

Uma Análise preditiva de desempenho dos cursos no ENADE com base no perfil socioeconômico e desempenho no ENEM dos alunos

Hugo Vieira Lucena de Souza¹, Davi Hirafuji Neiva¹, Rafael Pereira Calvalcanti¹, Rodrigo Lins Rodrigues², Alex Sandro Gomes¹, Paulo Jorge Leitão Adeodato¹

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 50.740-560 – Recife – PE – Brazil

²Departamento de Educação – Universidade Federal Rural de Pernambuco (UFRPE)
Caixa Postal 52171-900 – Recife – PE – Brazil

{hvls, dhn, rpc2, asg, pjla}@cin.ufpe.br,
rodrigo.linsrodrigues@ufrpe.br

Abstract. *The concept attributed to an undergraduate degree can influence its development, and also influence the decision of the student entering or not in a particular course. This concept, obtained through an evaluation process conducted by the National Institute of Educational Studies and Research (INEP), encompasses a complex process in which they assess the skills, abilities and social characteristics of students through a test carried out in public institutions and private higher education across the country. Therefore, this paper aims to analyze the performance of a course in ENADE, before performing the same, taking into account the performance of the students obtained in the test of the National Secondary Education Examination (ENEM) and their social and economic characteristics.*

Resumo. *O conceito atribuído para um curso de graduação pode influenciar o desenvolvimento do mesmo, além de também influenciar a decisão do aluno em ingressar ou não em um determinado curso. Este conceito, obtido mediante a um processo de avaliação realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), engloba um processo complexo em que são avaliadas as competências, habilidades e as características sociais de alunos através de uma prova realizada em instituições públicas e privadas de ensino superior de todo o país. Portanto, este trabalho busca analisar o desempenho de um curso no ENADE, antes da realização do mesmo, levando em consideração o desempenho dos alunos obtido na prova do Exame Nacional do Ensino Médio (ENEM) e suas características sociais e econômicas.*

1. Introdução

Todos os cursos de graduação no Brasil precisam ser avaliados para aferir sua qualidade perante o sistema de educação brasileiro. Esta avaliação é conhecida como Exame de Desempenho de Estudantes – ENADE (INEP, 2017a). O ENADE foi criado com o intuito de verificar o rendimento dos alunos concluintes nos cursos de graduação de instituições de ensino superior públicas e privadas do país, quando são observados os

conteúdos programáticos, as competências curriculares, as habilidades adquiridas para a sua formação.

De acordo com o INEP (2017, b) a primeira avaliação do ENADE ocorreu no ano de 2004 e o intervalo de aplicação das provas definido foi de três anos. Ao longo dos anos este processo de avaliação dos cursos sofreu algumas mudanças, a fim de melhorar os índices de qualidade do ensino, dentre elas a obrigatoriedade de realização para todos os concluintes de um determinado curso e uma avaliação socioeconômica em que são avaliados fatores como renda, escolaridade e a região.

Atualmente, o conceito é concedido através da prova, organizada em questões objetivas e subjetivas pelo curso e pelo questionário do estudante em que são avaliados os perfis socioeconômicos obtidos. Entretanto, muitas instituições de Ensino Superior buscam identificar quais aspectos quais características influenciam a criação de um perfil que poderia direcionar se um curso teria um bom desempenho ou não em uma futura avaliação do ENADE (Cruz et al., 2013). Diante desta observação, este artigo tem como objetivo identificar se um determinado curso terá um bom desempenho através da seguinte questão de pesquisa: dado um curso “X”, este curso terá um desempenho maior ou igual a três na avaliação do ENADE, com base no perfil socioeconômico de desempenho na prova do ENEM? Para que esta pergunta de pesquisa pudesse ser respondida utilizou-se a metodologia *Cross Industry Standard Process for Data Mining* (CRISP-DM) com o objetivo de extrair o conhecimento das bases de dados estudadas e construir o perfil dos alunos através destes dados.

Este artigo está organizado da seguinte forma: a seção 2 descreve as regras do ENADE, bem como as características para obtenção do conceito. A seção 3 descreve a etapa de entendimento dos dados. A seção 4 traz a análise descritiva dos dados, enquanto na seção 5 são apresentadas a árvore de decisão e a regressão logística realizadas. Na seção 6 são descritos os resultados dos testes da Curva ROC, Curva KS e do Teste *K-folder* e, por fim, são apresentadas as considerações finais.

2. Conceito ENADE

O Conceito ENADE é um indicador de qualidade que avalia o desempenho dos estudantes a partir dos resultados obtidos no exame (INEP, 2017b). Seu cálculo, no entanto, não é necessariamente realizado por curso, mas por Unidade de Observação. A Unidade de Observação consiste no conjunto de cursos que compõe uma área de enquadramento específica do de uma Instituição de Educação Superior em um determinado município.

As Unidades de Observação que contam com apenas um, ou sem nenhum concluinte participante, não obtêm o conceito, ficando com uma classificação atribuída como Sem Conceito (SC). Outro fator importante é o de que a nota é calculada para cada curso, que é definida por uma instituição de ensino superior (IES), por um município, por uma área de avaliação e, a partir de 2008, passou a considerar em seu processo apenas o desempenho dos alunos concluintes.

A nota do perfil ingressante foi substituída pela nota do aluno no Exame Nacional do Ensino Médio (ENEM). Desta forma, o conceito final do ENADE do curso é a média ponderada da nota padronizada dos concluintes na Formação Geral (FG) e no Componente Específico (CE). A parte referente à Formação Geral contribui com um

percentual de 25% da nota final, enquanto a referente ao Componente Específico contribui com 75%. Calculada a média ponderada, o conceito ENADE é distribuído conforme pode ser observado na Tabela 1.

TABELA 1. Distribuição dos conceitos, médias ponderadas e descrição (INEP, 2017)

Conceito	Notas finais (NC)	Descrição
1	0,0 a 0,94	Indicadores configuram um conceito <i>NÃO EXISTENTE</i> .
2	0,95 a 1,94	Indicadores configuram um conceito <i>INSUFICIENTE</i> .
3	1,95 a 2,94	Indicadores configuram um conceito <i>SUFICIENTE</i> .
4	2,95 a 3,94	Indicadores configuram um conceito <i>MUITO BOM</i> .
5	3,95 a 5	Indicadores configuram um conceito <i>EXCELENTE</i> .

Os conceitos descritos na Tabela 1 são obtidos com o uso do Questionário do Estudante, um dos instrumentos de coleta de informações do ENADE. Este instrumento, de caráter obrigatório, tem por objetivo construir o perfil socioeconômico do estudante e obter uma apreciação quanto ao seu processo formativo. As informações prestadas na pesquisa subsidiam estatisticamente o cálculo de indicadores educacionais do curso. Este questionário aborda aspectos como: bolsa de estudos, situação de trabalho, financiamento do curso, horas de estudo ao curso, etc. Para esta pesquisa foram selecionados dados das seguintes questões: Qual o seu estado civil? Qual a renda total de sua família, incluindo seus rendimentos? Qual alternativa abaixo melhor descreve sua situação financeira (incluindo bolsas)? Qual alternativa abaixo melhor descreve sua situação de trabalho (exceto estágio ou bolsas)? Que tipo de bolsa de estudos ou financiamento do curso você recebeu parte das mensalidades? Ao longo da sua trajetória acadêmica, você recebeu algum tipo de auxílio? No caso de haver mais de uma opção, marcar apenas a bolsa de maior duração. Ao longo da sua trajetória acadêmica, você recebeu algum tipo de bolsa? No caso de haver mais de uma opção, marcar apenas a de maior duração. Durante o curso de graduação, você participou de programas e ou atividades curriculares no exterior? Seu ingresso no curso de graduação se deu por meio de políticas de ação afirmativa ou inclusão social? Em que tipo de escola você cursou o ensino médio? Qual modalidade de ensino médio você concluiu? Quantas horas por semana, aproximadamente, você dedicou aos estudos, excetuando as horas de aula? Você teve oportunidade de aprendizado de idioma estrangeiro na Instituição?

3. Etapa de entendimento dos dados

A etapa de entendimento dos dados compreende a localização de uma base de dados e em seguida o seu estudo para a aquisição do conhecimento. Para a realização deste trabalho, foram utilizados os microdados do ENADE do ano de 2014 em que estão contidas informações das instituições de ensino, dos alunos e dos cursos. No mesmo ano foram avaliados quarenta e três cursos das áreas de ciências exatas e da natureza, ciências humanas e ciências sociais e da saúde.

Conforme o INEP (2017a) os cursos que possuem conceitos “1” e “2” não podem abrir novas turmas e correm o risco de serem fechados pois não possuem desempenho suficiente de acordo com as exigências curriculares do INEP. A partir do conceito “3”, os cursos recebem incentivos para que possam funcionar de maneira adequada, tais como a disponibilidade de bolsas de estudos e financiamentos por programas de crédito estudantil. Com isso, no contexto desse trabalho, considerou-se um *curso bom*, aquele que obteve uma avaliação maior ou igual a três, considerando a escala de conceitos estabelecida pelo INEP que varia de um a cinco. Um outro aspecto que deve ser destacado se refere à diferença na granularidade dos atributos. Para isso, este trabalho utilizou a média para os dados numéricos.

4. Análise descritiva dos dados

A análise descritiva dos dados constou no processo de verificação da base e na aplicação dos primeiros testes nas variáveis pré-selecionadas. Neste trabalho foi utilizada a ferramenta *Pentaho* para realizar a especificação do fluxo de trabalho na avaliação do desempenho e da ferramenta R para realizar a extração das informações estatísticas dos dados. A classe-alvo foi escolhida utilizando como critério os grupos com cursos que obtiveram um conceito inferior a três pois os mesmos estão propensos a obterem uma nota aceitável na avaliação seguinte do ENADE. Os cursos que representam a classe-alvo obtiveram um percentual de 37,14% do total dos cursos, enquanto a classe complementar a um percentual de 62,8%.

Na análise descritiva foram consideradas as variáveis que tiveram um nível de preenchimento com cem por cento dos valores, além daquelas que são categóricas e numéricas. Ao longo do processo foram avaliados os ganhos de informação, frequência, desvio-padrão, moda e suas respectivas quantidades para que se tornasse possível realizar a comparação. Para as variáveis categóricas, algo interessante a se destacar foram os dados reportados através da categoria-moda.

As análises relevaram que o curso que obteve uma maior frequência relativa da moda foi o de Licenciatura em Pedagogia, com 12,63%, enquanto o estado da federação que também obteve uma maior frequência relativa da moda foi o de São Paulo com 22,77%. A quantidade de categorias com frequência relativa acima de 5% teve uma variação considerável, quando foram avaliados os valores escalonados. A descrição das variáveis aponta que a Universidade do Estado da Bahia obteve uma maior incidência para esta categoria, baseando-se na frequência relativa da moda que chegou aos 1,94%. Nas variáveis numéricas pôde-se observar alguns aspectos importantes acerca dos dados. Larga parte dos alunos participantes da prova são do período noturno. O nível de preenchimento para esta análise chegou a 99,3%, enquanto para os períodos matutino e vespertino chegaram-se as respectivas taxas de 37% cada um.

5. Árvore de decisão e a Regressão Logística

A árvore de decisão é uma técnica que utiliza a estratégia de analisar um problema complexo para subdividi-lo em problema menores (Neves e Souza, 2009). A representação pode ser realizada através de nós, decompostos em níveis, onde são realizados os testes dos atributos quando são emparelhados aos valores que configuram as regras para garantir a tomada de decisão. A árvore de decisão gerada para a base de dados do ENADE apresenta quatro níveis, ordenada por variáveis de maior importância

de acordo com o percurso da árvore. As variáveis UF, tempo de início de graduação, média de estudo noturno, média de início de graduação e renda do aluno compõem as variáveis da árvore para inferir se um curso terá um bom desempenho ou não na prova do ENADE. A partir disto puderam-se extrair algumas regras:

- Regra 1: *if* UF == "MT", "PE", "PA" ou "BA" e o tempo de início na graduação == 3, 6, 8, 7, 4, ou 9, o curso terá uma nota boa no ENADE
- Regra 2: *if* UF == "MT", "PE", "PA" ou "BA" e o tempo de início na graduação == 5, 11 ou 2, o curso terá uma nota boa no ENADE
- Regra 3: *if* UF == "DF", "SE", "SP", "RN" ou "MS" e o tempo de início na graduação == 3 ou 2 e renda == e, b f ou c o curso terá um nota boa no ENADE
- Regra 4: *if* UF == "PI", "GO", ou "RR", o curso terá uma nota boa no ENADE
- Regra 5: *if* UF == "MG", "SC" ou "RJ" e média de estudo noturno == 0 ou 4 e tempo de início na graduação == 3, 5 ou 2, o curso terá uma nota boa no ENADE.

A árvore de decisão foi construída utilizando a ferramenta *SPSS* utilizando o algoritmo J48. O modelo que gera a árvore consistiu de nove variáveis: UF, Renda, Código de área, Inscritos, Participantes, Média de início graduação, Média noturno, Média vespertino e Média matutino. O primeiro nível foi o da variável alvo (cursos com menos de 3 no ENADE). O segundo nível foi a variável UF, seguida de Média de início.

A representação da árvore de decisão gerada para a base de dados do se inicia a partir do primeiro nó, conhecido como nó zero, que neste caso, é o alvo. Ao todo foram gerados trinta e dois nós onde foram estabelecidas as regras para determinar o valor-P, o valor do qui-quadrado e o valor DF. A Figura 1 contém um exemplo com a descrição do nó-alvo, nó 0, e as representações dos nós 1, 8 e 9.

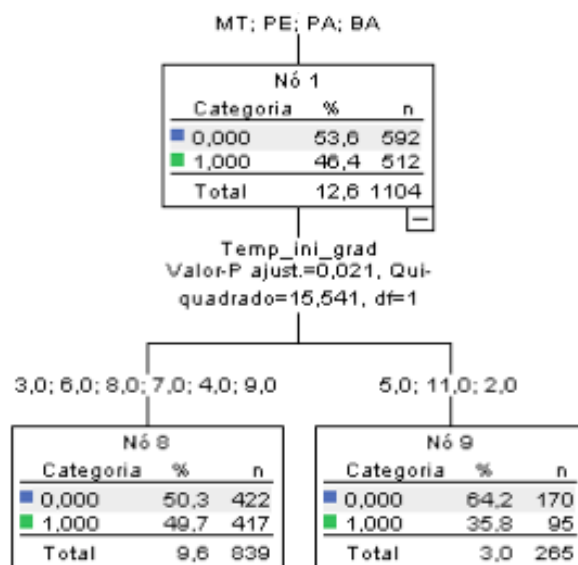


Figura 1. Árvore de decisão – Regra 1 e Regra 2

Na regressão logística o objetivo a ser buscado é o de encontrar o melhor modelo que possibilite descrever a relação existente entre as variáveis de saída e as variáveis independentes categóricas e binárias (Melo, 2017). O processo de regressão realizado para a base do ENADE foi realizado definindo-se em uma primeira etapa a análise de risco relativo por grupos de áreas dos cursos, para que na segunda etapa pudessem ser definidos os pesos da regressão. Na análise dos riscos foram criados quatro grupos de área: grupo 1, grupo 2, grupo 3 e o grupo de outros. Esta categorização da análise replica se um curso tem um índice de obter sucesso ou não, pode ser apresentado no Quadro 1.

QUADRO 1. Risco relativo por área no ENADE

Categorização Cod_area			
GRUPO 1	GRUPO 2	GRUPO 3	OUTROS
4004	21	79	73
5806	72	1502	76
5814	702	1601	701
6208	904	1602	903
6306	905	2001	1401
6307	906	2601	2401
	1402	3002	2501
	1501	3202	3001
	2402	3502	3201
	5710	4006	4004
	5902	5809	4005
		6008	4301
			5401
			5402
			5806
			5814
			6009
			6405

	BAIXO RISCO
	MÉDIO RISCO
	ALTO RISCO

Em cada grupo, como descrito no Quadro 1, estão descritas suas respectivas classificações. No grupo 1 as chances de que o curso terá um risco baixo de obter uma nota ruim, enquanto no grupo 2, grupo 3 e outros, as chances são respectivamente risco médio, alto risco e risco não determinado. Na segunda etapa da regressão logística, utilizou-se uma estratégia de definir as variáveis, as suas subcategorias e a relação com os pesos para os testes. Inicialmente, as variáveis propostas foram UF, código de área, inscritos, participantes, Dif_participantes, tempo de entrada na graduação e as médias de início da graduação nos turnos matutino, noturno e vespertino.

Uma variável de destaque foi a renda os estudantes, utilizada para verificar quais eram os impactos obtidos o processo de regressão estivesse concluído. As subcategorias foram escolhidas para serem representadas de acordo com o tipo de variável utilizada. Neste caso, para a variável UF foram definidas como subcategoria os estados brasileiros que estavam presentes na base, contabilizando um total de 26 estados e o distrito federal. As outras duas variáveis que receberam subcategoria foram o código da área, onde estão distribuídos os códigos, conforme é apresentado no Quadro 1, e a variável renda.

Na variável renda são propostas as subcategorias a, b, c, d, e, f onde são alimentados os valores classificados pelos estudantes para avaliar o resultado do curso mediante o teste prospectado pela prova. Por fim, os pesos são atribuídos considerando um valor padrão de 100 para pudessem ser calculados e contabilizados. Os resultados apontam que vários estados obtiverem um peso semelhante na possibilidade de obter uma boa nota, como pode ser observado no Quadro 2, enquanto outros obtiveram destaque em relação aos seus valores, tornando a escala de regressão variável.

Quadro 2. Variáveis, coeficientes e p-value dos pesos da Regressão Logística

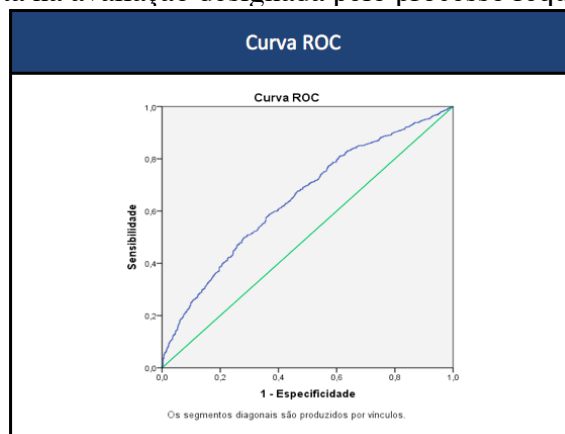
PESOS DA REGRESSÃO LOGÍSTICA		
VARIÁVEL	COEFICIENTE	P-VALOR
UF = RS	1,333	<0,001
UF = PR	0,894	<0,001
UF = RJ	0,842	<0,001
UF = MG	0,751	<0,001
UF = SC	0,701	<0,001
UF = SP	0,311	<0,001
PARTICIPANTES	0,01	<0,001
INSCRITOS	-0,011	<0,001
MED_INI_MATUTINO	-0,032	<0,001
COD_AREA = GRUPO2	-0,583	<0,001
COD_AREA = GRUPO1	-1,032	<0,001

No Quadro 2 percebe-se que há variáveis que possuem valores iguais, como é o caso de UF. Outras variáveis tais como participantes, inscritos, média de início matutino, e os códigos de área dos grupos 1 e grupo 2, obtiveram valores que oscilam mediante ao peso, apresentando uma boa nota. Nas variáveis de código de área o maior valor do coeficiente foi do grupo 2, enquanto dos estados foram do Rio Grande do Sul e do Paraná, potenciais candidatos a obterem uma boa nota. Com a Regressão Logística pôde representar quais aspectos influenciam para que os cursos obtenham uma boa nota.

6. Curva ROC, curva KS e a validação cruzada

Após os testes descritos iniciou-se o teste diagnóstico quantitativo contínuo dos valores de *cutoff point* (Lemos, 2015). Os autores construíram a curva *Receiver Operator Characteristic Curve* – ROC, seguindo um processo que consistiu em identificar as variáveis contínuas e em seguida estabelecer as classes para realizar os testes de sensibilidade e especificidade dos verdadeiros positivos (cursos que realmente obtiveram um bom desempenho no ENADE) e falsos negativos (os que foram classificados como baixo desempenho, porém deveriam ter tipo um bom desempenho).

A curva ROC replica os crescimentos diagonais produzidos por vínculos através da linha diagonal pontilhada, no centro do gráfico, e da linha curvada, direcionada através dos segmentos para validar a evolução, ou não, do índice de verdadeiros positivos. Na Figura verifica-se um crescimento acentuado da curva direcionando que há uma precisão importante a ser considerada quanto a previsão dos cursos avaliados obtiverem uma boa nota na avaliação designada pelo processo requerido.

**Figura 2. Gráfico da Curva ROC da Base do ENADE**

A curva descreve um estado contínuo, em que os valores do eixo x podem tender a gerar um padrão de comportamento, à media em que os cursos são avaliados. Isto implica dizer que há uma tendência que certos cursos obtenham notas boas em diversas situações, enquanto outros podem não adquirir o mesmo sucesso.

Esta tendência é percebida com o teste conhecido como Teste de *Kolmogorov-Smirnov* (Martinez et al., 2017) O teste KS tem como objetivo testar as probabilidades de grupos formados para verificar se os mesmos são ou não distintos. Para a base, realizou-se esta análise considerando como um mecanismo que possibilitasse aferir a performance dos modelos de classificação. A curva KS, representada na Figura 3, é 100 se os escores particionam a população em dois grupos separados na qual um grupo contém todos os positivos e o outro todos os negativos. Caso o KS não consiga dividir a população, seu valor é 0. Na maioria dos casos, o valor KS varia entre 0 e 100 e quanto maior o valor, se torna mais fácil comparar positivos de negativos.

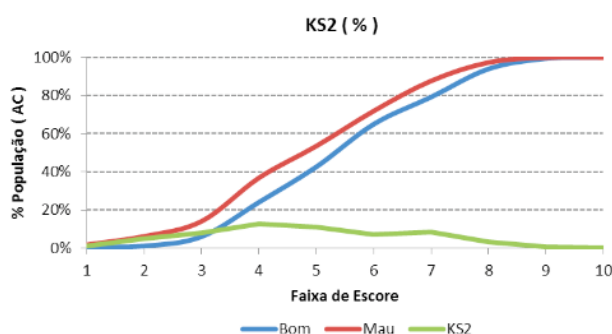


Figura 3. Curva KS – Base de dados ENADE

Com os testes KS obteve-se um KS máximo de 14%, o que equivale a maior distância entre a curva de bons acumulados e a curva de maus acumulados com relação ao score. É importante salientar que o valor KS máximo, possível de ser obtido é 1, que ocorre quando os cursos que obtêm uma boa nota, mediante ao score, são tidos como maus, ou os cursos que tiverem uma nota tida como ruim, ou vice-versa. Cada teste KS replica esta análise, através de valores atribuídos pelo KS *Fold*, ou dobras na curva, sempre que um curso for avaliado. Portanto, as avaliações no ENADE geram variações nas notas, que são restritas por pontos ou valores médios, que representam este crescimento entre os cursos que possuem um FDP mau ou FDP bom.

Todos estes testes são conhecidos como testes de validação dos modelos gerados para prever se as notas serão boas ou não, ou simplesmente teste de validação cruzada. Nos experimentos realizados para validação cruzada os autores realizaram o teste 10folds, onde o valor de cada *fold* pôde ser verificado, como é exibido na Tabela 3. Para cada *fold* definiu-se um valor máximo, além de ao final dos testes, também se calculou um valor de KS máximo médio com um total de 24,2 %.

Tabela 3. Teste de validação cruzada 10 folders

Fold	KS máximo
1	23,2%

2	22,4%
3	23,6%
4	26,6%
5	24,4%
6	22,8%
7	25,8%
8	25,2
9	27,2%
10	20,8%
KS Máximo médio	24,2%

Lendo os valores nas colunas percebe-se que o objetivo buscado foi de entender como a modelagem e a especificação das classes contribui para obter-se uma predição segura. Desta forma, estimar o quão preciso é este modelo na prática, ou seja, quando o objetivo principal é avaliar se os cursos avaliados no ano de 2014 realmente propiciam obter uma boa nota nas próximas avaliações a serem realizadas de acordo com o INEP.

7. Conclusões

Realizadas as análises, pôde-se concluir que o presente trabalho apresentou alguns resultados interessantes a serem discutidos. O primeiro, relativo ao impacto das variáveis replica que a variável que mais influencia em um possível resultado de um curso para obter uma nota superior ao conceito três é a de UF, que representa a unidade da federação. Outro aspecto importante percebido é que o tempo de egresso do aluno, representado pelo intervalo de tempo que compreende o término do ensino médio ao início do curso de graduação, infere diretamente no objetivo de um curso obter um conceito superior a três, considerando o modelo gerado pela árvore de decisão.

Entretanto, essa perspectiva só pode ser considerada através do modelo inicial proposto pela estrutura, visto que a esta variável fora removida no modelo final gerado pela análise, e por consequência da regressão logística realizada. Outra conclusão relevante foi a de que os cursos de ensino superior que pertencem ao grupo de médio e baixo risco são muito importantes para garantirem a obtenção de uma boa nota.

A justificativa pode ser explicada pelo fato de que foram obtidos valores mais negativos na regressão logística considerando o grupo de variáveis apresentados por esta regressão. Acerca da análise da curva KS, o valor médio máximo foi reduzido para um valor aproximado de 14% quando for gerado o modelo final. Isto representa que o modelo obteve uma baixa discriminação, considerando a classe-alvo que foi avaliada através da base de dados do ENADE no ano de 2014.

Desta forma, conclui-se que a análise dos dados da base do ENADE requer o conhecimento dos aspectos que ponderam a obtenção de boas notas, dentre eles considerando o grão curso, a partir do momento que esta análise é enriquecida. A alimentação da base foi realizada com dados provenientes com dados da base de dados do ENEM, visto que muitas variáveis socioeconômicas podem influenciar no conceito a ser obtido e com uma boa nota na prova realizada nos próximos anos pelo órgão.

Referências

Cruz, A. J.; Nossa, V.; Balassiano, M.; Teixeira, A.; Desempenho dos Alunos no Enade de 2009: um estudo empírico a partir do conteúdo curricular dos cursos de Ciências Contábeis no Brasil. *Advances in Scientific and Applied Accounting*, Vol. 2, 2013.

INEP, Instituto Nacional de Pesquisas Educacionais Anísio Teixeira. Cálculo do Conceito do ENADE. Disponível em: http://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2010/Nota_Tecnica_Conceito_Enade_2010.pdf. Acesso em Mai. 2017a.

INEP, Instituto Nacional de Pesquisas Educacionais Anísio Teixeira Instrumentos. Disponível em: <http://portal.inep.gov.br/superior-condicoesdeensino-manuais>. Acesso em Mai. 2017b.

Santos, N.; Cunha, J.; Cornachione, E.; Análise do Desempenho dos Cursos de Ciências Contábeis do Estado de Minas Gerais no ENADE/2006. Congresso dos Programas de Pós-graduação em Ciências Contábeis de São Paulo, vol. 3., 2009.

Lemos, K.; Miranda, G.; Alto e Baixo Desempenho no ENADE: que variáveis explicam? *Revista Ambiente Contábil*, volume 7, 2015.

Martinez, E.; Louzada Neto, F.; Pereira, B.; A curva ROC para testes diagnósticos. Disponível em: http://www.po.ufrj.br/basilio/publicacoes/artigos/2003_a_curva_ROC_para_testes_diagnosticos_cadernos_saude_coletiva_vol11_JanJun.pdf > Acesso em 16 Jun. 2017.

Melo, Daniel. Regressão Logística. Material institucional do Centro de Informática da Universidade Federal de Pernambuco. Disponível em: <http://www.cin.ufpe.br/~rmcrs/PRED/RegressaoLogistica> > Acesso em 25 Jun. 2017.

Neves, A. P.; Souza, D.; Desempenho dos Estudantes das Instituições Públicas e Privadas no ENADE: Um estudo no Estado de Roraima, 2009.

Queiroz, Sérgio. Árvores de Decisão e Sistemas Inteligentes. Material institucional do Centro de Informática da Universidade Federal de Pernambuco. Disponível em: <http://www.cin.ufpe.br/~if684/EC/aulas/Aula-arvores-decisao-SI.pdf> > Acesso em Jun. 2017.