas variáveis no modelo de regressão, levamos em conta os efeitos fixos desses atributos nos permitindo obter resultados mais acurados.

As Tabelas 2, 3 e 4 apresentam os resultados do modelo de regressão (1). Na coluna (a) apresentamos os resultados das regressões considerando os efeitos fixos das variáveis selecionados pelo *Machine Learning*. Adicionalmente, apresentamos na coluna (b) os

resultados tradicionais do modelo de regressão apenas com as variáveis escolhidas *ad doc*, renda_fam, idade, qt_pessoas, horas_estudo, sexo, quant_livro_leu e cor. Vemos que os resultados desses dois modelos, em geral, são diferentes. Isto acontece na medida em que um conjunto grande de informação não foi considerado no segundo caso onde apenas o modelo de regressão foi empregado com a escolha subjetiva das variáveis.

Tabela 2: Resultados econométricos

Variável Dependente:	nota_geral					
	(1)		(2)		(3)	
	(a)	(b)	(a)	(b)	(a)	(b)
curso_experi	-0,426*** (0,088)	-1,025*** (0,015)				
curso_teorico			-0,162* (0,087)	-0,840*** (0,014)		
disp_prof					0,110 (0,083)	-0,408*** (0,014)
renda_fam	0,0002*** (0,00002)	0,0005*** (0,00000)	0,0002*** (0,00002)	0,0005*** (0,00000)	0,0002*** (0,00002)	0,0005*** (0,00000)
idade	-0,208*** (0,019)	-0,292*** (0,003)	-0,211*** (0,019)	-0,297*** (0,003)	-0,212*** (0,019)	-0,303*** (0,003)
qt_pessoas	-0,355*** (0,055)	-0,694*** (0,011)	-0,348*** (0,055)	-0,714*** (0,011)	-0,347*** (0,055)	-0,751*** (0,011)
horas_estud	0,213*** (0,022)	0,488*** (0,004)	0,207*** (0,022)	0,482*** (0,004)	0,206*** (0,022)	0,476*** (0,004)
sexoM	1,808*** (0,196)	2,496*** (0,038)	1,830*** (0,197)	2,508*** (0,038)	1,863*** (0,197)	2,657*** (0,038)
quant_livro_leu	0,167*** (0,032)	0,190*** (0,007)	0,162*** (0,032)	0,183*** (0,007)	0,160*** (0,032)	0,164*** (0,007)
Raça [preta]	-1,106*** (0,332)	-0,985*** (0,067)	-1,124*** (0,333)	-1,088*** (0,067)	-1,125*** (0,334)	-1,138*** (0,068)
Raça [parda]	-1,232*** (0,208)	-1,442*** (0,042)	-1,245*** (0,204)	-1,514*** (0,043)	-1,255*** (0,205)	-1,585*** (0,043)
Raça [amarela]	-1,952*** (0,552)	-2,081*** (0,120)	-1,983*** (0,554)	-2,102*** (0,121)	-2,003*** (0,555)	-2,147*** (0,121)
Raça [indígena]	-3,742** (1,670)	-3,708*** (0,316)	-3,816** (1,684)	-3,784*** (0,316)	-3,843** (1,684)	-3,938*** (0,317)
Raça [não declarado]	-0,055 (0,704)	1,522*** (0,131)	0,082 (0,705)	1,559*** (0,132)	0,072 (0,710)	1,688*** (0,132)
Efeito Fixo						
(Variáveis <i>Machine Learning</i>)*	Sim	Não	Sim	Não	Sim	Não
Observações	839,724	843,366	837,619	840,961	835,863	839,058
R ²	0,867	0,083	0,867	0,081	0,867	0,078
R ² Ajustado	0,226	0,082	0,225	0,081	0,225	0,078
Erro Robusto (clustering)	aluno	aluno	aluno	aluno	aluno	aluno

Nota: ***1%; **5% e *10% de nível de significância. O modelo (a) considera como feitos fixos o conjunto de variáveis selecionados pelo procedimento *feature selection* tal como apresentado na Figura 1, sendo eles: *categad, educ_mae, educ_pai, bolsa_finan, orgacad, grupo, fam_con_ens_super, ativ_curricu, bolsa_acad, sit_trab, estud_idioma, est_civil, mod_ens_med_con, inst_ref_ban, esc_med_con*. O modelo (b) considera as variáveis apresentadas na equação (1) sem os efeitos fixos.

Tabela 3: Resultados econométricos

Variável Dependente:	nota_geral					
	(4)		(5)		(6)	
	(a)	(b)	(a)	(b)	(a)	(b)
curso_eventos	0,165** (0,083)	-0,299*** (0,014)				
proj_ini_cientif			-0,253*** (0,074)	-0,264*** (0,013)		
prog_extensao					-0,083 (0,075)	-0,197*** (0,013)
renda_fam	0,0002*** (0,00002)	0,0005*** (0,00000)	0,0002*** (0,00002)	0,0005*** (0,00000)	0,0002*** (0,00002)	0,0005*** (0,00000)
idade	-0,212*** (0,019)	-0,303*** (0,003)	-0,212*** (0,019)	-0,305*** (0,003)	-0,214*** (0,019)	-0,305*** (0,003)
qt_pessoas	-0,346*** (0,055)	-0,759*** (0,011)	-0,350*** (0,056)	-0,762*** (0,011)	-0,351*** (0,056)	-0,768*** (0,011)
horas_estud	0,204*** (0,022)	0,471*** (0,004)	0,207*** (0,022)	0,471*** (0,004)	0,203*** (0,022)	0,470*** (0,004)
sexoM	1,838*** (0,198)	2,644*** (0,038)	1,787*** (0,201)	2,627*** (0,039)	1,810*** (0,200)	2,636*** (0,039)
quant_livro_leu	0,162*** (0,032)	0,158*** (0,007)	0,161*** (0,032)	0,152*** (0,007)	0,161*** (0,032)	0,153*** (0,007)
Raça [preta]	-1,107*** (0,334)	-1,128*** (0,068)	-1,157*** (0,340)	-1,127*** (0,068)	-1,093*** (0,339)	-1,124*** (0,068)
Raça [parda]	-1,271*** (0,205)	-1,582*** (0,043)	-1,278*** (0,208)	-1,591*** (0,043)	-1,317*** (0,208)	-1,602*** (0,043)
Raça [amarela]	-2,002*** (0,556)	-2,165*** (0,121)	-1,993*** (0,564)	-2,126*** (0,122)	-1,990*** (0,565)	-2,139*** (0,122)
Raça [indígena]	-3,821** (1,682)	-3,936*** (0,317)	-3,992** (1,715)	-3,923*** (0,319)	-3,922** (1,712)	-3,964*** (0,318)
Raça [não declarado]	0,090 (0,709)	1,748*** (0,132)	0,056 (0,723)	1,782*** (0,133)	0,111 (0,723)	1,804*** (0,133)
Efeito Fixo						
(Variáveis <i>Machine Learning</i>)*	Sim	Não	Sim	Não	Sim	Não
Observações	835,509	838,772	823,616	826,602	827,407	830,493
R ²	0,867	0,077	0,868	0,077	0,868	0,077
R ² Ajustado	0,225	0,077	0,225	0,077	0,225	0,077
Erro Robusto (clustering)	aluno	aluno	aluno	aluno	aluno	aluno

Nota: ***1%; **5% e *10% de nível de significância. O modelo (a) considera como feitos fixos o conjunto de variáveis selecionados pelo procedimento *feature selection* tal como apresentado na Figura 1, sendo eles: *categad, educ_mae, educ_pai, bolsa_finan, orgacad, grupo, fam_con_ens_super, ativ_curricu, bolsa_acad, sit_trab, estud_idioma, est_civil, mod_ens_med_con, inst_ref_ban, esc_med_con*. O modelo (b) considera as variáveis apresentadas na equação (1) sem os efeitos fixos.

Tabela 4: Resultados econométricos

Variável Dependente:	nota_geral							
	(7)		(8)		(9)		(10)	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
inst_social	-0,004 (0,131)	-0,656*** (0,018)						
curso_moni			-0,241*** (0,075)	-0,488*** (0,012)				
prof_estra					0,226** (0,100)	-0,425*** (0,016)		
curso_equipe							0,016 (0,106)	-0,486*** (0,017)
renda_fam	0,0002*** (0,00003)	0,0004*** (0,00000)	0,0002*** (0,00002)	0,0005*** (0,00000)	0,0002*** (0,00002)	0,0005*** (0,00000)	0,0002*** (0,00002)	0,0005*** (0,00000)
idade	-0,301*** (0,035)	-0,350*** (0,004)	-0,215*** (0,020)	-0,307*** (0,003)	0,208*** (0,019)	-0,304*** (0,003)	-0,210*** (0,019)	-0,304*** (0,003)
qt_pessoas	-0,323*** (0,088)	-0,698*** (0,015)	-0,338*** (0,056)	-0,747*** (0,012)	-0,352*** (0,055)	-0,749*** (0,011)	-0,352*** (0,055)	-0,743*** (0,011)
horas_estud	0,171*** (0,035)	0,431*** (0,006)	0,209*** (0,023)	0,481*** (0,004)	0,206*** (0,022)	0,475*** (0,004)	0,208*** (0,022)	0,479*** (0,004)
sexoM	0,043 (0,348)	2,075*** (0,051)	1,840*** (0,203)	2,631*** (0,039)	1,869*** (0,197)	2,632*** (0,038)	1,851*** (0,197)	2,619*** (0,038)
quant_livro_leu	0,132** (0,052)	0,038*** (0,009)	0,165*** (0,033)	0,157*** (0,007)	0,158*** (0,032)	0,160*** (0,007)	0,161*** (0,032)	0,162*** (0,007)
Raça [preta]	-1,051* (0,563)	-1,321*** (0,090)	-1,087*** (0,342)	-1,167*** (0,068)	-1,067*** (0,333)	-1,117*** (0,068)	-1,121*** (0,332)	-1,091*** (0,067)
Raça [parda]	-1,257*** (0,338)	-1,632*** (0,056)	-1,235*** (0,210)	-1,582*** (0,043)	-1,234*** (0,204)	1,572*** (0,043)	-1,239*** (0,204)	-1,549*** (0,043)
Raça [amarela]	-1,384 (0,891)	-1,758*** (0,158)	-1,972*** (0,568)	-2,126*** (0,122)	-2,002*** (0,555)	-2,115*** (0,121)	-1,955*** (0,552)	-2,134*** (0,121)
Raça [indígena]	-2,889 (2,537)	-4,714*** (0,398)	-3,641** (1,724)	-3,964*** (0,319)	-3,920** (1,687)	-3,896*** (0,317)	-3,779** (1,674)	-3,887*** (0,316)
Raça declarado] [não]	-0,614 (1,115)	1,544*** (0,167)	0,008 (0,729)	1,744*** (0,134)	0,058 (0,706)	1,699*** (0,132)	0,022 (0,703)	1,700*** (0,132)
Efeito Fixo (Variáveis <i>Machine Learning</i>)*	Sim	Não	Sim	Não	Sim	Não	Sim	Não
Observações	463,880	463,880	817,671	820,389	837,666	840,857	840,157	843,735
R ²	0,900	0,078	0,869	0,079	0,867	0,078	0,867	0,078
R ² Ajustado	0,238	0,078	0,225	0,079	0,225	0,078	0,226	0,078
Erro Robusto (clustering)	aluno							

Nota: ***1%; **5% e *10% de nível de significância. O modelo (a) considera como feitos fixos o conjunto de variáveis selecionados pelo procedimento feature selection tal como apresentado na Figura 1, sendo eles: *categad, educ_mae, educ_pai, bolsa_finan, orgacad, grupo, fam_con_ens_super, ativ_curricu, bolsa_acad, sit_trab, estud_idioma, est_civil, mod_ens_med_con, inst_ref_ban, esc_med_con*. O modelo (b) considera as variáveis apresentadas na equação (1) sem os feitos fixos.

Por exemplo, na Tabela 2, o resultado na coluna (a) para a variável *curso_expe* foi -0.426, enquanto na coluna (b) mostra -1.025. Embora em ambos os casos os resultados sejam estatisticamente significativos ao nível de 5%, o valor para o caso (b) resulta ser aproximadamente duas vezes o obtido no caso (a). Podemos ver também que para a variável *disp_prof*, na Tabela 2 no caso (b), achamos evidências dessa variável sob o desempenho acadêmico, porém na análise com os efeitos fixos escolhidos pelo *machine learning* evidencia que não há evidências dessa variável na explicação do desempenho. Esta análise evidencia a importância de considerar no estudo a maior quantidade de informação para ter as estimativas mais robustas.

Consideramos, portanto, os resultados do modelo (a) com efeitos fixos. Na Tabela 2 observamos que a variável *curso_experi* tem um efeito negativo estatisticamente significativo ao nível de 5%. Isso significa que à medida que os alunos concordam com essa prática ela não faz com que a nota do aluno diminua. Em relação a variável *curso_teorico* vemos que ela não impacta ao nível de 5% a nota, embora exista um efeito negativo ao nível de 10%. Com respeito a variável *disp_prof* vemos que não tem um efeito no desempenho ao nível de 5%.

A variável *curso_eventos* apresentados na Tabela 3 tem um efeito positivo ao nível de 5%, a variável *proj_ini_cientif* tem um efeito negativo estatisticamente significante ao nível de 5%. A variável *prog_extensao* não afeta o desempenho.

Na Tabela 4 podemos ver que a variável *inst_social* não tem um efeito no desempenho, a variável *curso_moni* tem um efeito negativo e a variável *prof_extra* um impacto positivo e a variável *curso_equipe* não afeta o desempenho.

Em resumo, observamos que as práticas que contribuem positivamente no desempenho do estudante são aquelas relacionadas com a utilização de tecnologias da informação e comunicação como estratégia de ensino, e as condições do curso em oferecer condições para os estudantes participarem de eventos internos e/ou externos à instituição.

Além disso em todas as estimativas observamos que em todas as variáveis de práticas educativas não tradicionais as variáveis de renda família, idade, quantidade de pessoas que moram com o estudante, sexo, raça são estatisticamente significantes ao nível de 5%. tem uma significância em relação ao desempenho acadêmico do aluno. Observamos que na variável renda família tem um valor positivo, isso significa que quando melhor as condições financeiras da família do estudante, melhor será sua nota. Notamos que quanto mais o estudante dedica para os estudos, isso eleva sua nota. Por outro lado, a cor do aluno influencia negativamente para seu desempenho. Outra revelação importante é o sexo do aluno, no qual mostrou não haver interferência na nota do aluno em relação a sua sexualidade.

A tabela a seguir apresenta o resumo dos resultados. Vemos que apenas duas práticas tiveram um efeito positivo no desempenho: a disponibilidade dos professores em atender os alunos fora do horário das aulas (*prof_extra*); e as condições do curso em oferecer condições para os estudantes participarem de eventos internos e/ou externos à instituição (*curso_eventos*).

Por outro lado, em relação as perguntas se o curso propiciou experiências de aprendizagem inovadoras (*curso_experi*); oferecer oportunidades para os estudantes participarem de projetos de iniciação científica e de atividades que estimularam a investigação acadêmica (*curso_experi*); e se o curso disponibilizou monitores ou tutores para auxiliar os estudantes (*curso_experi*), todas elas tiveram um efeito negativo no desempenho. As outras dimensões não tiveram efeito algum na nota.

Tabela 4: Resumo dos resultados

Descrição	Variável	$\alpha = 5\%$	$\alpha = 10\%$
O curso propiciou experiências de aprendizagem inovadoras.	curso_experi	<0	<0
O curso favoreceu a articulação do conhecimento teórico com atividades práticas.	curso_teorico	-	<0
Os professores apresentaram disponibilidade para atender os estudantes fora do horário das aulas.	disp_prof	-	-
Os professores apresentaram disponibilidade para atender os estudantes fora do horário das aulas.	curso_eventos	>0	>0
Foram oferecidas oportunidades para os estudantes participarem de projetos de iniciação científica e de atividades que estimularam a investigação acadêmica.	proj_ini_cientif	<0	<0
Foram oferecidas oportunidades para os estudantes participarem de programas, projetos ou atividades de extensão universitária.	prog_extensao	-	-
A instituição promoveu atividades de cultura, de lazer e de interação social.	inst_social	-	-
O curso disponibilizou monitores ou tutores para auxiliar os estudantes.	curso_moni	<0	<0
Os professores utilizaram tecnologias da informação e comunicação (TICs) como estratégia de ensino (projektor multimídia, laboratório de informática, ambiente virtual de aprendizagem).	prof_estra	>0	>0
No curso você teve oportunidade de aprender a trabalhar em equipe.	curso_equipe	-	-

5. CONCLUSÕES

Este estudo tem como objetivo estudar a percepção dos estudantes em relação aos efeitos dessas práticas educativas no desempenho dos estudantes do IES no Brasil. Empregamos técnicas de aprendizagem de máquinas juntamente com técnicas econométricas para identificar e estimar as variáveis que afetam no desempenho acadêmico do aluno.

Os resultados deste trabalho evidenciam que as práticas que contribuem positivamente no desempenho do estudante são aquelas relacionadas com o uso de tecnologias da informação e comunicação como estratégia de ensino, projetor multimídia, laboratório de informática, ambiente virtual de aprendizagem. Evidenciamos também que oferecer aos estudantes incentivos para participar de eventos internos e/ou externos à instituição tem um efeito positivo no desempenho acadêmico.

Esses resultados podem trazer uma nova perspectiva para as IES para melhorar as condições de ensino nos cursos, mudando o currículo para preparar o aluno para área de atuação e desenvolvimento profissional.

O uso de métodos de aprendizagem de máquinas, uma área da Inteligência Artificial, nos permite automatizar o processo subjetivo de escolha de quais variáveis são importantes em qualquer análise econométrica. Ao utilizar uma seleção de características, como a rede elástica, estamos aptos a identificar os atributos que melhor qualificam as IES com suas práticas educativas proporcionando assim, uma melhor evidencia dos resultados alcançados.

6. REFERÊNCIAS

- Aithal, P. S. and Kumar, P. M., How Innovations and Best Practices Can Transform Higher Education Institutions: A Case Study of SIMS (February 17, 2015). International Journal of Management (IJM), Volume 6, Issue 2, February (2015), pp. 83-98.
- Alencar, E. M. L. S.; Fleith, D. S. Inventário de Práticas Docentes que Favorecem a Criatividade no Ensino Superior. Psicologia: Reflexão e Crítica, 2004, 17(1), pp.105-110.
- Barbel, N. A. N. As metodologias ativas e a promoção da autonomia de estudantes. Semina: Ciências Sociais e Humanas, Londrina, v. 32, n. 1, p. 25-40, jan./jun. 2011.

- Batts, D., Colaric, S. M., & McFadden, C. (2006). Online courses demonstrate use of seven principles. *International Journal of Instructional Technology & Distance Learning*, 3(12), 15-26.
- Bonwell, C. C. & Eison, J. A. (1991). Active learning: creating excitement in the classroom. ASHE-ERIC higher education reports. ERIC Clearinghouse on Higher Education, The George Washington University, Washington, DC.
- Cornachione Junior, E. B.; Cunha, J. V. A.; de Luca, M. M. M.; Ott, E. O bom é meu, o ruim é seu: perspectivas da teoria da atribuição sobre o desempenho acadêmico de alunos da graduação em Ciências Contábeis. *Rev. contab. finanç.* 2010.
- Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *American Association for Higher Education Bulletin*, 39(7), 3-7.
- Chickering, A. W., & Gamson, Z. F. (1991). Applying the seven principles for good practice in higher education. San Francisco, CA: Jossey-Bass.
- Chickering, A. W., & Ehrmann, S. C. (1996). Implementing the seven principles: Technology as a lever. *American Association for Higher Education Bulletin*, 49(2), 3-6.
- Cyrino, L. G.; Pereira, M. L. T. Trabalhando com estratégias de ensino-aprendizado por descoberta na área da saúde: a problematização e a aprendizagem baseada em problemas. *Cad. Saúde Pública*, Rio de Janeiro, 20(3):780-788. 2004.
- Dean, T.; Lee-Post, A.; Hapke, H. Universal Design for Learning in Teaching Large Lecture Classes *Journal of Marketing Education*. 2016.
- Estévez-Ayres, Iria; Alario-Hoyos, Carlos; Pérez-Sanagustín, Mar; Pardo, Abelardo; Crespo-García, Raquel M.; Leony, Derick; Parada G., Hugo A.; Delgado-Kloos, Carlos. A methodology for improving active learning engineering courses with a large number of students and teachers through feedback gathering and iterative refinement. *ERIC - International Journal of Technology and Design Education*. 2015.
- Gaižiūnienė, Lina. Innovative Study Methods Adaptation: The Power of An Effective Teacher. *The European Journal of Social and Behavioural Sciences EJSBS Volume XXIV*. 2018.
- Guerra, C. J. O., & Teixeira, A. J. C. Os impactos da adoção de metodologias ativas no desempenho dos discentes do curso de ciências contábeis de instituição de ensino superior mineira. *Revista De Educação E Pesquisa Em Contabilidade (REPeC)*, 10(4). 2016.
- Huffman, W. H.; Huffman, A. H. Beyond basic study skills: The use of technology for success in college. Elsevier. 2011.
- Hutchins, H. M. (2003). Instructional immediacy and the seven principles: Strategies for facilitating online courses. *Online Journal of Distance Learning Administration*, 6(3), 1-10.
- Kolb, D. A. (1981). Learning styles and disciplinary differences. In A. Chickering (Ed.), *The modern American college* (pp. 232-255). San Francisco, CA: Jossey-Bass.
- Limberger, J. B. Metodologias ativas de ensino-aprendizagem para educação farmacêutica: um relato de experiência. *Interface*. 2013.
- Malerva, I. S.; Escorza, Y. H. Estrategias de aprendizaje y desempeño académico en estudiantes de Medicina. *Educación Médica*. 2018.
- McCabe, D. B.; Meuter, M. L. A Student View of Technology in the Classroom: Does It Enhance the Seven Principles of Good Practice in Undergraduate Education? *Journal of Marketing Education*. 2011.

Mello, M. B. J. B.; Leme, M. I. S. Motivação de alunos dos cursos superiores de tecnologia. *Psicol. Esc. Educ.* 2016.

Santana, A. L. A.; Araújo, A. M. P. Aspectos do perfil do professor de Ciências Contábeis e seu reflexo no Exame Nacional de Desempenho dos estudantes (ENADE) – um estudo nas universidades federais do Brasil. Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

Silva, T. C.; Tabak, B. M.; Ferreira, I. M. Modeling Investor Behavior Using Machine Learning: Mean-Reversion and Momentum Trading Strategies. Hindawi. 2019.

Souza, C. S.; Iglesias, A.G; Pazin Filho, A. Estratégias inovadoras para métodos de ensino tradicionais – aspectos gerais. *Revista FMRP USP.* 2014.