



UFTM – UNIVERSIDADE FEDERAL DO TRIÂNGULO MINEIRO

ALAN LOPES MELO

**USO DA TÉCNICA DE MINERAÇÃO DE DADOS COMO UMA FERRAMENTA DE
GESTÃO DA EVASÃO NO ENSINO SUPERIOR**

UBERABA

2019

ALAN LOPES MELO

**USO DA TÉCNICA DE MINERAÇÃO DE DADOS COMO UMA FERRAMENTA DE
GESTÃO DA EVASÃO NO ENSINO SUPERIOR**

Projeto de Defesa apresentado ao Programa de Mestrado Profissional em Inovação Tecnológica, da Universidade Federal do Triângulo Mineiro, como requisito parcial para obtenção do título de Mestre em Inovação Tecnológica.

Orientador: Prof. Dr. Julio Cesar de Souza Inácio Gonçalves

UBERABA

2019

**Catálogo na fonte: Biblioteca da Universidade Federal do
Triângulo Mineiro**

M485u Melo, Alan Lopes
Uso da técnica de mineração de dados como uma ferramenta de
gestão da evasão no ensino superior / Alan Lopes Melo. -- 2019.
94 f. : il., fig., graf., tab.

Dissertação (Mestrado Profissional em Inovação Tecnológica) --
Universidade Federal do Triângulo Mineiro, Uberaba, MG, 2019
Orientador: Prof. Dr. Julio Cesar de Souza Inácio Gonçalves

1. Ensino superior. 2. Universidades e faculdades - Administração.
3. Evasão universitária. 4. Mineração de dados (Computação). 5. Re
des neurais (Computação). I. Gonçalves, Julio Cesar de Souza Inácio.
II. Universidade Federal do Triângulo Mineiro. III. Título.

CDU 378

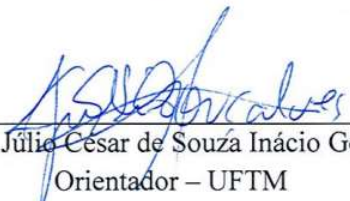
ALAN LOPES MELO

USO DA TÉCNICA DE MINERAÇÃO DE DADOS COMO UMA
FERRAMENTA DE GESTÃO DA EVASÃO NO ENSINO SUPERIOR

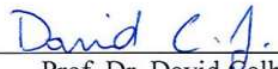
Trabalho de conclusão apresentado ao
Programa de Mestrado Profissional em
Inovação Tecnológica da Universidade Federal
do Triângulo Mineiro, como requisito para
obtenção do título de mestre.

Uberaba, 14 de fevereiro de 2019

Banca Examinadora:



Prof. Dr. Júlio César de Souza Inácio Gonçalves
Orientador – UFTM



Prof. Dr. David Calhau Jorge
Membro Titular – UFTM



Prof. Dr. Marcelo Ponciano da Silva

Dedico à minha esposa que me apoiou e me ajudou durante esta e tantas outras jornadas.

AGRADECIMENTOS

Primeiramente a Deus e a nossa Mãe Maria, que sempre ilumina os dias de trabalho, permitindo que eu tenha força de vontade para correr atrás dos nossos objetivos.

Aos meus familiares pelo apoio constante, nas horas mais necessárias.

Ao meu orientador Professor Julio Cesar de Souza Inácio Gonçalves, que encarou junto comigo o desafio deste trabalho.

RESUMO

Para este estudo é proposta uma pesquisa aplicada *ex post facto*, do tipo quantitativa e descritiva, com base nas informações sobre os alunos existentes na base de dados do Sistema Acadêmico (SISCAD) da UFTM. Esta pesquisa tem por objetivo identificar um padrão de informações a respeito dos alunos que pode ser relacionado à possibilidade de evasão do ensino superior. Com este estudo, foi possível a criação de uma ferramenta automatizada, baseada na técnica de mineração de dados utilizando-se redes neurais artificiais, capazes de identificar os alunos com tendência à evasão. O modelo criado foi capaz de identificar 63,8% dos alunos que evadiram. Dentre todos os alunos analisados durante a validação do modelo, este foi capaz de identificar corretamente a situação de 70,5% dos alunos (entre evadidos e não evadidos). Também foi possível definir que o modelo consegue identificar os alunos com tendência à evasão com 36 dias de antecedência da ocorrência da evasão, na média. Foi proposto um mecanismo de geração de alertas, mediante a produção de um relatório que pode ser disponibilizado diretamente na ferramenta de gestão acadêmica (SISCAD). Dessa forma, os agentes da Instituição, como os coordenadores de curso e os servidores da Pró-Reitoria de Assuntos Comunitários e Estudantis (PROACE) da UFTM, poderão atuar junto aos alunos com essa tendência de evasão e tentar revertê-la.

Palavras-chave: Sistema de gestão acadêmica. Evasão. Mineração de dados. Redes neurais artificiais. Alerta de evasão.

ABSTRACT

For this study, the proposed approach is an *ex post facto* applied research, of the quantitative and descriptive type, based on the information about the students existing in the database of the Academic System (SISCAD) of UFTM. This research aims to identify a pattern of information about students that may be related to the possibility of dropout of higher education. With this study, it was possible to create an automated tool, based on the technique of data mining using artificial neural networks, capable of identifying students with a tendency to evasion. The model was able to identify 63.8% of the students who dropout. Among all the students analyzed during the validation of the model, it was able to correctly identify the situation of 70.5% of the students (between evaded and not evaded). It was also possible to define that the model can identify the students a tendency to evasion 36 days in advance of the occurrence of evasion, on average. An alert generation mechanism has been proposed, through the production of a report that can be made available directly in the academic management tool (SISCAD). In this way, the Institution's agents, such as the course coordinators and the servers of the Pro-Rector's Office for Community and Student Affairs (PROACE) of the UFTM, will be able to act together with the students with this tendency of evasion and try to reverse it.

Keywords: Academic management system. Evasion. Data mining. Artificial neural networks. Avoidance alert.

LISTA DE ILUSTRAÇÕES

Figura 1 – Evolução do percentual de evasão anual total nas Instituições de Ensino Superior do Brasil	11
Figura 2 – Gráfico da evolução do percentual de evasão anual total nas instituições públicas de ensino superior	18
Figura 3 – Resumo esquemático da descoberta de conhecimento em dados (KDD).....	20
Figura 4 – Resumo esquemático da descoberta de conhecimento em dados (KDD) que se pretende realizar	21
Figura 5 – Diagrama representativo da equação do modelo matemático do NA de McCulloch e Pitts	25
Figura 6 – Exemplo de diagrama representativo de <i>underfitting</i> e <i>overfitting</i> em relação à determinação de uma função de $f(x)$	28
Figura 7 – Exemplo de diagrama representativo de <i>underfitting</i> e <i>overfitting</i> em relação a um problema de classificação de padrões de duas variáveis.....	28
Figura 8 – Dados dos atributos de entrada sem normalização	40
Figura 9 – Dados dos atributos de entrada após processo de normalização.....	40
Figura 10 – Processo de treinamento da MLP e verificação do erro nos registros de teste	44
Figura 11 – Diagrama representativo da MLP	45
Figura 12 – Detalhamento da quantidade de alunos ativos e a quantidade de alunos evadidos em cada um dos cursos	50
Figura 13 – Detalhamento da quantidade de alunos em tendência à evasão e a quantidade de alunos evadidos em cada um dos cursos.....	50
Figura 14 – Detalhamento da quantidade de alunos evadidos em comparação com a quantidade de alunos identificados como em tendência à evasão e a quantidade de alunos corretamente identificados como em tendência	51
Figura 15 – Detalhamento da quantidade de alunos identificados em comparação com a quantidade de alunos identificados pela primeira vez como em tendência à evasão por curso.....	52
Figura 16 – Distribuição da evasão na UFTM ao longo dos anos e dos meses, de fevereiro de 2013 a novembro de 2018	54
Figura 17 – Comparativo da distribuição da evasão na UFTM nos meses, ano a ano, de 2013 a 2018	55
Figura 18 – Comparativo da distribuição da evasão na UFTM em meses desde o início do semestre, ano/semestre a ano/semestre, de 2013 a 2018.....	55
Figura 19 – Diagrama representando um resumo dos resultados obtidos	59
Figura 20 – Modelo de relatório proposto listando os alunos identificados com tendência à evasão..	60

LISTA DE TABELAS

Tabela 1 – Apresentando as principais funções de ativação utilizadas, seu nome, fórmula e respectivo gráfico.....	26
Tabela 2 – Resultado da validação do modelo em relação ao total de alunos analisados, no período de abril a novembro de 2018, em relação à situação do aluno em dezembro/2018.....	45
Tabela 3 – Resultado da validação do modelo em relação aos alunos evadidos, no período de abril a novembro de 2018, em relação à situação do aluno em dezembro/2018.....	46
Tabela 4 – Dados da validação das cópias gravadas mensalmente, nas datas referenciadas pelos nomes das colunas, em relação à situação do aluno na cópia do banco de dados gravado em 13/12/2018.....	47
Tabela 5 – Dados de ingresso, matrículas, concluintes e evasão da graduação presencial da educação superior de 2000 a 2017 no Brasil.....	68
Tabela 6 – Dados de ingresso, matrículas, concluintes e evasão da graduação presencial pública da educação superior de 2000 a 2017 no Brasil.....	68
Tabela 7 - Dados de ingresso, matrículas, concluintes e evasão da graduação presencial privada da educação superior de 2000 a 2017 no Brasil.....	69
Tabela 8 - Dados de ingresso, matrículas, concluintes e evasão da graduação presencial pública da educação superior de 2000 a 2017 da região Sudeste.....	69
Tabela 9 - Dados de ingresso, matrículas, concluintes e evasão da graduação presencial pública da educação superior de 2000 a 2017 de Minas Gerais.....	69
Tabela 10 - Dados de ingresso, matrículas, concluintes e evasão da graduação presencial pública da educação superior de 2000 a 2017 da UFTM.....	70
Tabela 11 – Dados comparativos da distribuição da evasão na UFTM nos meses, ano a ano, de 2013 a 2018.....	72
Tabela 12 – Dados comparativos da distribuição da evasão na UFTM em meses desde o início do semestre, ano/semestre a ano/semestre, de 2013 a 2018.....	72
Tabela 13 – Dados comparativos da identificação de alunos em tendência à evasão pelo modelo, por cursos no mês de agosto de 2018.....	73

SUMÁRIO

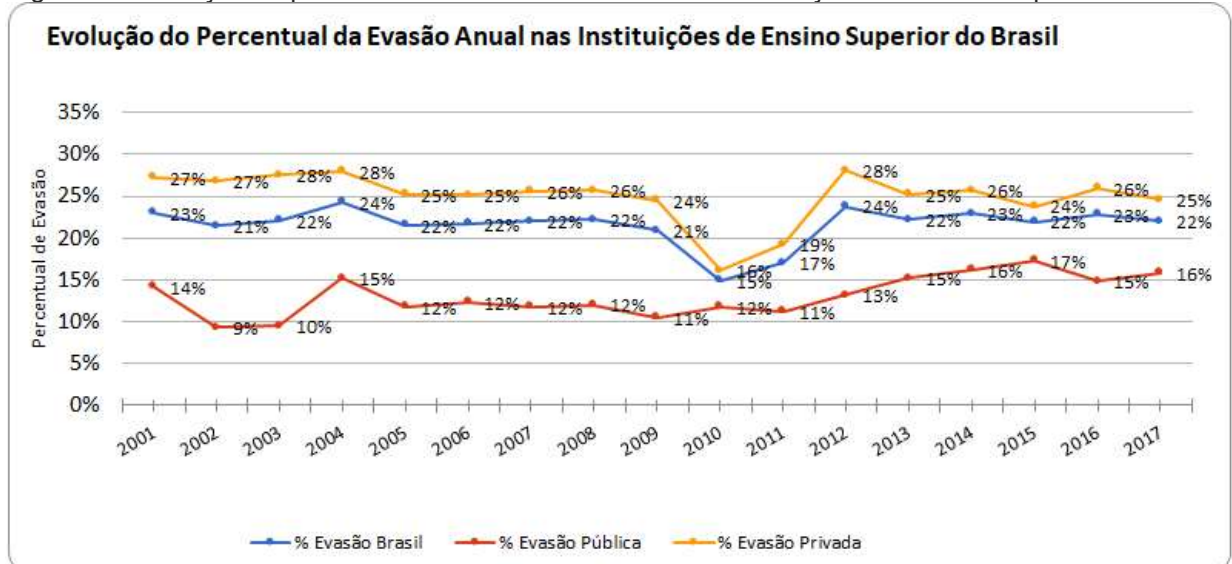
1 INTRODUÇÃO	11
1.1 OBJETIVOS	12
1.1.1 Objetivo geral	12
1.1.2 Objetivos específicos	12
2 REVISÃO DE LITERATURA	13
2.1 O PANORAMA BRASILEIRO SOBRE A EVASÃO DAS INSTITUIÇÕES DE ENSINO SUPERIOR	13
2.2 A REGIÃO SUDESTE, MINAS GERAIS E A UFTM	18
2.3 SISCAD	19
2.4 <i>DATA MINING</i>	19
2.5 REDES NEURAIS ARTIFICIAIS (RNA) E A <i>MULTILAYES PERCEPTRON</i> (MLP)	23
3 MATERIAL E MÉTODOS	30
3.1 CONHECIMENTO DO NEGÓCIO	30
3.2 OBTENÇÃO DOS DADOS	30
3.3 CONHECIMENTO DOS DADOS	31
3.4 LIMPEZA DOS DADOS	35
3.5 SELEÇÃO DO MODELO ANALÍTICO E DO ALGORITMO DE PROCESSAMENTO	36
3.6 TRANSFORMAÇÃO DOS DADOS – NORMALIZAÇÃO	38
3.7 <i>DATA MINING</i>	40
3.8 TESTE	41
3.9 VALIDAÇÃO DO MODELO	41
4 RESULTADOS OBTIDOS E DISCUSSÃO	44
5 CONCLUSÃO E SUGESTÃO DE TRABALHOS FUTUROS	61
5.1 CONCLUSÃO	61
5.2 SUGESTÕES DE TRABALHOS FUTUROS	62
REFERÊNCIAS	63
APÊNDICE A - DADOS DE INGRESSO, MATRÍCULAS, CONCLUINTES E EVASÃO DA GRADUAÇÃO PRESENCIAL DA EDUCAÇÃO SUPERIOR	68
APÊNDICE B - BREVE DESCRIÇÃO DO CURRÍCULO DO AUTOR	71
APÊNDICE C – DADOS COMPARATIVO DA DISTRIBUIÇÃO DA EVASÃO NA UFTM NOS MESES	72
APÊNDICE D – DADOS COMPARATIVO DA IDENTIFICAÇÃO DE ALUNOS EM TENDÊNCIA À EVASÃO PELO MODELO, POR CURSOS NO MÊS DE aGOSTO/2018	73
ANEXO A – PEDIDO DE AUTORIZAÇÃO DE ACESSO A DADOS	74
ANEXO B – COMPROVANTE DE ENVIO DO PROJETO AO CEP	75
ANEXO C – PARECER CONSUBSTANCIADO DO CEP	76
ANEXO D – RELATÓRIO DE REPROVAÇÕES POR DISCIPLINA E ANO/SEMESTRE ORDENADAS PELO PERCENTUAL DE REPROVAÇÕES	80
ANEXO E – RELATÓRIO DE RELAÇÃO DE REPROVAÇÕES POR DISCIPLINA	85

1 INTRODUÇÃO

A evasão dos alunos das instituições de ensino superior do Brasil é um problema crônico, ou seja, um problema antigo, persistente e ainda sem solução definitiva.

No Brasil a evasão anual se mantém praticamente constante acima de 20%. Nas instituições públicas, desde 2003, a evasão anual está consistentemente em patamares acima de 10%, e a situação nas instituições privadas é ainda pior (INEP, 2017). Na Figura 1 é apresentado um detalhamento da evolução do percentual de evasão anual no Brasil.

Figura 1 – Evolução do percentual de evasão anual total nas Instituições de Ensino Superior do Brasil



Fonte: O autor (2018); Dados INEP – Censo da Educação Superior.

A UFTM encontra-se em uma situação ainda mais crítica que a média nacional, pois apesar de ser uma instituição de ensino superior pública, apresenta evasão anual em quase todos os anos superior a 20% (UFTM, 2017).

Muitos estudos têm sido feitos de forma a tentar entender as causas e as possíveis soluções da evasão (BAGGI; LOPES, 2011; TONTINI; WALTER, 2014; GÓMEZ; TORRES, 2015; GILIOLI, 2016). A literatura sobre o assunto é bastante ampla e em sua maioria é discutida uma revisão do ensino superior, como um todo: reforma de grades, maior aderência do curso ao mercado de trabalho, maior acompanhamento dos alunos, etc. Essas soluções passam por questões estruturantes do ensino superior e, portanto, complexas e de longo prazo. Mas há

ações que podem ser feitas no curto prazo a fim de se reduzir a evasão, como o que foi realizado por Tontini e Walter (2014), onde os coordenadores de curso mantiveram contato com os alunos identificados como em tendência à evasão e conseguiram reduzir o índice de evasão.

É proposta uma abordagem então de curto prazo, em que se utilizando dados já coletados de registro e controle acadêmico dos alunos na UFTM, possa ser realizada uma triagem dos alunos em relação à sua tendência à evasão. Esses alunos então identificados como em tendência à evasão poderão ser encaminhados para agentes da instituição a fim de se tentar reverter essa tendência.

1.1 OBJETIVOS

Baseado nesse contexto foram propostos os seguintes objetivos:

1.1.1 Objetivo geral

Criar uma ferramenta, baseada na técnica de mineração de dados, para auxiliar na redução da evasão acadêmica do ensino superior na UFTM.

1.1.2 Objetivos específicos

- Determinar os sinais apresentados pelo acadêmico antes de evadir, sinais estes presentes nos dados do aluno de registro e controle acadêmico.
- Identificar quais os acadêmicos apresentam elevada probabilidade de evasão.
- Criar um mecanismo de geração de alertas dos acadêmicos em tendência de evasão.

2 REVISÃO DE LITERATURA

2.1 O PANORAMA BRASILEIRO SOBRE A EVASÃO DAS INSTITUIÇÕES DE ENSINO SUPERIOR

Ao analisar a Figura 1, pode ser verificado que a evasão da educação superior é historicamente elevada no Brasil, sendo quase sempre superior a 20%. Embora nas instituições privadas a taxa de evasão seja maior que nas instituições públicas, em ambos os casos a evasão é bastante expressiva e inclusive vem se agravando a partir do ano 2010 nas instituições públicas.

O gráfico representado na Figura 1 foi gerado pelo próprio autor, baseado nas informações disponibilizadas pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) através das Sinopses Estatísticas da Educação Superior – Graduação de 2000 a 2015.

Para a confecção desse gráfico foi utilizada a metodologia de cálculo da evasão anual proposta por Silva Filho e outros (2007) do Instituto Lobo para o Desenvolvimento da Educação, da Ciência e da Tecnologia. Essa metodologia foi posteriormente incorporada pelas instituições de ensino a partir do Fórum de Pró-Reitores de Planejamento e Administração (FORPLAD) de 2015, transformando-se em um dos indicadores institucionais.

Essa metodologia define que a evasão anual é o percentual de acadêmicos que “tendo terminado um período letivo sem concluir o curso não volta a se matricular” (SILVA FILHO; LOBO, 2012, p. 2). Portanto, o cálculo é feito pelo percentual da diferença entre a quantidade de acadêmicos que eram esperados terem se matriculado e a quantidade que efetivamente se matriculou. A quantidade de acadêmicos que eram esperados terem se matriculado é dada pela quantidade de acadêmicos matriculados no ano desejado, menos a quantidade de alunos ingressantes no ano. O percentual de acadêmicos que efetivamente se matriculou é definido pela quantidade de matrículas no ano anterior, menos a quantidade de concluintes do ano anterior. Esse cálculo é expresso na equação (1), extraído dos indicadores FORPLAD (2015).

$$Ev_p = \left(1 - \frac{M_p - I_p}{M_{p-1} - C_{p-1}} \right) \times 100 \quad (1)$$

Em que: M_p é o número de matriculados no período p ; I_p é o número de ingressantes no período p ; M_{p-1} é o número de matriculados no período anterior; C_{p-1} é o número de concluintes no período anterior.

O impacto dessa evasão é importante, entre outros motivos, devido ao investimento necessário para prover a formação superior, sendo o custo médio anual por aluno de graduação muito elevado. Foi realizado um estudo por Ducati e Alberton (2016) na Universidade Federal de Santa Catarina (UFSC) em 2015, em que foi estimado que o custo médio anual por aluno nessa instituição era de R\$ 19.656,46. Um outro estudo feito por Kassai e outros (2010), na Universidade de São Paulo (USP), definiu um custo médio anual de R\$ 28.700,00. Há uma grande diferença de valores, inclusive destacada pelos autores, devido aos cursos verificados nessa análise, havendo grande variação de valores nos cursos, mesmo dentro da mesma instituição. Essa variação ainda pode ter se acentuado devido aos estudos referenciados serem de modelos de gestão diferentes, sendo um estadual e o outro federal. Mesmo assim, em ambos os casos os valores são expressivos e pode-se dizer que esse valor investido na formação do aluno é perdido quando da ocorrência de evasão, de forma total ou parcialmente (no caso dos alunos que reingressam no ensino superior essa perda pode ser revertida parcialmente através de aproveitamento de disciplina). A situação se torna ainda mais grave no caso dos alunos que frequentam a universidade pública e gratuita, pois o governo, e, portanto, a população, arca com estes custos. A população, que apesar de custear a formação do aluno, não recebe a contrapartida gerada pela sua formação (maior disponibilidade de profissionais capacitados, melhores serviços e produtos e um incremento na produtividade do país). O aluno também arca com um prejuízo pessoal imensurável com a não obtenção do diploma.

Existe ainda o questionamento se há um retorno do valor investido na formação parcial de um aluno do ensino superior que não obteve o diploma como, por exemplo, uma correlação entre o tempo investido no ensino superior e o valor de salário médio recebido. Há estudos que demonstram que a variação do salário médio recebido durante o período que o aluno está cursando o ensino superior é bastante baixa, principalmente se comparada ao valor médio de rendimentos do aluno que ocorre pouco tempo depois da obtenção do diploma (SPENCE, 2001). Conforme pode ser verificado no estudo realizado por Spence (2001), demonstrou-se que na prática a

posse do diploma de curso superior possui grande influência nos critérios de contratação e remuneração do candidato, trabalho esse que lhe rendeu, juntamente com dois outros autores de trabalhos independentes, o prêmio Nobel de economia em 2001.

O governo brasileiro, tendo em vista essa preocupante situação, tem buscado soluções, conforme pode ser verificado no decreto nº 6.096, de 24 de abril de 2007, de criação do Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI), que em seu Art. 2º, inciso I, define que uma das diretrizes será a “**redução das taxas de evasão**, ocupação de vagas ociosas e aumento de vagas de ingresso, especialmente no período noturno” (BRASIL, 2007, grifo nosso). Infelizmente, como pode ser observado no gráfico da Figura 1, essas ações até o momento tiveram baixa efetividade.

Para entender melhor o problema é necessário contextualizar o que é a evasão. Bueno (1993, p. 13) define a evasão como “uma postura ativa do aluno que decide desligar-se por sua própria responsabilidade”. A Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras (1996) expande esse conceito ao definir três tipos de evasão: a) a evasão do curso, onde o acadêmico se desvincula do curso, o que pode ocorrer por não realização da matrícula, por trancamento de matrícula, por mudança de curso ou por cancelamento de matrícula por parte da instituição por não atendimento a alguma norma; b) evasão da instituição, que se caracteriza pelo desligamento do vínculo do acadêmico não somente com o curso ao qual está matriculado mas também com a instituição de ensino; c) evasão do sistema, que é o desligamento permanente ou temporário do acadêmico com o ensino superior como um todo.

Dentre os trabalhos que buscam tratar o assunto da evasão, a grande maioria busca identificar e tratar as causas principais da evasão. Esses estudos demonstram que a tratativa do problema envolve ações complexas e de longo prazo, como discutido pela revisão teórica elaborada por Gilioli (2016) para a Câmara dos Deputados. Neste estudo técnico, Gilioli (2016) discute vários trabalhos de outros autores a fim de buscar uma relação da forma de ingresso do Sistema de Seleção Unificada (SISU) com a evasão. Enquanto realiza essa revisão teórica, Gilioli (2016) acaba por fazer uma sumarização de diversos trabalhos anteriores, que indicam que a solução definitiva do problema da evasão passaria por uma reestruturação dos

modelos de ensino e das grades curriculares dos cursos de graduação, para dessa forma se tornarem mais aderentes ao mercado de trabalho.

Porém, para tentar reduzir a evasão no curto prazo há poucos estudos realizados e muitos deles envolvem a produção de trabalhos adicionais de triagem pelos integrantes da Instituição que teriam que empregar parte de seu tempo para aplicação das técnicas propostas. Um exemplo é um estudo realizado por Tontini e Walter (2014), onde foi possível reduzir em 18% a evasão no semestre subsequente ao da aplicação de questionários, a partir da intervenção junto aos alunos pelos coordenadores de curso. Para que esse trabalho continuasse a apresentar bons resultados, demandaria a reaplicação e reavaliação periódica de questionários para os alunos da instituição.

Não foi possível encontrar nenhum estudo no contexto brasileiro que verifique se os alunos que evadem do curso de graduação emitem sinais indiretos antes de deixarem o curso. A maioria dos trabalhos no sentido de verificação da evasão focam na aplicação de questionários com perguntas diretas sobre o abandono do curso. Esses questionários apresentam-se normalmente em dois formatos distintos. Alguns pesquisadores preferem questionários com uma abordagem direta, ou seja, perguntar ao aluno diretamente sobre sua intenção de evadir-se, utilizando perguntas como, por exemplo, “você já pensou em evadir-se da instituição?”. Há outros pesquisadores que têm preferência por questionários com perguntas indiretas, onde são feitos questionamentos diversos, mas sem indagar diretamente a intenção do aluno em evadir-se da instituição, mas apresentando perguntas como a opinião do aluno sobre a qualidade do curso, a aplicação profissional do aprendizado e a competência dos professores. Esse tipo de pesquisa, embora possa ter um bom resultado, pode apresentar um viés, pois o aluno questionado pode se sentir constrangido de responder com sinceridade a todas as perguntas. Além disso esses modelos de questionários são trabalhosos para sua elaboração, aplicação e análise, dificultando muito o processo de triagem dos alunos. Mesmo quando aplicados de modo online e automatizados, esses questionários não apresentam uma adesão de 100% dos alunos, ou seja, muitos não respondem. Nesse tipo de aplicação de questionários então torna-se impossível a identificação da situação dos alunos que não responderam, em relação à sua propensão a evadir-se.

Conforme apresentado no estudo biblioteconômico realizado por Machado et al. (2015), poucos estudos no mundo têm sido realizados através de técnicas

computacionais de inteligência artificial para se tentar identificar e reverter a evasão. De acordo com as publicações encontradas por Machado et al. (2015), pôde-se verificar que não há nenhum estudo brasileiro que utilize os dados já existentes dos sistemas de gestão acadêmica das instituições de ensino para aplicar técnica de mineração de dados e se definir os padrões de evasão.

Por outro lado, o Ministério da Educação dá indicações claras que é necessário que as instituições de ensino superior utilizem sistemas de gestão acadêmica, principalmente através da portaria nº 315/2018 (MINISTÉRIO DA EDUCAÇÃO, 2018a), na qual se dispõe que todos os documentos do acervo acadêmico sejam transformados em digitais até 2020. Além disso há também a portaria nº 1.095/2018 (MINISTÉRIO DA EDUCAÇÃO, 2018b) que estabelece parâmetros para a emissão digital de históricos escolares e diplomas. Desta forma espera-se que as instituições de ensino superior, que porventura ainda não utilizem um sistema de gestão acadêmica, deverão fazê-lo em breve, até mesmo para se adequarem às novas legislações.

Havendo base de dados bem estruturada e preenchida para fins de gestão acadêmica, estas bases podem ser utilizadas para outros fins. Essa pesquisa irá determinar que essas bases de dados podem ser utilizadas na criação de um modelo capaz de identificar os alunos com elevada propensão de se evadirem do curso de ensino superior, ou seja, que é possível, através dessas informações de gestão acadêmica, fazer uma análise de padrões relacionados à ocorrência de evasão do aluno. Através dessa análise de padrões da ocorrência de evasão criar um modelo capaz de definir quais os alunos estão em tendência a evasão.

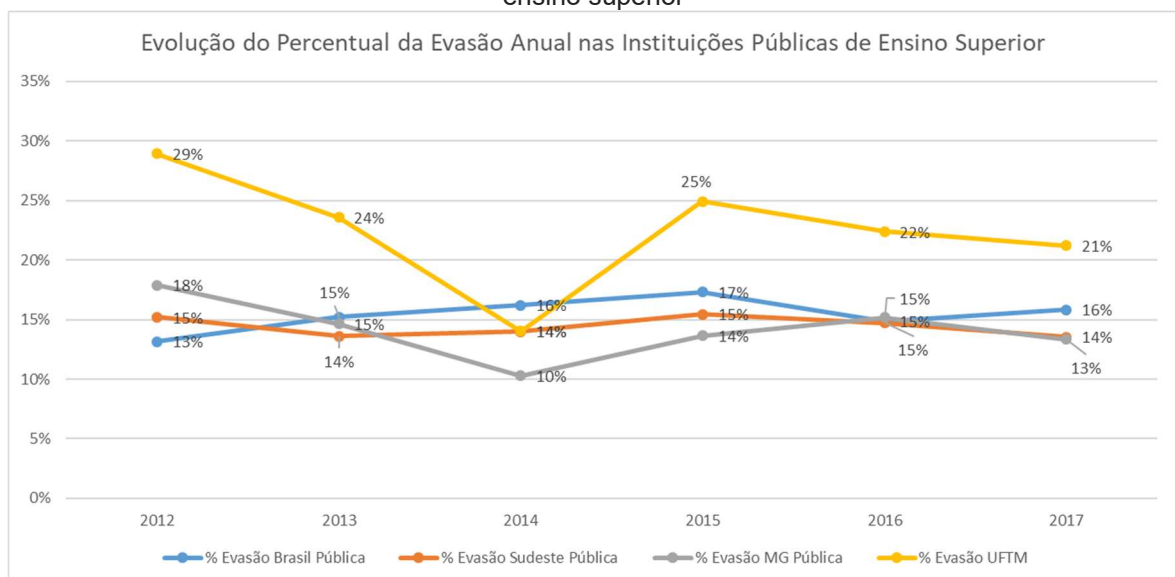
Acredita-se que, embora situações como baixa renda, distância da família, curso que o aluno frequenta, entre outros, tenham uma importância grande em relação à sua evasão do curso (GÓMEZ; TORRES, 2015), outros sinais também devem existir e talvez sejam até mais determinísticos para se definir o momento no qual o aluno irá se evadir. Acredita-se que possam haver sinais como aumento de faltas, queda no desempenho do curso, ocorrências de reprovações, ocorrências de trancamentos anteriores para o aluno, etc., que sejam capazes de definir o momento provável no qual o aluno deverá evadir-se da instituição de ensino.

2.2 A REGIÃO SUDESTE, MINAS GERAIS E A UFTM

A UFTM apresenta um quadro ainda mais grave que a situação nacional. Ao se comparar os dados de evasão da UFTM pode-se verificar, conforme apresentado na Figura 2, que ela apresenta a média de evasão anual um pouco maior que as médias da evasão das instituições públicas nacionais, da região Sudeste e de Minas Gerais. Com base nesse indicador pode-se observar que ações em relação à evasão na UFTM são necessárias de forma tão urgente como em outras instituições de ensino.

É interessante notar que no gráfico houve uma aparente queda na evasão no âmbito da UFTM no ano de 2014, mas essa aparente queda refere-se apenas a um problema da fórmula da evasão anual (Equação 1). Quando há um grande aumento no número de ingressantes, como o que ocorreu em 2015 na UFTM, resultado do aumento do número de ingressantes de maneira geral na Instituição, mas também devido à criação do *campus* de Iturama, com os cursos de Licenciatura em Química e Licenciatura em Ciências Biológicas, a equação apresenta esse aumento como se fosse uma aparente queda na evasão, mesmo não sendo o que ocorreu na realidade. Essa nova oferta de cursos, conjuntamente com o aumento de vagas nos demais cursos ocasionou um aumento de 36% na quantidade de ingressos da instituição em relação ao ano anterior (1.811 ingressantes em 2015, enquanto que em 2014 o número foi de 1.325). Essa é a razão da aparente queda da evasão na Figura 2, de 2014 a 2015 (UFTM, 2018).

Figura 2 – Gráfico da evolução do percentual de evasão anual total nas instituições públicas de ensino superior



Fonte: O autor (2018) / Dados INEP – Censo da Educação Superior e UFTM (2018).

2.3 SISCAD

O Sistema Acadêmico (SISCAD) é um sistema de gestão acadêmico desenvolvido pela Universidade Federal do Mato Grosso do Sul (UFMS) com base nos processos e necessidades daquela Instituição. Existia na UFTM um outro sistema de gestão acadêmico que já estava no limite da sua capacidade de utilização, principalmente por não ser um sistema web, mas também por ocorrer uma mudança do modelo de matrícula da UFTM, que deixaria de ser seriado para ser na forma de matrícula por disciplinas. No final de 2012 foi solicitada a participação de alguns membros da comunidade acadêmica para montarem uma comissão para avaliar um sistema de gestão acadêmica que a UFMS possuía. O sistema existente na UFMS era um sistema web que funcionava com a matrícula por disciplina. Essa análise tinha como objetivo verificar se o sistema existente na UFMS seria aderente às regras e processos da UFTM. Embora houvesse algumas ressalvas em relação à aderência aos processos internos, a comissão foi a favor da implantação do sistema e então firmou-se uma parceria de cooperação técnica entre a UFMS e a UFTM para implantação e manutenção deste sistema.

Desde sua implantação o SISCAD tem sido modificado para atender melhor às necessidades da UFTM, além de permitir o registro de mais informações acadêmicas e melhoria de sua utilização.

O SISCAD hoje é o sistema em que ficam armazenadas todas as informações do acadêmico desde o seu ingresso até o término de seu vínculo com a UFTM. Nele são registradas as disciplinas do aluno, suas notas, faltas e todo o seu histórico acadêmico.

Essas informações aqui descritas são de conhecimento do autor durante a sua vivência profissional, por ser um dos membros da equipe responsável pela manutenção do SISCAD. Maiores detalhes sobre a experiência profissional do autor podem ser vistos no apêndice B.

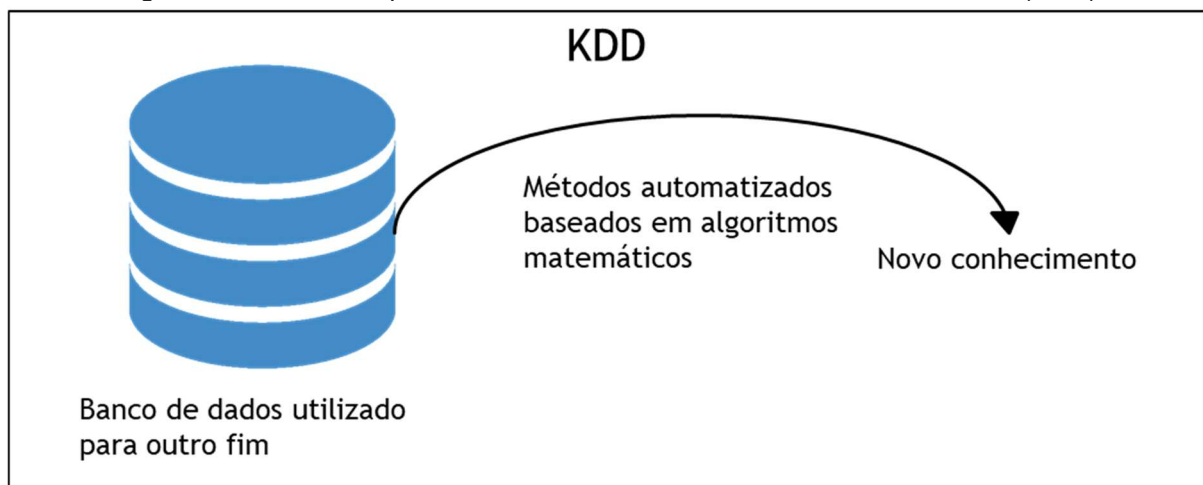
2.4 DATA MINING

Segundo a Oracle (2019), *data mining* pode ser entendido como o processo de se identificar padrões e tendências não conhecidas, através da análise de grandes volumes de dados, utilizando-se de métodos automatizados baseados em algoritmos

matemáticos. Em muitos casos o termo *data mining* é utilizado como sinônimo de KDD (do inglês *Knowledge Discovery in Data* – Descoberta de Conhecimento em Dados).

O *data mining* permite que se utilizem informações já coletadas para outros fins, também possam ser utilizadas para se identificar correlações não conhecidas. Com base nessas novas correlações, é possível fazer previsões com certo grau de confiança. A Figura 3 apresenta um resumo esquemático desse processo de descoberta de conhecimento em dados.

Figura 3 – Resumo esquemático da descoberta de conhecimento em dados (KDD)



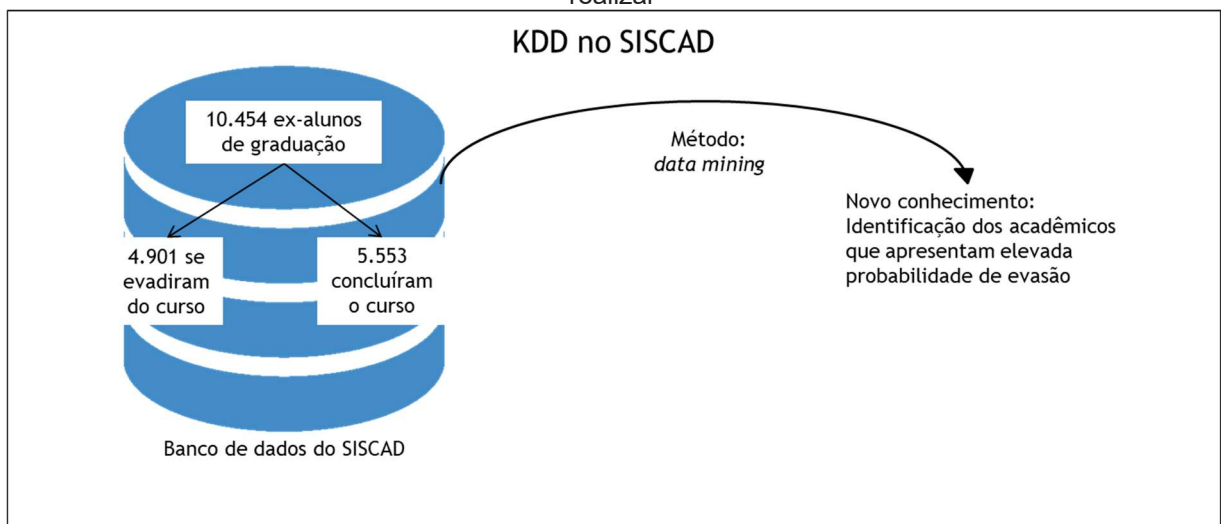
Fonte: O autor (2018).

O sistema SISCAD possui 10.454¹ ex-alunos de graduação cadastrados, sendo que 4.901¹ se evadiram do curso e 5.553¹ concluíram o curso. Esses ex-alunos possuem uma grande quantidade de informações cadastradas. São informações cadastrais (nome, e-mail, endereço, nomes dos pais, método de ingresso, etc.) e informações de controle e registro acadêmico (notas, frequências, reprovações, trancamentos, etc.). Essas informações podem então ser utilizadas para identificar padrões e características até então não exploradas. Busca-se então verificar se é possível caracterizar um padrão que diferencie os alunos que conseguem obter o diploma daqueles que se evadem da instituição de ensino superior sem concluir o curso (UFTM, 2017). A Figura 4 apresenta um resumo esquemático desse processo de descoberta de conhecimento em dados, que se pretende realizar no SISCAD a fim de se obter um modelo capaz de identificar os alunos com tendência à evasão.

¹ Dados obtidos através de consulta ao banco de dados do SISCAD fornecido, mediante autorização da PROENS e DTI.

É interessante observar que ao se fazer a proporcionalidade de alunos evadidos (4.901²) em relação ao total de ex-alunos de graduação cadastrados (10.454²), tem-se um valor de 46% de alunos evadidos. Ao se comparar esse dado com os dados da UFTM apresentados na Figura 2, é importante observar que os dados lá mencionados utilizam a fórmula da evasão anual (Equação 1), enquanto que esse resultado de 46% seria algo como a proporção total da evasão em relação aos ex-alunos.

Figura 4 – Resumo esquemático da descoberta de conhecimento em dados (KDD) que se pretende realizar



Fonte: O autor (2018).

De acordo com a literatura (AGGARWAL, 2015; ZENTUT, 2019; BRAMER, 2016; FAYYAD, 1996 apud OPREA et al., 2017), o processo de *data mining* pode ser subdividido nas seguintes atividades:

1. Conhecimento do negócio: é necessário um conhecimento claro do ambiente onde o problema que está sendo pesquisado se encontra. É necessário conhecimento das regras que regem esse ambiente, quais as interações ocorrem e como elas se manifestam. Essa necessidade existe, pois sem esse conhecimento se torna inviável a seleção dos atributos que podem ter relevância para o problema em questão. Além disso, também é necessário saber para cada um dos atributos quais os valores são possíveis de existirem e quais são erros de medição.

² Dados obtidos através de consulta ao banco de dados do SISCAD fornecido, mediante autorização da PROENS e DTI.

2. Obtenção dos dados: o processo de obtenção dos dados pode ser realizado através de *hardware* especializado, como redes de sensores, coleta manual de dados, aplicação de questionários, ou mesmo através de *softwares* para pesquisas em documentos. Quando os dados são coletados, esses dados podem não estar em um formato próprio para serem processados. Um exemplo seria no caso de aplicação de formulários de pesquisa: se os formulários forem aplicados em papel, esses precisam ser digitalizados. Se já estiverem em formato digital, precisam ser transformados em tabelas. Embora essa etapa aconteça antes do *data mining* propriamente dito, ela pode ter um grande impacto na qualidade dos dados, e pouca atenção nessa etapa pode causar grandes dificuldades nas etapas posteriores.
3. Conhecimento dos dados: etapa altamente interligada com o processo de obtenção dos dados. Essa etapa pode vir antes ou depois da etapa de obtenção dos dados. Nas pesquisas do tipo *ex post facto*, ou seja, a partir de fatos passados, essa etapa ocorre após a obtenção dos dados e seu principal objetivo é o descarte das informações que foram armazenadas, mas não possuem relevância para o problema a ser estudado. Para as demais pesquisas essa etapa acontece anteriormente à obtenção dos dados e deve nortear o recolhimento destes, de forma a evitar coletas desnecessárias. Nessa etapa são definidas, entre todas as variáveis disponíveis no ambiente pesquisado, quais serão analisadas e quais serão ignoradas.
4. Limpeza dos dados: durante os processos de coleta de dados podem ocorrer problemas que dificultam ou distorcem os resultados. Esses problemas podem ter as mais diversas origens: não preenchimento de informações, mudanças nos formulários de coleta de informação, falha no processo de preenchimento ou armazenagem dos dados. Para que os resultados sejam coesos é necessário que os dados de entrada sejam de boa qualidade, e por isso existem técnicas a serem aplicadas para tentar corrigir os dados incorretos ou incompletos. A técnica de correção a ser utilizada irá variar bastante para cada atributo, sendo as mais comuns as técnicas estatísticas. As técnicas estatísticas mais utilizadas são o preenchimento dos valores faltantes pelo valor médio do atributo ou pelo

valor de maior frequência. Em alguns casos mais graves a amostra inteira pode ser removida do conjunto de dados.

5. Seleção do modelo analítico e do algoritmo de processamento: nesta etapa é feita a seleção dentre todos os modelos analíticos e algoritmos de processamento, qual deles se mostra mais aderente ao problema em questão, ao ambiente em que está inserido e ao resultado esperado.
6. Transformação dos dados (também conhecida como normalização): alguns dos modelos existentes trabalham tanto com dados categóricos, quanto dados numéricos, outros somente dados numéricos. Alguns modelos aceitam dados contínuos e outros somente dados discretos. A etapa de transformação dos dados tem como objetivo transformar todos os dados em valores aceitos pelo modelo selecionado.
7. *Data mining*: é o processo de aplicação do modelo selecionado aos dados obtidos e transformados. Normalmente se realiza essa etapa buscando identificar padrões, regras ou criar modelos de predições numéricas, com base no conjunto de dados. É gerada então uma fórmula, um modelo ou um fluxo que é capaz de identificar esses padrões, regras ou realizar as predições numéricas.
8. Teste: etapa em que é verificada a coesão/precisão do modelo. É usual para essa etapa separar uma parte dos dados existentes em dois conjuntos de dados, chamados de conjunto de treinamento e conjunto de teste. É feita uma predefinição do tamanho percentual de cada um dos conjuntos, que poderá ser um requisito definido pelo problema ou definido a critério do pesquisador. Normalmente essa separação é feita tendo o conjunto de treinamento com 50% a 90% dos dados disponíveis e o conjunto de testes com 10% a 50% dos dados. A definição de qual amostra deverá pertencer a cada conjunto de dados é feita de maneira aleatória para evitar que algum dos conjuntos possua um viés (SILVA; SPATTI; FLAUZINO, 2010, p. 51; BRAMER, 2016, p. 80).

2.5 REDES NEURAS ARTIFICIAIS (RNA) E A *MULTILAYES PERCEPTRON* (MLP)

Segundo Silva, Spatti e Flauzino (2010, p. 121), um dos modelos de classificação (reconhecimento) de padrões que se destaca é o modelo de Rede Neural

Artificial do tipo *perceptron* de múltiplas camadas (PMC) (MLP – do inglês, *Multilayers Perceptron*). De fato, pode-se comprovar essa eficiência através da quantidade de trabalhos que utilizam MLP como um modelo para classificação de padrões. Também está quase sempre presente em trabalhos de comparação de performance de modelos de classificação de padrões e/ou de *data mining*, como os trabalhos de Kumar, Sehgal e Chauhan (2012) ou no trabalho de Romero et al. (2008) ou no trabalho de Popelka et al. (2012).

Além de apresentar uma boa performance, o modelo de MLP apresenta outras vantagens. É um modelo fácil de usar e de implementar, além de poder ser aplicado a muitos problemas reais de diferentes áreas (NIKAM, 2015; SILVA; SPATTI; FLAUZINO, 2010).

Entre os problemas da rede estão: pode ser necessário muito tempo de processamento para treinamento da MLP, em especial em redes muito complexas; não há uma definição clara de qual arquitetura da MLP terá um melhor resultado, sendo muitas vezes uma definição empírica ou baseada em tentativa e erro (MCCAFFREY, 2017); alguns problemas podem ser resolvidos por mais de um tipo de topologia de rede, sendo a escolha da melhor topologia de rede neural artificial para cada problema também dependente de tentativa e erro (ROMERO et al., 2008); apresenta uma característica de “caixa-preta”, onde pode ser bastante difícil a compreensão de como a MLP chegou em determinado resultado. Apesar disso alguns trabalhos, como o de Katz et al. (2017), buscam formas viáveis de “abrir” a caixa-preta e possibilitar esse tipo de investigação.

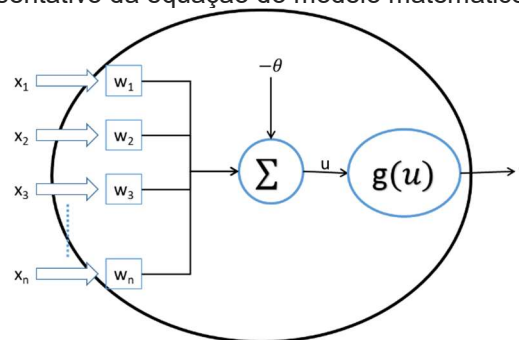
Rede neural artificial é um conceito que inicia sua construção com a proposta de McCulloch e Pitts (1943) de um modelo matemático para descrever o funcionamento de um neurônio biológico, que foi denominado de neurônio artificial (NA). Esse modelo matemático proposto se mantém até hoje com poucas alterações. O modelo proposto por McCulloch e Pitts (1943) pode ser resumido através da equação 2, e no diagrama representado na Figura 5, que expressam o mesmo conceito (SILVA; SPATTI; FLAUZINO, 2010).

$$u = \sum_{i=1}^n w_i \cdot x_i - \theta \quad (2)$$

$$y = g(u)$$

Em que: x_i representa os sinais de entrada; w_i são denominados pesos sinápticos, são valores que são multiplicados aos sinais de entrada para que o resultado do NA seja o esperado; Σ é denominado combinador linear. É o somatório de todos os sinais de entrada multiplicados pelos seus respectivos pesos sinápticos; θ é denominado limiar de ativação. É um patamar, uma espécie de peso adicional, que deve ser produzido pelo combinador linear para que esse possa gerar um valor em direção a saída do NA; u é denominado potencial de ativação. É o valor que deverá ser processado pela função de ativação para gerar a saída y ; $g(u)$ é a função de ativação. É uma função matemática que irá determinar a saída do NA; y conhecido como sinal de saída. É o valor de resposta do processamento do NA (SILVA; SPATTI; FLAUZINO, 2010).

Figura 5 – Diagrama representativo da equação do modelo matemático do NA de McCulloch e Pitts

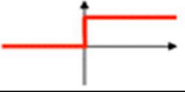
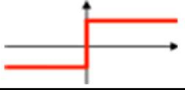

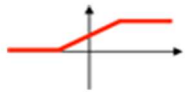




Fonte: Silva, Spatti e Flauzino (2010)

As funções de ativação mais comumente utilizadas podem ser vistas na Tabela 1 criada por Raschka (2017).

Posteriormente, Rosenblatt (1958) propôs o conceito do *perceptron*. O *perceptron* é uma rede neural artificial de um único NA, onde todos os valores de entrada eram processados sempre na direção da saída da rede neural artificial (RNA). Esse modelo de processamento dos dados é utilizado sempre na direção da saída da rede, sem ocorrência de valores de saída de camadas de NAs sendo utilizados como entrada em camadas anteriores da RNA. Posteriormente formou-se uma categoria de RNAs que ficou conhecida como *feedforward* (do inglês algo como alimentação à frente) (SILVA; SPATTI; FLAUZINO, 2010).

Tabela 1 – Apresentando as principais funções de ativação utilizadas, seu nome, fórmula e respectivo gráfico

Função de ativação	Equação	Exemplo	Gráfico
Degrau	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron	
Sinal	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron	
Linear	$\phi(z) = z$	Adaline, regressão linear	
Função definida por partes lineares	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Máquina de vetores de suporte	
Logística (<i>sigmoid</i>)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Regressão logística, Redes neurais de múltiplas camadas	
Tangente Hiperbólica	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	Regressão logística, Redes neurais de múltiplas camadas	

Fonte: Raschka (2017), traduzido pelo Autor

Em uma RNA, os NAs são organizados na forma de grupos. Normalmente cada grupo de NA possui uma ligação com todos os NAs do próximo grupo. A cada ligação é aplicado um peso sináptico, normalmente definido no modelo pela letra w , que vem do inglês *weight* (peso). Existem três grupos ou camadas especiais: a primeira denominada camada de entrada, onde ficam somente os valores de entrada da RNA com seus respectivos pesos; a segunda chamada de camada oculta (no inglês *hidden layer*); e a terceira, a camada de saída, que são os valores de resposta da RNA. Caso exista pelo menos uma camada oculta (pode-se adicionar quantas camadas ocultas quanto se desejar), a RNA é chamada de profunda (no inglês *deep*) ou multicamadas (no inglês *multilayers*) (SILVA; SPATTI; FLAUZINO, 2010).

Então, resumidamente, uma RNA do tipo *multilayers perceptron* é uma RNA onde existe pelo menos uma camada oculta (*multilayers*) (além das duas obrigatórias: a camada de entrada e a camada de saída) e que os dados são sempre processados em direção à saída (*perceptron*) da RNA.

Existem RNAs cujo padrão de processamento dos dados inclui a alimentação de NA com os dados de saída de NA em camadas posteriores no modelo, sendo uma retroalimentação do NA com os seus dados de saída. Exemplos desses tipos de RNA são as *recurrent neural networks* (RNN) e as *long/short term memory* (LSTM). Essas RNAs são bastante utilizadas para o processamento de dados temporais, que não é

o objetivo deste trabalho. De fato, alguns autores como Van Veen (2016) destacam a existência de pelo menos 27 tipos diferentes de RNA. Por isso não é objetivo desse trabalho esgotar esse assunto.

Após o processamento dos dados pela RNA, que inicialmente deve ter todos os seus pesos distribuídos aleatoriamente, verifica-se qual foi o erro obtido e se realizam ajustes nos pesos. Esse processo de ajustes dos pesos baseado no valor do erro obtido é chamado de *backpropagation*. Durante o processo de *backpropagation* é verificado o erro resultante do processamento, e são feitos pequenos ajustes nos pesos sinápticos da RNA a fim de se chegar mais próximo do resultado esperado. Esse processo de *backpropagation* é bastante complexo e pré-implementado em todas as ferramentas que trabalham RNA e já foi muito bem descrito por diversos autores como Silva, Spatti e Flauzino (2010, p. 95-120).

O processo de treinamento da RNA normalmente é realizado em ciclos. Cada ciclo é uma apresentação de todos os registros para a RNA e posteriores ajustes. Esses ajustes podem ser feitos após a apresentação de todos os registros, após a apresentação de cada registro, ou após a apresentação de um lote de registros. Não há um método que defina quais serão os melhores valores da quantidade de ciclos e do tamanho do lote de registros a ser apresentado à RNA antes de serem feitos os ajustes. É um processo em que devem ser verificados os melhores valores através de tentativa e erro.

Também em relação à quantidade de camadas ocultas da RNA, ou da quantidade de NA em cada camada oculta, não há métodos que definam quais serão os melhores a não ser o método de tentativa e erro. Assim como na camada de entrada, a camada de saída da RNA possui sua quantidade de NAs definida pelo problema, ou seja, pela quantidade de parâmetros de saída necessários.

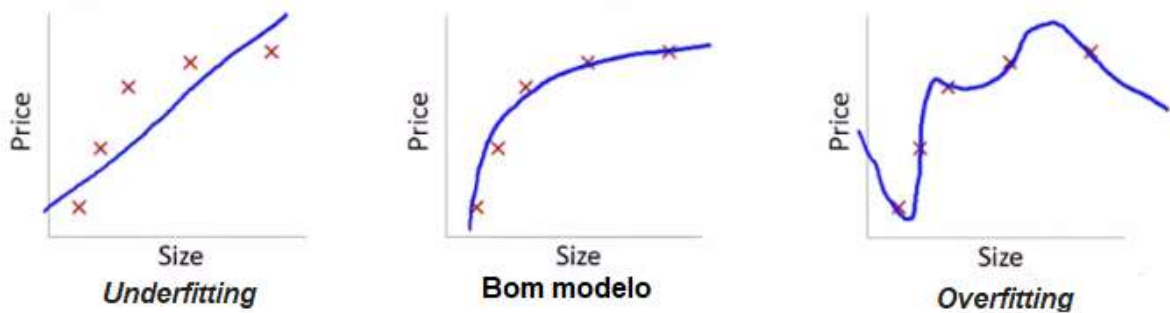
Para problemas de classificação, onde a saída esperada é do tipo sim ou não, alguns autores, como Silva, Spatti e Flauzino (2010, p. 121-132) indicam que o resultado do modelo gerado é melhor se forem utilizados dois parâmetros de saída do tipo percentual, ao invés de utilizar um único parâmetro do tipo sim ou não.

Há casos ainda que é necessária a utilização de um processo de *dropout* durante o treinamento da RNA em todas as camadas internas da RNA a fim de se reduzir o *overfitting*. O *overfitting* é um processo onde o modelo é treinado mais do que deveria, se tornando extremamente aderente aos dados de treinamento. Ao tornar-se muito aderente aos dados de treinamento acaba tendo pouca precisão

quando encontra novos dados (NG, 2019). A Figura 6 demonstra de forma gráfica como seria um exemplo de *underfitting*, *overfitting* e um bom modelo em relação à determinação de uma função de $f(x)$. E a Figura 7 demonstra de forma gráfica como seria um exemplo de *underfitting*, *overfitting* e um bom modelo em relação a um problema de classificação de padrões de duas variáveis.

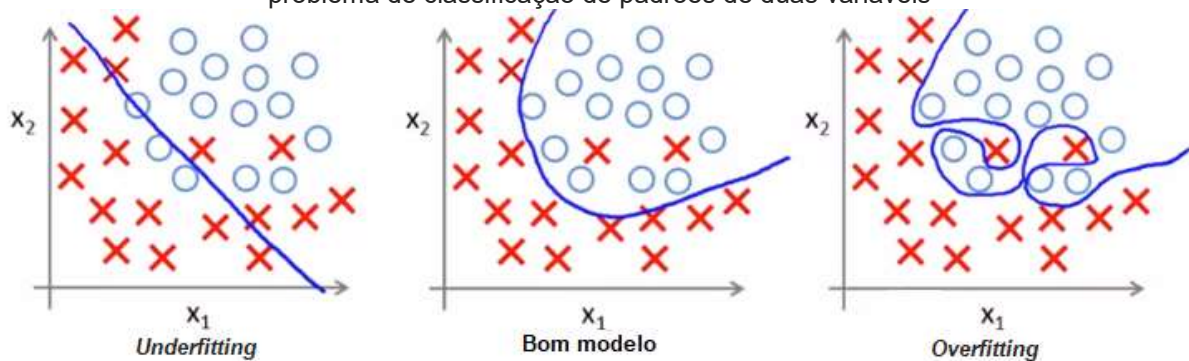
Em relação às funções de ativação, nenhuma função de ativação é utilizada na camada de entrada, pois nenhum processamento ocorre nessa camada. Para as camadas ocultas alguns autores sugerem a utilização da sigmoide ou da tangente hiperbólica, sendo a mais indicada a Tangente Hiperbólica (SILVA; SPATTI; FLAUZINO, 2010, p. 95-120).

Figura 6 – Exemplo de diagrama representativo de *underfitting* e *overfitting* em relação à determinação de uma função de $f(x)$



Fonte: Ng (2019).

Figura 7 – Exemplo de diagrama representativo de *underfitting* e *overfitting* em relação a um problema de classificação de padrões de duas variáveis



Fonte: Ng (2019).

Para a camada de saída, como este é um problema de reconhecimento de padrões e classificação, é interessante a utilização de alguma função que converta as saídas para o formato percentual. Para esse tipo de problema é recomendável a utilização da função softmax (DAHALL, 2019; MICROSOFT, 2019b). Na equação (3) é apresentada a função softmax (DAHALL, 2019; MICROSOFT, 2019b).

$$p_j = \frac{e^{v_j}}{\sum_{i=1}^n e^{v_i}} \quad (3)$$

Em que: tem-se e^{v_j} como sendo o número de Euler elevado ao potencial de ativação do NA j ; $\sum_{i=1}^n e^{v_i}$ sendo o somatório do número de Euler elevado ao potencial de ativação de todos os NA da camada de saída da RNA.

3 MATERIAL E MÉTODOS

3.1 CONHECIMENTO DO NEGÓCIO

A etapa de conhecimento do negócio foi realizada pela própria vivência profissional do autor dentro da Instituição. Por trabalhar diretamente com o sistema de gestão acadêmica anteriormente à realização do projeto de pesquisa, o autor possui um conhecimento prévio sobre o sistema e sobre os processos de gestão acadêmica do sistema. Um breve resumo da história profissional do autor pode ser verificado no apêndice B.

3.2 OBTENÇÃO DOS DADOS

A obtenção dos dados foi realizada através de solicitação de acesso via memorando para os setores responsáveis pela posse dos dados (Pró-Reitoria de Ensino – PROENS) e pela armazenagem (Departamento de Tecnologia da Informação – DTI). Cópia dessa autorização se encontra no anexo A. Após esse processo foi feita a análise do projeto pelo Comitê de Ética, através do número CAAE 71405317.0.0000.5154, que está no anexo B. O projeto então foi aceito pelo Comitê de Ética, através do parecer número 2.254.582, que se encontra no anexo C. Após o recebimento das autorizações foi disponibilizada uma cópia do banco de dados do SISCAD (UFTM, 2017).

Como esse banco de dados possui informações pessoais de alunos, o autor se compromete a manter sigilo sobre as informações ali contidas, além de se responsabilizar por não divulgar e não repassar nenhuma informação pessoal de alunos a terceiros. As informações pessoais dos alunos serão utilizadas somente com o objetivo de realização da pesquisa e não serão divulgadas informações pessoais ou que permitam a identificação pessoal de alunos. Todas as informações extraídas do banco de dados e utilizadas para publicação do estudo serão informações estatísticas e agregadas, que não possibilitam identificação individual de alunos.

3.3 CONHECIMENTO DOS DADOS

Como o autor já possuía experiência anterior com o banco de dados, e como os dados já tinham como origem um sistema, e, portanto, se encontravam bastante estruturados, essa etapa foi relativamente simples, sendo o enfoque maior a escolha de quais atributos seriam selecionados para processamento. Foram selecionados os seguintes atributos de entrada do modelo:

- Cidade do *campus*: Uberaba ou Iturama. Foi adicionado esse parâmetro pois pode haver uma diferença na tendência de evasão dos alunos a depender do *campus* que o aluno está matriculado;
- Instituto (ICENE, ICS, ICBN, ICTE, IELACHS ou Iturama): instituto no qual o aluno está matriculado;
- Tipo do Curso: tipo do curso no qual o aluno está matriculado. Na instituição de ensino são oferecidos cursos dos seguintes tipos: Aperfeiçoamento, Bacharelado e Licenciatura Plena;
- Área do conhecimento: área do conhecimento do curso no qual o aluno está matriculado, segundo definição do CNPQ (2019);
- Curso: curso no qual o aluno está matriculado;
- Idade: idade do aluno no momento em que ele deixou de ser um aluno da instituição;
- Tipo de ingresso (Enem, vestibular, outros processos seletivos, etc.): esse parâmetro foi adicionado pois alguns estudos questionam se o tipo de ingresso do aluno pode influenciar a tendência à evasão do aluno ou não (GÓMEZ; TORRES, 2015; GILIOLI, 2016);
- Ação afirmativa: ação afirmativa de ingresso do aluno, conforme lei nº 12.711/2012 (BRASIL, 2012);
- Quantidade de reprovações: quantas reprovações o aluno teve durante o curso;
- Quantidade de reprovações na mesma disciplina: quantidade máxima de reprovações na mesma disciplina;
- Quantidade de reprovações por falta: quantas reprovações por faltas o aluno teve durante o curso;

- Quantidade de reprovações por nota: quantas reprovações por nota o aluno teve durante o curso;
- Percentual médio de faltas: percentual médio de faltas nas disciplinas apresentado pelo aluno;
- Coeficiente de rendimento: coeficiente de rendimento conforme fórmula definida na regulamentação da graduação da instituição;
- Nota média: nota média do aluno nas disciplinas;
- Média da quantidade de disciplinas por semestre: quantidade média de disciplinas matriculadas pelo aluno a cada semestre;
- Média de carga horária das disciplinas por semestre: quantidade média da soma das cargas horárias das disciplinas matriculadas pelo aluno a cada semestre;
- Quantidade de disciplinas no último semestre: quantidade de disciplinas em que o aluno esteve matriculado no último semestre dele;
- Quantidade de reprovações no último semestre: quantidade de reprovações das disciplinas em que o aluno esteve matriculado no último semestre dele;
- Quantidade de reprovações por falta no último semestre: quantidade de reprovações por falta nas disciplinas em que o aluno esteve matriculado no último semestre dele;
- Quantidade de reprovações por nota no último semestre: quantidade de reprovações por nota nas disciplinas em que o aluno esteve matriculado no último semestre dele;
- Carga horária das disciplinas no último semestre: soma do total de carga horária das disciplinas nas quais o aluno esteve matriculado no último semestre dele;
- Percentual médio das faltas no último semestre: percentual médio de faltas em todas as disciplinas nas quais o aluno esteve matriculado no último semestre dele;
- Média das notas no último semestre: média das notas em todas as disciplinas nas quais o aluno esteve matriculado no último semestre dele;
- Percentual médio de faltas nos últimos 60 dias: percentual médio de faltas em todas as disciplinas nas quais o aluno esteve matriculado nos últimos 60 dias;

- Nota mínima nos últimos 60 dias: nota mínima em todas as disciplinas nas quais o aluno esteve matriculado apresentada pelo aluno nos últimos 60 dias;
- Nota média nos últimos 60 dias: nota média em todas as disciplinas nas quais o aluno esteve matriculado apresentada pelo aluno nos últimos 60 dias;
- Nota máxima nos últimos 60 dias: nota máxima em todas as disciplinas nas quais o aluno esteve matriculado apresentada pelo aluno nos últimos 60 dias;
- Possui trancamentos (Sim/Não): indicativo se o aluno possuiu trancamento de matrícula em algum momento;
- Quantidade de trancamentos: quantidade de trancamentos que o aluno possuiu durante a sua permanência no curso;
- Quantidade de disciplinas com alta reprovação percentual: quantidade de disciplinas com alta reprovação percentual (mais de 40% dos alunos matriculados reprovam na disciplina) nas quais o aluno está matriculado. A lista com essas disciplinas está disponível no anexo D.
- Quantidade de disciplinas com alta reprovação sequencial: quantidade de disciplinas com alta reprovação sequencial nas quais o aluno está matriculado. A lista com essas disciplinas está disponível no anexo E.
- Quantidade de disciplinas com média reprovação sequencial: quantidade de disciplinas com média reprovação sequencial nas quais o aluno está matriculado. A lista com essas disciplinas está disponível no anexo E.

Alguns dos parâmetros, em especial os parâmetros em relação aos dados dos últimos 60 dias, foram inseridos com o objetivo de possibilitar uma análise mais imediata no modelo gerado, ou seja, possibilitar uma melhor identificação de “quando” o aluno deve evadir. Embora haja regulamentação específica sobre esse assunto, pois a resolução nº 10/2012, do CONSU/UFTM (CONSU – Conselho Universitário da UFTM) (UFTM, 2012), que estabelece a regulamentação dos cursos de graduação da UFTM, através do seu artigo 201, determina que é necessário o preenchimento das informações de frequências com periodicidade quinzenal, na prática sabe-se que nem sempre isso ocorre. Mesmo assim, optou-se por incluir essas informações de curto prazo (últimos 60 dias), pois espera-se que quando preenchidas elas permitam um

melhor acerto de quando o aluno deverá evadir. Acredita-se que conforme a gestão incentive o preenchimento das informações de notas e frequências através de ações educativas e administrativas no aspecto de fazer valer a regulamentação existente, melhores serão os resultados do modelo gerado em relação a esses parâmetros dos últimos 60 dias.

Em relação ao parâmetro adicionado de “Quantidade de disciplinas com alta reprovação percentual”, verificou-se que existe no sistema SISCAD um relatório chamado de “Relatório de Reprovações por Disciplina e Ano/Semestre ordenadas pelo percentual de reprovações”. Esse relatório lista as disciplinas com percentual de reprovação maior que 40%, ordenadas de forma decrescente pelo maior percentual. Caso o aluno esteja matriculado em alguma dessas disciplinas, há uma alta probabilidade de ele ser reprovado. Definiu-se então, adicionar um parâmetro constando quantas dessas disciplinas o aluno está matriculado. Essa listagem de disciplinas utilizadas para esse parâmetro está disponível no anexo D.

Existe também no sistema SISCAD um relatório chamado “Relatório de Relação de Reprovações por Disciplina”, que lista até quantas vezes foi necessário que alunos repetissem a mesma disciplina para obterem a aprovação nessa disciplina. Há casos de alunos que precisam fazer a mesma disciplina até sete vezes (anexo E). A regulamentação da graduação na UFTM (UFTM, 2012), através do seu artigo 148, estabelece que poderá ocorrer o desligamento do discente por ato administrativo quando o aluno reprovar mais de quatro vezes na mesma disciplina. Foram então consideradas como “disciplinas com alta reprovação sequencial” todas aquelas em que existem casos de alunos que reprovaram mais de quatro vezes, segundo o “Relatório de Relação de Reprovações por Disciplina”, gerando potenciais jubilações.

De forma similar ao parâmetro “quantidade de disciplinas com alta reprovação sequencial”, entendeu-se que a existência de disciplinas em que existem alunos que tiveram que refazê-la por quatro vezes, pode ser também um potencial ponto de atenção a ser analisado. Essas disciplinas então foram analisadas através do parâmetro “quantidade de disciplinas com média reprovação sequencial”. Tanto as disciplinas utilizadas no parâmetro “quantidade de disciplinas com alta reprovação sequencial”, como as utilizadas no parâmetro “quantidade de disciplinas com média reprovação sequencial” encontram-se listadas no anexo E.

Foram selecionados os seguintes atributos de saída do modelo:

- Probabilidade de evasão: valor entre 0 e 1 indicando a probabilidade de evasão estimada pelo modelo;
- Probabilidade de conclusão de curso: valor entre 0 e 1 indicando a probabilidade de conclusão de curso estimada pelo modelo.

Os valores de saída do modelo devem ser ajustados de forma tal que a soma da probabilidade de evasão e da probabilidade de conclusão de curso sempre seja igual a 1.

3.4 LIMPEZA DOS DADOS

Durante o processo de checagem dos dados verificou-se que: o banco de dados possui 17.187 alunos registrados, sendo destes 10.454 alunos inativos. Dentre os alunos inativos, 5.553 foram inativos por terem obtido o diploma e por isso não eram mais alunos da instituição, e 4.901 estavam inativos por terem se evadido da Instituição (UFTM, 2017).

Devido à vivência do autor com o sistema SISCAD, sabe-se que o sistema que funcionava anteriormente na UFTM não possuía todas as informações, e mesmo as informações que existiam nesse sistema antigo, nem sempre estavam em um formato compatível com o modelo de dados do SISCAD. Então nesse processo de migração dos dados do sistema antigo para o SISCAD algumas informações foram simplificadas ou descartadas. Uma das informações que foram descartadas é em relação à frequência dos alunos, onde foram importadas somente o total de frequências e faltas dos alunos, mas foi descartada a informação de em quais datas individuais o aluno recebeu presença ou falta.

Sabe-se também que após a implantação do SISCAD na UFTM, houve um processo de melhoria da utilização do sistema, além de um processo evolutivo de melhorias no sistema ao longo dos anos. Esse processo evolutivo proporcionou o aumento progressivo das informações dos alunos que são armazenadas no sistema, além da melhoria da consistência dos dados.

Os cursos da UFTM apresentam em média 4 anos e meio de duração mínima (UFTM, 2017), então buscou-se localizar uma base de dados, baseado na data do último contato do aluno com a UFTM, ou seja, selecionar os dados de alunos que tenham passado a maior parte do seu tempo na UFTM com seus dados registrados

diretamente no SISCAD. Essa base de dados selecionada deve ser de tal forma que possua um bom conjunto de dados consistentes, mas também que não fosse uma base de dados pequena. Como a implantação do SISCAD ocorreu em torno de 2013, foram feitas verificações utilizando-se os dados dos alunos que tiveram seu último contato com a UFTM a partir de 2014, 2015 e 2016 e 2017. Para o conjunto de dados dos alunos que tiveram seu último contato com a UFTM a partir de 2017, obteve-se um conjunto de dados pequeno (1150 alunos). Para o conjunto de dados dos alunos que tiveram seu último contato com a UFTM a partir de 2014 e 2015, obteve-se um conjunto de dados com pouca consistência, sendo necessário a realização de correção de erros em uma parte considerável dos registros. Foi escolhido então se trabalhar com os dados dos alunos que tiveram seu último contato com a UFTM a partir de 2016.

Através desse processo definiu-se que o conjunto de dados a ser utilizado para processamento são todos os dados do último semestre dos alunos inativos, que tiveram seu último contato com a instituição a partir do ano de 2016. Nesse conjunto de dados foram encontrados 2.849 alunos inativos. Desses alunos inativos, 979 estavam inativos por terem obtido o diploma e por isso não eram mais alunos da instituição, e 1.870 estavam inativos por terem se evadido da instituição (UFTM, 2017).

A criação do modelo foi feita então utilizando-se dados relativos a 27% dos alunos inativos registrados no banco de dados do sistema. Em relação ao total de alunos registrados no sistema SISCAD, ou seja, em relação aos 17.187 alunos registrados foi utilizado para a criação do modelo 16% dos alunos.

3.5 SELEÇÃO DO MODELO ANALÍTICO E DO ALGORITMO DE PROCESSAMENTO

Devido aos bons resultados das MLPs em outros trabalhos (SILVA; SPATTI; FLAUZINO, 2010; KUMAR; SEHGAL; CHAUHAN, 2012; ROMERO et al., 2008; POPELKA et al., 2012; NIKAM, 2015), optou-se por fazer alguns testes com esse modelo. O tempo de processamento para o treinamento da MLP não foi um impeditivo da utilização da técnica, ficando em 20 minutos e 48 horas para treinamento de uma MLP em um computador de capacidade de processamento mediano. Sendo em torno de 20 a 30 minutos para o treinamento de uma MLP utilizando-se 10 ciclos e em torno

de 48 horas para treinamento ao se utilizar 2000 ciclos de treinamento. A questão do modelo se apresentar como uma caixa-preta no caso do problema pesquisado não foi considerado com um requisito prioritário, entre outros motivos pelo fato dos parâmetros de entrada estarem muito mais ligados a “sintomas”, do que causas da evasão. Baseado nisso, considerou-se que saber quais os parâmetros foram de maior importância para determinação do resultado, não seria um critério crítico para determinar o sucesso ou insucesso desse trabalho.

Para o processo de verificação do erro foi utilizado o modelo de *ClassificationError* (MICROSOFT, 2019b), e para o processo de ajustes dos pesos foi utilizado o modelo de *backpropagation*.

Para o processo de treinamento da RNA foram feitos testes com 10, 15, 20, 30, 40, 50, 100, 1.000 e 2.000 ciclos de treinamento. Também se fez a verificação de se tentar realizar o treinamento da RNA somente após a apresentação de todos os registros, após a apresentação de cada registro, ou após a apresentação de lotes de registros. Os melhores resultados obtidos foram utilizando-se o valor de 40 ciclos de treinamento e a realização de ajustes após a apresentação de lotes de 10 registros.

Em relação à quantidade de camadas ocultas da RNA, foram feitas diversas tentativas. Os melhores resultados foram obtidos ao se utilizar duas camadas ocultas, cada uma com 15 NAs. Na camada de entrada da RNA a quantidade de NAs é definida pela quantidade de parâmetros de entrada, sendo um NA para cada parâmetro, totalizando 33 NAs na camada de entrada. Assim como na camada de entrada, a camada de saída da RNA possui sua quantidade de NAs definida pelo problema, ou seja, pela quantidade de parâmetros de saída necessários. No caso deste trabalho a camada de saída foi definida como tendo dois NAs.

Foi utilizado um processo de *dropout* durante o treinamento da RNA em todas as camadas internas da RNA a fim de se reduzir o *overfitting*.

Para a camada de saída foi utilizada a função utilização da função softmax, conforme recomendado por alguns autores (DAHALL, 2019; MICROSOFT, 2019b).

Esses hiperparâmetros influenciam podem apresentar grande influência no tempo necessário para o treinamento da MLP, como sua capacidade de convergência além do erro final obtido. A fim de obter um conjunto adequado de hiperparâmetros foram criados e verificados cerca de duas centenas de modelos de MLP, onde os melhores resultados foram obtidos com os hiperparâmetros aqui descritos.

Através desse processo pôde-se fazer o treinamento da RNA, gerando os valores de pesos sinápticos e valores dos limiares de ativação. No total são 782 parâmetros que em conjunto à distribuição dos NA e com as funções utilizadas descrevem o modelo gerado.

3.6 TRANSFORMAÇÃO DOS DADOS – NORMALIZAÇÃO

Cada modelo analítico de *data mining* necessita dos dados em determinado formato. Alguns modelos aceitam variáveis categóricas ou numéricas, outros aceitam somente variáveis numéricas. Alguns trabalham somente com valores numéricos discretos, enquanto outros conseguem processar valores numéricos contínuos. Os dados então devem ser ajustados para o formato aceito pelo modelo analítico selecionado.

O modelo analítico da MLP trabalha somente com valores numéricos, inclusive valores contínuos. Esse modelo não consegue processar valores categóricos, por isso esses valores devem ser convertidos. Para a transformação dos dados categóricos em valores numéricos, nos casos onde existe uma tabela no banco de dados com chave primária numérica, foi utilizada a própria chave da tabela. Nos demais casos foi definida uma relação “de-para” para valores numéricos inteiros. Um exemplo em que essa situação ocorreu é no caso do parâmetro da área de conhecimento, onde foi feita a seguinte relação:

- O valor da área de conhecimento Ciências Agrárias foi convertido para o valor numérico 1;
- O valor da área de conhecimento Ciências Biológicas foi convertido para o valor numérico 2;
- O valor da área de conhecimento Ciências da Saúde foi convertido para o valor numérico 3;
- O valor da área de conhecimento Ciências Exatas e da Terra foi convertido para o valor numérico 4;
- O valor da área de conhecimento Ciências Humanas foi convertido para o valor numérico 5;
- O valor da área de conhecimento Ciências Sociais Aplicadas foi convertido para o valor numérico 6;

- O valor da área de conhecimento Engenharias foi convertido para o valor numérico 7;
- O valor da área de conhecimento Linguística, Letras e Artes foi convertido para o valor numérico 8;
- O valor da área de conhecimento Outros, por ser um valor especial foi convertido para o valor numérico 0.

Além disso para o processamento dos valores na MLP é recomendável que todos os valores de entrada estejam na mesma escala numérica. Essa necessidade existe para que determinado parâmetro não possua um peso inicial muito grande, atrapalhando a performance geral da MLP. Um caso onde valores podem ter uma escala bastante diferenciada é, por exemplo, a diferença em escala entre a idade de uma pessoa em anos (normalmente um valor menor que 100) e seu salário (normalmente na casa das centenas ou milhares).

Um modelo bastante utilizado para a normalização dos dados é a da distância do valor do atributo até sua média, medida em desvios-padrões do atributo. Assumindo-se que a maioria dos atributos deve seguir a distribuição normal, essa equação 4 produz valores bem ajustados, sendo quase que em sua totalidade dentro da faixa [-3,3] (AGGARWAL, 2015, p. 37);

$$z_j^i = \frac{x_j^i - \mu_j}{\sigma_j} \quad (4)$$

Em que: j representa o atributo da “posição” j, na lista de atributos; o valor μ_j é definida como sendo a média dos valores desse atributo; o valor σ_j é definido como sendo o desvio-padrão dos valores desse atributo; o valor x_j^i representa o valor do atributo j na i-ésima linha da amostra; dessa forma o valor z_j^i será o resultado do valor, após a normalização (AGGARWAL, 2015, p. 37).

Nas Figuras 8 e 9 são apresentadas algumas amostras conforme se apresentam antes da normalização e após a normalização.

Figura 8 – Dados dos atributos de entrada sem normalização

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	1	25	12	4	303	38	36	5	17	2	12	5	20	4,93	4,81	5,1	280,7	0
2	1	25	12	4	302	21	24	1	14	3	14	0	61	1,66	1,84	5,3	360	0
3	1	25	12	4	301	30	36	5	6	5	4	2	23	5,19	5,56	3,4	174	0
4	1	20	12	5	203	23	24	5	6	1	6	0	83	0	0	6	345	0
5	1	10	2	3	104	23	36	5	0	0	0	0	4	8,26	8,18	8,1	465	4
6	1	30	2	7	317	24	36	5	3	1	0	3	5	7,66	7,69	7	506,7	3
7	1	25	12	4	303	20	24	5	4	1	4	0	41	4,52	3,77	6	390	0
8	1	20	12	8	201	24	36	5	2	1	1	1	4	7,56	7,34	6,5	330	2
9	1	30	2	7	314	25	36	5	9	3	2	7	7	7,15	7,3	7,1	528,3	1
10	1	10	2	3	106	19	24	5	5	2	2	3	8	6,86	7,14	8	320	0

	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
1	0	0	0	0	0	0,13	0	10	10	10	0	0	0	0	0	1	0
2	0	0	0	0	0	0,3	0	10	10	10	0	1	0	0	0	1	0
3	0	0	0	0	0	5,6	0	10	10	10	0	0	0	0	0	1	0
4	0	0	0	0	0	0	0	10	10	10	0	0	0	0	0	1	0
5	0	0	0	435	0	9,27	0	8,8	9,42	10	0	0	0	0	0	0	1
6	0	0	0	570	6	8,63	0	7,5	9	10	0	0	0	0	0	0	1
7	0	0	0	0	0	3,77	0	10	10	10	1	1	0	0	0	1	0
8	0	0	0	90	0	7	0	7	7	7	0	0	1	1	0	0	1
9	0	0	0	60	10	3,6	20	5,4	7,2	9	0	0	0	0	0	0	1
10	0	0	0	0	0	6,23	0	10	10	10	0	0	0	0	0	1	0

Fonte: O autor (2018).

Figura 9 – Dados dos atributos de entrada após processo de normalização

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	-0,2186195	0,5595393	1,474174	-0,3548894	0,2569902	1,848759	0,7634473	0,4573431	1,290367	0,6568687	1,922746	0,32871	0,1237069	0,08325902	0,02657668	-0,1501397	-0,2951566	-0,5804321
2	-0,2186195	0,5595393	1,474174	-0,3548894	0,2528703	-0,4193061	-0,8432459	-3,13954	0,9293953	1,458676	2,344936	-0,6327835	1,798496	-0,8808983	-0,8535204	-0,0800393	0,1129644	-0,5804321
3	-0,2186195	0,5595393	1,474174	-0,3548894	0,2487504	0,7814343	0,7634473	0,4573431	-0,0331957	3,062291	0,2339897	-0,2481861	0,2462525	0,1599199	0,2488234	-0,7459925	-0,8442931	-0,5804321
4	-0,2186195	-0,0195117	1,474174	0,1295763	-0,1550005	-0,1524749	-0,8432459	0,4573431	-0,0331957	-0,1449389	0,6561788	-0,6327835	2,697163	-1,370348	-1,398766	0,1653118	0,0357662	-0,5804321
5	-0,2186195	-1,177614	-0,6701747	-0,8393551	-0,5628713	-0,1524749	0,7634473	0,4573431	-0,7551389	-0,9467465	-0,6103886	-0,6327835	-0,5298694	1,065107	1,025205	0,9013655	0,6533517	1,350276
6	-0,2186195	1,13859	-0,6701747	1,098508	0,3146689	-0,0190593	0,7634473	0,4573431	-0,3941673	-0,1449389	-0,6103886	-0,0558874	-0,4890209	0,8881977	0,8800042	0,5158135	0,8679627	0,8675987
7	-0,2186195	0,5595393	1,474174	-0,3548894	0,2569902	-0,5527217	-0,8432459	0,4573431	-0,2738434	-0,1449389	0,2339897	-0,6327835	0,9815258	-0,0376292	-0,2816055	0,1653118	0,2673607	-0,5804321
8	-0,2186195	-0,0195117	1,474174	1,582973	-0,1632403	-0,0190593	0,7634473	0,4573431	-0,5144912	-0,1449389	-0,399294	-0,4404848	-0,5298694	0,8587128	0,7762891	0,3405627	-0,041432	0,3849217
9	-0,2186195	1,13859	-0,6701747	1,098508	0,3023092	0,1143563	0,7634473	0,4573431	0,327776	1,458676	-0,1881995	0,7133073	-0,4073238	0,7378246	0,7644359	0,5508636	0,9791279	-0,0977552
10	-0,2186195	-1,177614	-0,6701747	-0,8393551	-0,5546315	-0,6861373	-0,8432459	0,4573431	-0,1535196	0,6568687	-0,1881995	-0,0558874	-0,3664753	0,6523183	0,7170232	0,8663152	-0,0928974	-0,5804321

	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
1	-0,0980228	-0,0947995	-0,0561781	-0,5616707	-0,1924431	-1,029662	-0,1453623	0,4599621	0,1521609	-0,0607556	-0,3892593	-0,327818	-0,181384	-0,2339043	-0,2316214	1	0
2	-0,0980228	-0,0947995	-0,0561781	-0,5616707	-0,1924431	-0,9855396	-0,1453623	0,4599621	0,1521609	-0,0607556	2,568074	1,178558	-0,181384	-0,2339043	-0,2316214	1	0
3	-0,0980228	-0,0947995	-0,0561781	-0,5616707	-0,1924431	0,3900513	-0,1453623	0,4599621	0,1521609	-0,0607556	-0,3892593	-0,327818	-0,181384	-0,2339043	-0,2316214	1	0
4	-0,0980228	-0,0947995	-0,0561781	-0,5616707	-0,1924431	-1,063403	-0,1453623	0,4599621	0,1521609	-0,0607556	-0,3892593	-0,327818	-0,181384	-0,2339043	-0,2316214	1	0
5	-0,0980228	-0,0947995	-0,0561781	0,9397947	-0,1924431	1,342583	-0,1453623	-0,2182893	-0,1827927	-0,0607556	-0,3892593	-0,327818	-0,181384	-0,2339043	-0,2316214	0	1
6	-0,0980228	-0,0947995	-0,0561781	1,405767	0,5569015	1,176474	-0,1453623	-0,9530619	-0,4253453	-0,0607556	-0,3892593	-0,327818	-0,181384	-0,2339043	-0,2316214	0	1
7	-0,0980228	-0,0947995	-0,0561781	-0,5616707	-0,1924431	-0,0849168	-0,1453623	0,4599621	0,1521609	-0,0607556	2,568074	1,178558	-0,181384	-0,2339043	-0,2316214	1	0
8	-0,0980228	-0,0947995	-0,0561781	-0,2510227	-0,1924431	0,753415	-0,1453623	-1,235667	-1,580358	-0,273801	-0,3892593	-0,327818	3,483603	2,171848	-0,2316214	0	1
9	-0,0980228	-0,0947995	-0,0561781	-0,354572	1,056465	-0,1290396	2,60364	-2,140002	-1,464857	-0,1317707	-0,3892593	-0,327818	-0,181384	-0,2339043	-0,2316214	0	1
10	-0,0980228	-0,0947995	-0,0561781	-0,5616707	-0,1924431	0,553565	-0,1453623	0,4599621	0,1521609	-0,0607556	-0,3892593	-0,327818	-0,181384	-0,2339043	-0,2316214	1	0

Fonte: O autor (2018).

Pode-se observar na Figura 9 que os dados adquirem uma característica bastante uniforme em relação à escala após o processo de normalização.

3.7 DATA MINING

Para o processo de execução do *data mining* foi utilizada a ferramenta *Cognitive Toolkit* (CNTK 2.6) da Microsoft®. Foi escolhida essa ferramenta por ser uma ferramenta gratuita, poderosa, de fácil utilização e de fácil integração com muitas das linguagens de programação comerciais modernas (atualmente C#/.Net, C++, Python, Java) e com versões para ser utilizada em diferentes sistemas operacionais (Microsoft Windows e Linux) (MICROSOFT, 2019a).

3.8 TESTE

O conjunto de dados existente após a etapa de limpeza, contendo 2.849 alunos, foi subdividida em dois conjuntos. Como não existe uma definição precisa de como deve ser a divisão desses dois conjuntos, somente uma indicação aproximada, onde é sugerido que o conjunto de treinamento tenha 50% a 90% dos dados disponíveis e o conjunto de testes tenha 50% a 10% dos dados (SILVA; SPATTI; FLAUZINO, 2010, p. 51; BRAMER, 2016, p. 80), optou-se por fazer a divisão dos dados em 80% dos dados para treinamento e 20% dos dados para teste.

O conjunto de dados de treinamento foi estipulado como sendo da proporção de 80% dos dados originais, ficando com 2.279 dados. Por sua vez, o conjunto de testes ficou com os 20% restante dos dados, totalizando dados de 570 alunos. A divisão de quais registros fazem parte do conjunto de dados de teste e quais os dados fazem parte do conjunto de treinamento foi feita de forma aleatória.

A aferição do erro resultante do modelo é realizada pela própria ferramenta CNTK, na sequência do processo de treinamento. A fórmula de cálculo do erro utilizada foi o modelo *ClassificationError*. Essa fórmula de *ClassificationError* responde para um modelo de reconhecimento de padrões (reconhecimento de em qual classe o registro se encontra), a distância do valor encontrado pelo modelo até o valor da classe esperada, que sempre deve ser igual a 1 (MICROSOFT, 2017b).

Após a aferição do erro em todos os registros, é feita a média do erro, que representa o erro total do modelo.

3.9 VALIDAÇÃO DO MODELO

Nesse ponto tem-se um modelo matemático capaz de identificar alunos em tendência de evasão. Esse modelo foi gerado com base em uma cópia do banco de dados do SISCAD, disponibilizado pelo DTI da UFTM. Essa cópia do banco de dados foi retirada em 30 de outubro de 2017.

Foi solicitado ao DTI cópias do banco de dados do SISCAD do ano de 2018, de janeiro a novembro, além de uma cópia única do banco de dados do mês de dezembro de 2018. Com base nesse pedido o DTI disponibilizou 100 cópias do banco de dados do SISCAD.

A intenção dessa solicitação de cópias do banco de dados de 2018 é simular como o modelo gerado se comportaria se estivesse já integrado ao sistema SISCAD, produzindo relatórios dos alunos com tendência de evasão em tempo real. A cópia do banco de dados do mês de dezembro foi utilizada como um indicador da situação “final” do aluno para essa validação. Nessa cópia do banco de dados, gravado no dia 13 de dezembro de 2018, todos os alunos cadastrados foram classificados em dois grupos conforme a sua última situação registrada: evadidos ou não evadidos.

Além da cópia do banco de dados de dezembro, o DTI também pôde disponibilizar algumas outras cópias do banco de dados conforme seguem: uma cópia por mês (último dia do mês) gravada nos meses de abril, maio, junho e julho; quatro cópias para o mês de agosto (01/08/2018, 10/08/2018, 20/08/2018 e 31/08/2018); cópias diárias nos meses de setembro, outubro e novembro.

Foi realizado o processamento do modelo criado sobre os dados dos alunos ativos em cada uma das cópias dos bancos de dados gravados entre abril e novembro e confrontada a situação prevista pelo modelo (evasão ou não evasão) em relação à situação do aluno encontrada na cópia do banco de dados gravada em dezembro. Esse processo foi realizado a fim de se definir o percentual de acerto do modelo criado.

Foi também realizado o processamento do modelo criado sobre os dados dos alunos ativos em cada uma das cópias dos bancos de dados gravadas diariamente entre setembro e novembro e confrontada a situação prevista pelo modelo (evasão ou não evasão) em relação à situação do aluno encontrada na cópia do banco de dados de dezembro. Nessa análise foi feito o processamento dos dados do banco de dados gravado no dia 31 de agosto de 2018, e todos os alunos que foram identificados com tendência de evasão foram retirados da análise do processamento dos bancos de dados gravados de setembro a novembro. Esse procedimento de análise dos dados gravados de setembro a novembro foi realizado com o intuito de se definir com quantos dias de antecedência há a ocorrência de evasão, o modelo identifica essa tendência à evasão. Dessa análise foram removidos os alunos identificados com tendência à evasão dos dados gravados em 31 de agosto de 2018, pois não se sabe qual seria a primeira data a identificar a tendência de evasão desses alunos.

Como não se obteve um bom resultado da média de dias na avaliação das cópias gravadas diariamente no período de setembro a novembro, repetiu-se o procedimento analisando-se as cópias disponíveis do banco de dados que foram

gravadas em agosto. De forma similar ao que foi feito para as cópias de setembro-novembro, foi feito o processamento dos dados de 30 de julho de 2018 e removidos os alunos identificados com tendência à evasão da análise dos dados gravados em agosto.

Foi feita a validação do modelo utilizando-se dados referentes a 6.985 alunos.

4 RESULTADOS OBTIDOS E DISCUSSÃO

Durante a fase de teste da MLP (item 3.8) obteve-se um erro médio do modelo de 0,0087 (0,87%) nos 570 registros de teste, conforme pode ser observado destacado em vermelho na Figura 10.

Figura 10 – Processo de treinamento da MLP e verificação do erro nos registros de teste

```
C:\WINDOWS\system32\cmd.exe
Finished Epoch[23 of 40]: [Training] ce = 0.03751219 * 2279; err = 0.01228609 * 2279; totalSamplesSeen = 52417; learningRatePerSample = 0.0049999999; epochTime=2.58488s
Finished Epoch[24 of 40]: [Training] ce = 0.04661507 * 2279; err = 0.01448004 * 2279; totalSamplesSeen = 54696; learningRatePerSample = 0.0049999999; epochTime=2.55505s
Finished Epoch[25 of 40]: [Training] ce = 0.04360091 * 2279; err = 0.01272488 * 2279; totalSamplesSeen = 56975; learningRatePerSample = 0.0049999999; epochTime=2.53778s
Finished Epoch[26 of 40]: [Training] ce = 0.03383676 * 2279; err = 0.01184730 * 2279; totalSamplesSeen = 59254; learningRatePerSample = 0.0049999999; epochTime=2.56287s
Finished Epoch[27 of 40]: [Training] ce = 0.04345378 * 2279; err = 0.01404125 * 2279; totalSamplesSeen = 61533; learningRatePerSample = 0.0049999999; epochTime=2.55971s
Finished Epoch[28 of 40]: [Training] ce = 0.03546226 * 2279; err = 0.01184730 * 2279; totalSamplesSeen = 63812; learningRatePerSample = 0.0049999999; epochTime=2.54837s
Finished Epoch[29 of 40]: [Training] ce = 0.044483614 * 2279; err = 0.01404125 * 2279; totalSamplesSeen = 66091; learningRatePerSample = 0.0049999999; epochTime=2.54891s
Finished Epoch[30 of 40]: [Training] ce = 0.03295653 * 2279; err = 0.00921457 * 2279; totalSamplesSeen = 68370; learningRatePerSample = 0.0049999999; epochTime=2.56165s
Finished Epoch[31 of 40]: [Training] ce = 0.02807841 * 2279; err = 0.00570426 * 2279; totalSamplesSeen = 70649; learningRatePerSample = 0.0049999999; epochTime=2.52879s
Finished Epoch[32 of 40]: [Training] ce = 0.03642007 * 2279; err = 0.01360246 * 2279; totalSamplesSeen = 72928; learningRatePerSample = 0.0049999999; epochTime=2.54171s
Finished Epoch[33 of 40]: [Training] ce = 0.04240581 * 2279; err = 0.01448004 * 2279; totalSamplesSeen = 75207; learningRatePerSample = 0.0049999999; epochTime=2.55377s
Finished Epoch[34 of 40]: [Training] ce = 0.04102773 * 2279; err = 0.01579640 * 2279; totalSamplesSeen = 77486; learningRatePerSample = 0.0049999999; epochTime=2.60177s
Finished Epoch[35 of 40]: [Training] ce = 0.04286823 * 2279; err = 0.01360246 * 2279; totalSamplesSeen = 79765; learningRatePerSample = 0.0049999999; epochTime=2.53369s
Finished Epoch[36 of 40]: [Training] ce = 0.03904687 * 2279; err = 0.01316367 * 2279; totalSamplesSeen = 82044; learningRatePerSample = 0.0049999999; epochTime=2.52921s
Finished Epoch[37 of 40]: [Training] ce = 0.03524109 * 2279; err = 0.01096972 * 2279; totalSamplesSeen = 84323; learningRatePerSample = 0.0049999999; epochTime=2.51668s
Finished Epoch[38 of 40]: [Training] ce = 0.02780721 * 2279; err = 0.01009215 * 2279; totalSamplesSeen = 86602; learningRatePerSample = 0.0049999999; epochTime=2.53236s
Finished Epoch[39 of 40]: [Training] ce = 0.03440005 * 2279; err = 0.01140851 * 2279; totalSamplesSeen = 88881; learningRatePerSample = 0.0049999999; epochTime=2.50749s
Finished Epoch[40 of 40]: [Training] ce = 0.03592059 * 2279; err = 0.01228609 * 2279; totalSamplesSeen = 91160; learningRatePerSample = 0.0049999999; epochTime=2.51526s

#####
#
# WriteProbs command (write action)
#
#####
Minibatch[0]: ActualMBSize = 570
Written to TestProbs_txt*
Total Samples Evaluated = 570

#####
#
# DumpWeights command (dumpNode action)
#
#####
Warning: node name '__AllNodes__' does not exist in the network. dumping all nodes instead.

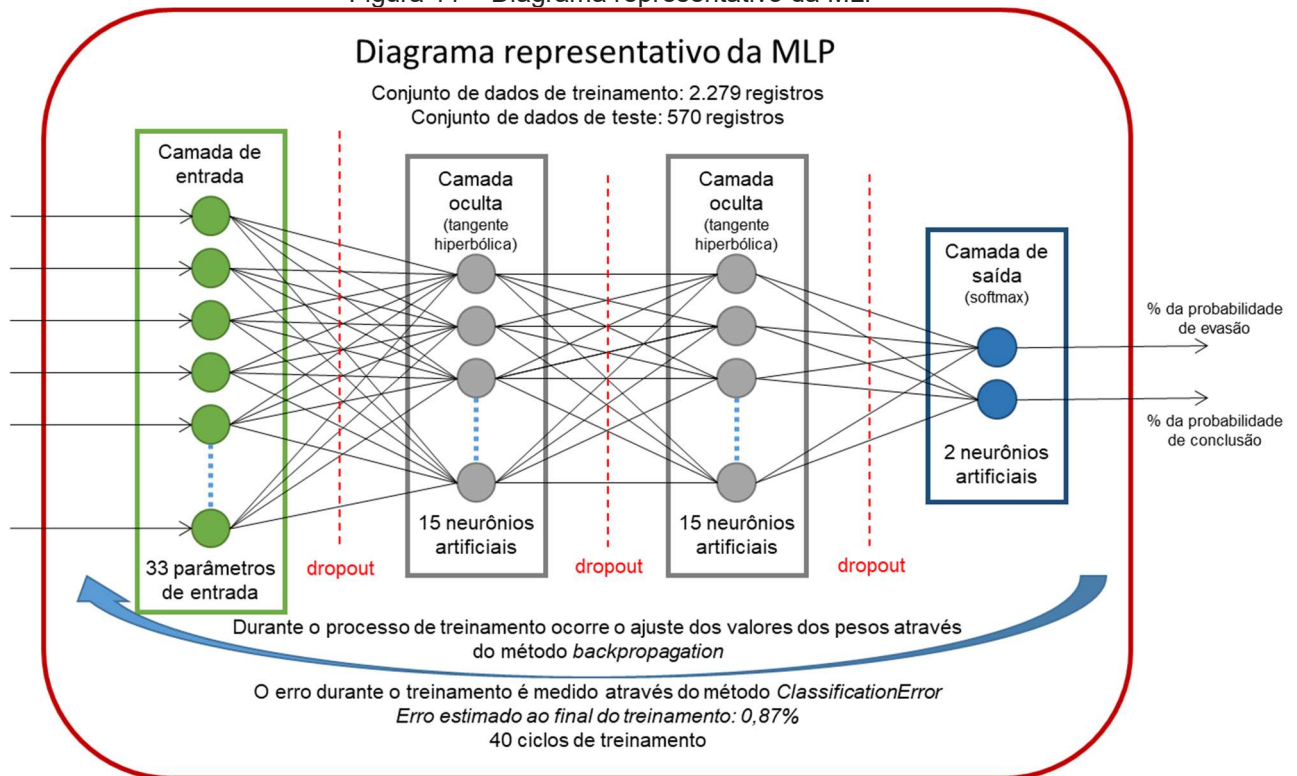
#####
#
# Test command (test action)
#
#####
evalNodeNames are not specified, using all the default evalnodes and training criterion nodes.
Final Results: Minibatch[1-1]: err = 0.00877193 * 570; ce = 0.03937965 * 570; perplexity = 1.04016531
```

Fonte: O autor (2018).

A Figura 11, apresenta um diagrama resumido do processo de geração da MLP e do modelo gerado.

A validação do modelo gerado foi realizada ao se processar as cópias existentes dos bancos de dados gravadas mensalmente no período de abril a novembro de 2018. A situação prevista pelo modelo foi comparada à situação real do aluno, encontrada na cópia do banco de dados gravada em 13 de dezembro de 2018. Nessa validação foram utilizados dados referentes a 6.985 alunos. Desses, 527 alunos evadiram-se da Instituição. Foi verificado que o modelo detectou 333 dos 527 alunos que se evadiram (63,2%) nos dados analisados.

Figura 11 – Diagrama representativo da MLP



Fonte: O autor (2018).

Nessa validação, dos 6.985 alunos analisados, 4.924 tiveram sua situação identificada corretamente pelo modelo (70,5%). O modelo alertou para a tendência à evasão a respeito de 2.200 alunos (31,5%), dos alunos analisados. A Tabela 2 sumariza essas informações em relação ao total de alunos analisados e a Tabela 3 em relação aos alunos evadidos.

Tabela 2 – Resultado da validação do modelo em relação ao total de alunos analisados, no período de abril a novembro de 2018, em relação à situação do aluno em dezembro/2018

	Quantidade	Percentual
Dados de alunos analisados	6.985	100,0%
Alunos que se evadiram no período analisado	527	7,5%
Alunos evadidos corretamente apontados pelo modelo	333	4,8%
Alunos apontados pelo modelo com tendência à evasão	2.200	31,5%
Alunos não apontados pelo modelo e que não se evadiram	4.591	65,7%
Alunos apontados corretamente pelo modelo em relação à evasão ou não evasão	4.924	70,5%

Fonte: O autor (2018).

Tabela 3 – Resultado da validação do modelo em relação aos alunos evadidos, no período de abril a novembro de 2018, em relação à situação do aluno em dezembro/2018

	Quantidade	Percentual
Alunos que se evadiram no período analisado	527	100,0%
Alunos evadidos corretamente apontados pelo modelo	333	63,2%
Alunos apontados pelo modelo com tendência à evasão	2.200	417,5%

Fonte: O autor (2018).

Embora o preenchimento periódico de notas e frequências dos alunos seja obrigatório, sabe-se que na prática nem sempre isso ocorre. Muitas vezes, os dados de frequência e de notas dos alunos são preenchidos somente no final de cada semestre. De certa forma foi ousado a utilização de parâmetros de curto prazo, ou seja, os parâmetros que referenciam os últimos 60 dias. Caso não houvesse informação suficiente preenchida nesses parâmetros que referenciam os últimos 60 dias para se definir a tendência do aluno, esses parâmetros poderiam não só não auxiliar, mas até mesmo atrapalhar o modelo, no sentido de deixá-lo mais complexo sem nenhum ganho adicional. Os parâmetros de curto prazo utilizados foram: percentual médio de faltas nos últimos 60 dias, nota mínima nos últimos 60 dias, nota média nos últimos 60 dias, nota máxima nos últimos 60 dias.

Felizmente os parâmetros selecionados foram relevantes, e nota-se que esses parâmetros tiveram grande influência na definição do momento em que o aluno demonstra tendência à evasão. Como são os únicos parâmetros utilizados que são internos e variáveis ao longo do semestre, e visto que a situação em relação à tendência de evasão do aluno indicada pelo modelo também muda conforme vai se passando o semestre, pode-se dizer que possivelmente esses parâmetros têm muita influência nesse processo. Essa atuação fica mais evidente na Tabela 4, onde pode-se ver os resultados detalhados da análise dos dados gravados mensalmente para cada uma das cópias dos dados analisados de abril a novembro.

Ao se observar o atributo “Qtd. alunos identificados como tendência à evasão que realmente evadiram, identificados pela primeira vez”, na Tabela 4, onde pode-se ver com clareza quais os “novos” alunos são identificados pelo modelo mês a mês. Esse atributo da tabela destaca a importância dos parâmetros de curto prazo para o modelo e também a importância da atualização periódica dos dados de notas e frequências. Acredita-se que quanto mais atualizados estiverem esses dados de notas e frequências, melhor será a performance do modelo, ao mesmo tempo reduzindo a

Tabela 4 – Dados da validação das cópias gravadas mensalmente, nas datas referenciadas pelos nomes das colunas, em relação à situação do aluno na cópia do banco de dados gravado em 13/12/2018

Dados	30/04/2018	31/05/2018	30/06/2018	30/07/2018	31/08/2018	30/09/2018	31/10/2018	30/11/2018
Qtd. alunos analisados	6267	6254	6275	6525	6315	6106	6076	6155
Qtd. alunos evadidos	474	461	456	412	244	36	7	0
Qtd. alunos identificados como tendência à evasão	691	572	618	1208	1235	760	597	666
Qtd. alunos identificados como tendência à evasão que realmente evadiram, identificados pela primeira vez	157	4	19	134	19	0	0	0
Qtd. alunos identificados como possibilidade de evasão que realmente evadiram	157	140	145	262	165	25	2	0
Qtd. alunos identificados como não tendo tendência à evasão e que realmente não evadiram	5259	5361	5346	5167	5001	5335	5474	5489
Qtd. Alunos identificados corretamente	5416	5501	5491	5429	5166	5360	5476	5489
Qtd. Alunos identificados incorretamente (falsos positivos e falsos negativos)	851	753	784	1096	1149	746	600	666
Qtd. alunos identificados como não tendo tendência à evasão e que evadiram (falsos negativos)	317	321	311	150	79	11	5	0
Qtd. alunos identificados com tendência à evasão e que não evadiram (falsos positivos)	534	432	473	946	1070	735	595	666
% alunos identificados corretamente com tendência à evasão em relação ao número de alunos que evadiram	33,1%	30,4%	31,8%	63,6%	67,6%	69,4%	28,6%	-
% alunos identificados corretamente com tendência à evasão em relação ao total de alunos analisados	2,5%	2,2%	2,3%	4,0%	2,6%	0,4%	0,0%	0,0%
% alunos identificados corretamente em relação ao total de alunos analisados	83,9%	85,7%	85,2%	79,2%	79,2%	87,4%	90,1%	89,2%
% de acertos em relação ao total de alunos analisados	86,4%	88,0%	87,5%	83,2%	81,8%	87,8%	90,1%	89,2%
% alunos identificados incorretamente em relação ao total de alunos analisados	13,6%	12,0%	12,5%	16,8%	18,2%	12,2%	9,9%	10,8%
% alunos identificados incorretamente com tendência à evasão em relação ao total de alunos analisados	5,1%	5,1%	5,0%	2,3%	1,3%	0,2%	0,1%	0,0%
% alunos identificados incorretamente como não evadidos em relação ao total de alunos analisados	8,5%	6,9%	7,5%	14,5%	16,9%	12,0%	9,8%	10,8%
Média da quantidade de dias da primeira detecção até a evasão:				34,6				
Desvio-padrão da média da quantidade de dias da primeira detecção até a evasão:				23,17				
Qtd. alunos identificados que se evadiram no período:				176				

Fonte: O autor (2019).

quantidade de alunos indicados como em tendência à evasão e aumentando o percentual de alunos indicados pelo modelo que realmente evadiram.

Ao se observar o atributo “Qtd. alunos identificados como tendência à evasão que realmente evadiram, identificados pela primeira vez”, na Tabela 4, onde pode-se ver com clareza quais os “novos” alunos são identificados pelo modelo mês a mês. Esse atributo da tabela destaca a importância dos parâmetros de curto prazo para o modelo e também a importância da atualização periódica dos dados de notas e frequências. Acredita-se que quanto mais atualizados estiverem esses dados de notas e frequências, melhor será a performance do modelo, ao mesmo tempo reduzindo a quantidade de alunos indicados como em tendência à evasão e aumentando o percentual de alunos indicados pelo modelo que realmente evadiram.

É importante observar que o número de alunos marcados pelo sistema como em tendência de evasão em relação ao número de alunos que realmente evadiram, foi quatro vezes maior (417,5% maior), como apresentado na Tabela 3. Esse valor ressalta dois problemas importantes do modelo criado: o caso dos falsos negativos e o caso dos falsos positivos.

No caso dos falsos negativos (37% dos alunos que evadiram), ou seja, dos alunos em tendência real de evasão e que não foram identificados pelo modelo, pode-se dizer que nem todos os indícios da tendência à evasão vão estar caracterizados nos dados do histórico escolar do aluno. A evasão é uma situação complexa que possui muitas causas e muitos motivos, e o modelo embora possa proporcionar um norteamento a ser trabalhado, não deve ser considerado como a única forma de avaliação da tendência à evasão. A instituição deve focar esforços em diversas frentes, e através de diversos métodos para identificar e combater esse fenômeno da evasão.

No caso dos falsos positivos identificados, ou seja, dos alunos que não evadiram, porém foram identificados pelo modelo como em tendência à evasão. Nesse caso duas situações podem ter ocorrido: de alunos que estavam em tendência real à evasão e por algum motivo essa tendência foi revertida, e alunos que não estavam em tendência à evasão e foram identificados incorretamente pelo modelo. A princípio acredita-se que não é um problema de forma direta a utilização do modelo gerado, porém pode ser um problema de forma indireta. Acredita-se que no caso dos alunos que não estavam em tendência à evasão, mas foram identificados pelo modelo, ao serem abordados por agentes da Instituição, como coordenadores de

curso, possivelmente não haverá prejuízo. Sempre há espaço para melhorias e um acompanhamento mais próximo da situação do aluno possivelmente melhorará ou manterá da mesma forma o seu rendimento escolar. Tontini e Walter ainda salientam que durante a abordagem realizada pelos coordenadores de curso a “receptividade no que tange a essa ação foi positiva, visto que muitos alunos agradeceram e se mostraram felizes pela preocupação demonstrada pela universidade” (TONTINI e Walter, 2014, p. 105).

O prejuízo é mais no sentido indireto, pois como foram identificados pelo modelo mais de quatro vezes (417,5%) a quantidade de alunos que realmente evadiram-se da Instituição, pode ocorrer de não ser viável para os agentes da Instituição, como os coordenadores de curso, atender a todos esses alunos, e dessa forma terem que priorizar o atendimento de alguns em detrimento de outros. É possível que nessa priorização seja selecionado um aluno que não está em tendência de evasão, em detrimento de atender um aluno que realmente está em tendência à evasão. Seria muito interessante se o número de alunos identificados pelo modelo fosse menor.

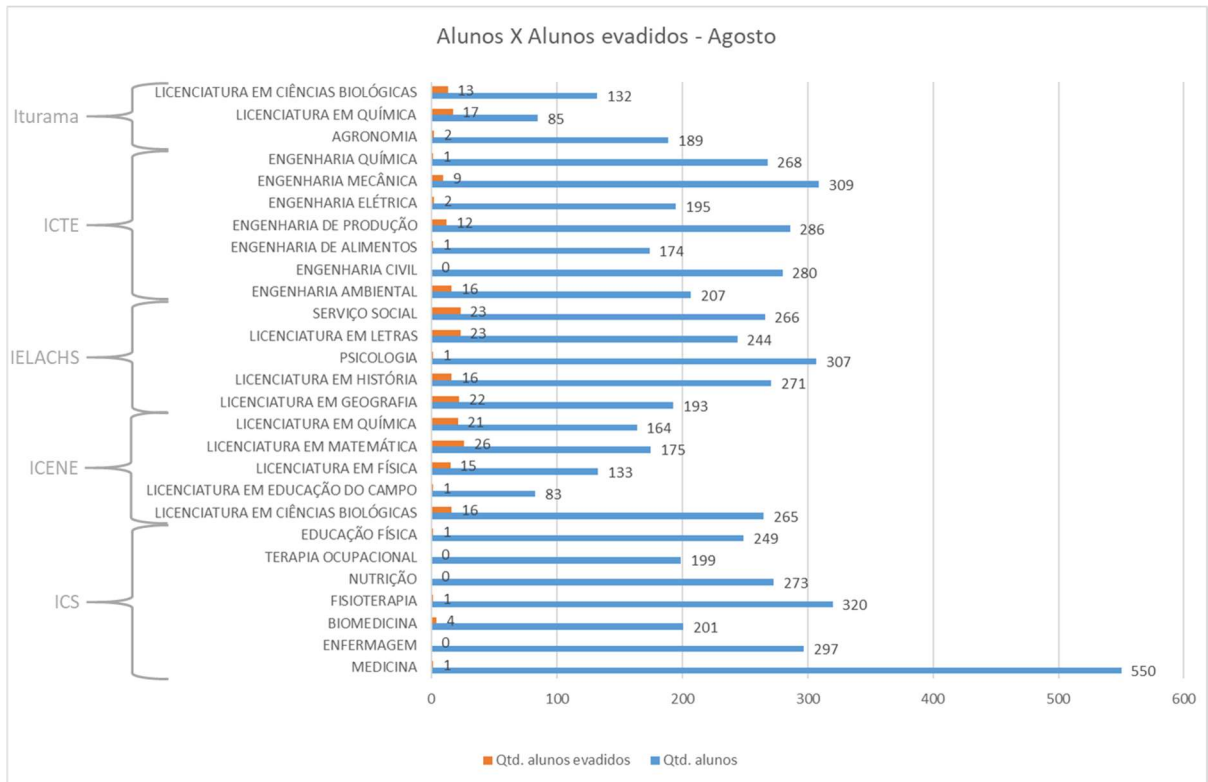
É importante destacar, que embora o percentual de falsos positivos seja alto, os casos de falsos negativos são muito mais graves que os falsos positivos, pois no caso de alunos não identificados com em tendência à evasão pelo modelo, esses alunos podem ficar sem o devido acompanhamento e atenção pelos agentes da Instituição.

Para se entender melhor o problema dos falsos positivos, foram criados os gráficos das Figuras 12, 13 e 14. Para a criação desses gráficos foram utilizados como base os dados da validação do banco de dados gravado em 31 de agosto de 2018, que é o mês onde o sistema identificou um maior número de alunos com tendência à evasão, conforme pode ser visto na Tabela 4.

A Figura 12 detalha a quantidade de alunos ativos em cada curso e a quantidade de alunos evadidos em cada um deles, na relação do banco de dados gravado em 31 de agosto de 2018 para o banco de dados gravado em 13 de dezembro de 2018.

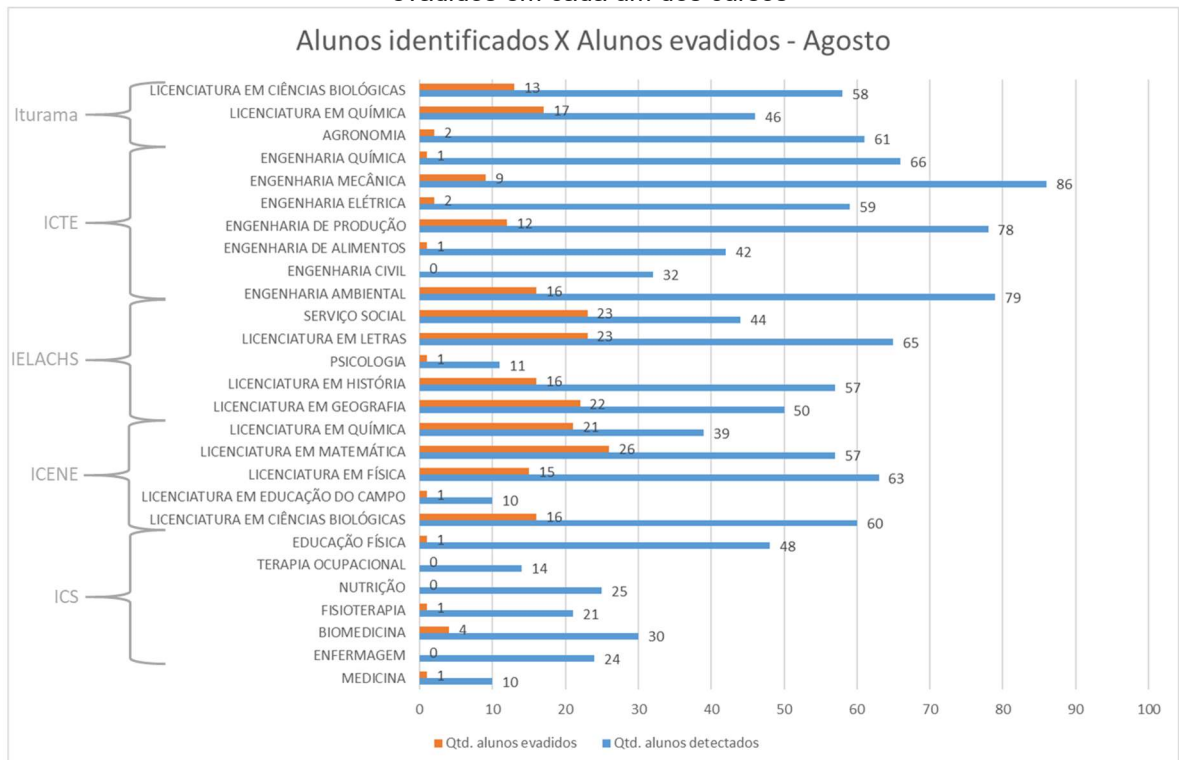
A Figura 13 detalha a quantidade de alunos identificados como em tendência à evasão pelo modelo em comparação com a quantidade de alunos evadidos em cada um deles, na relação do banco de dados gravado em 31 de agosto de 2018 para o banco de dados gravado em 13 de dezembro de 2018.

Figura 12 – Detalhamento da quantidade de alunos ativos e a quantidade de alunos evadidos em cada um dos cursos



Fonte: O autor (2018).

Figura 13 – Detalhamento da quantidade de alunos em tendência à evasão e a quantidade de alunos evadidos em cada um dos cursos

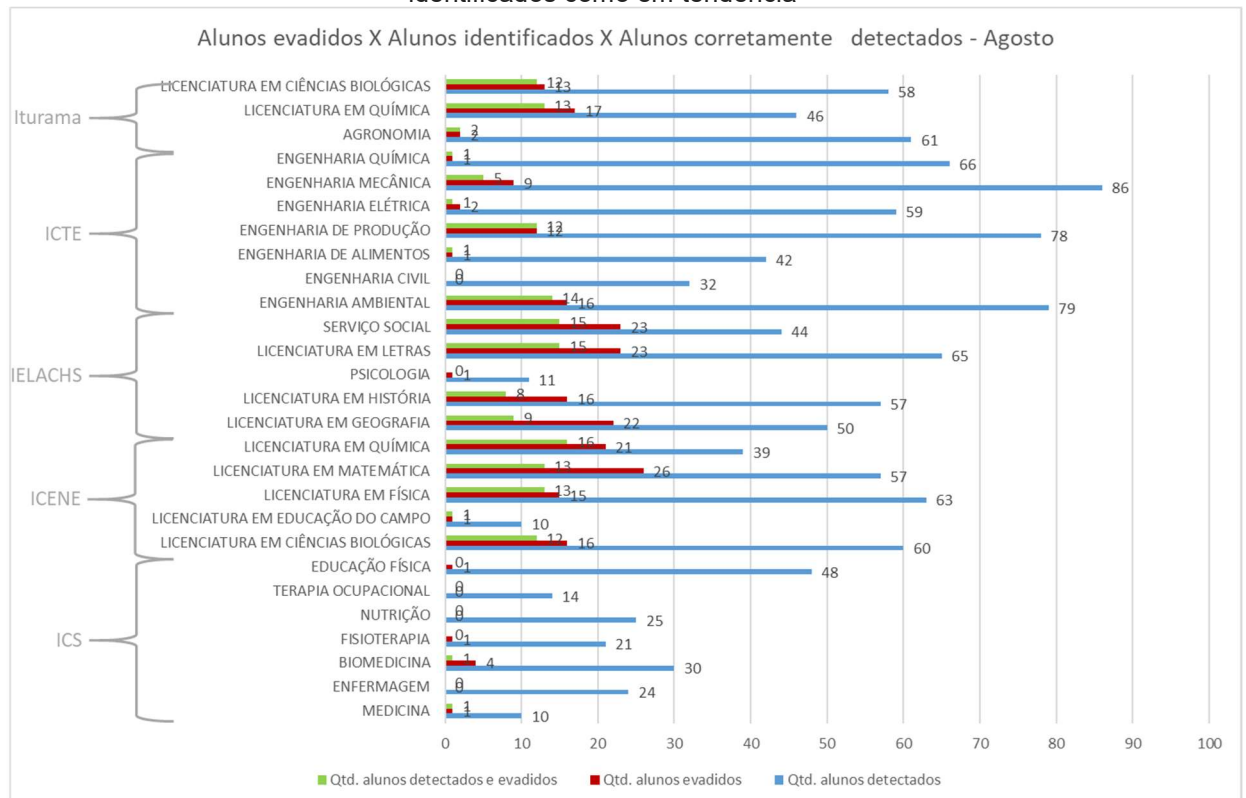


Fonte: O autor (2018).

O gráfico da Figura 14 apresenta um comparativo entre a quantidade de alunos detectados como em tendência à evasão pelo modelo, a quantidade de alunos evadidos e a quantidade de alunos corretamente detectados como em tendência à evasão em cada um dos cursos, na relação do banco de dados gravado em 31 de agosto de 2018 para o banco de dados gravado em 13 de dezembro de 2018.

Ao se analisar os gráficos das Figuras 12, 13 e 14, pode-se perceber que para alguns cursos como, por exemplo, Engenharia de Produção, Mecânica e Ambiental, o modelo identificou mais de 70 alunos por curso. Ao se analisar essa informação com foco da figura do coordenador de curso, pode-se perceber que são muitos alunos a serem contatados num único mês.

Figura 14 – Detalhamento da quantidade de alunos evadidos em comparação com a quantidade de alunos identificados como em tendência à evasão e a quantidade de alunos corretamente identificados como em tendência



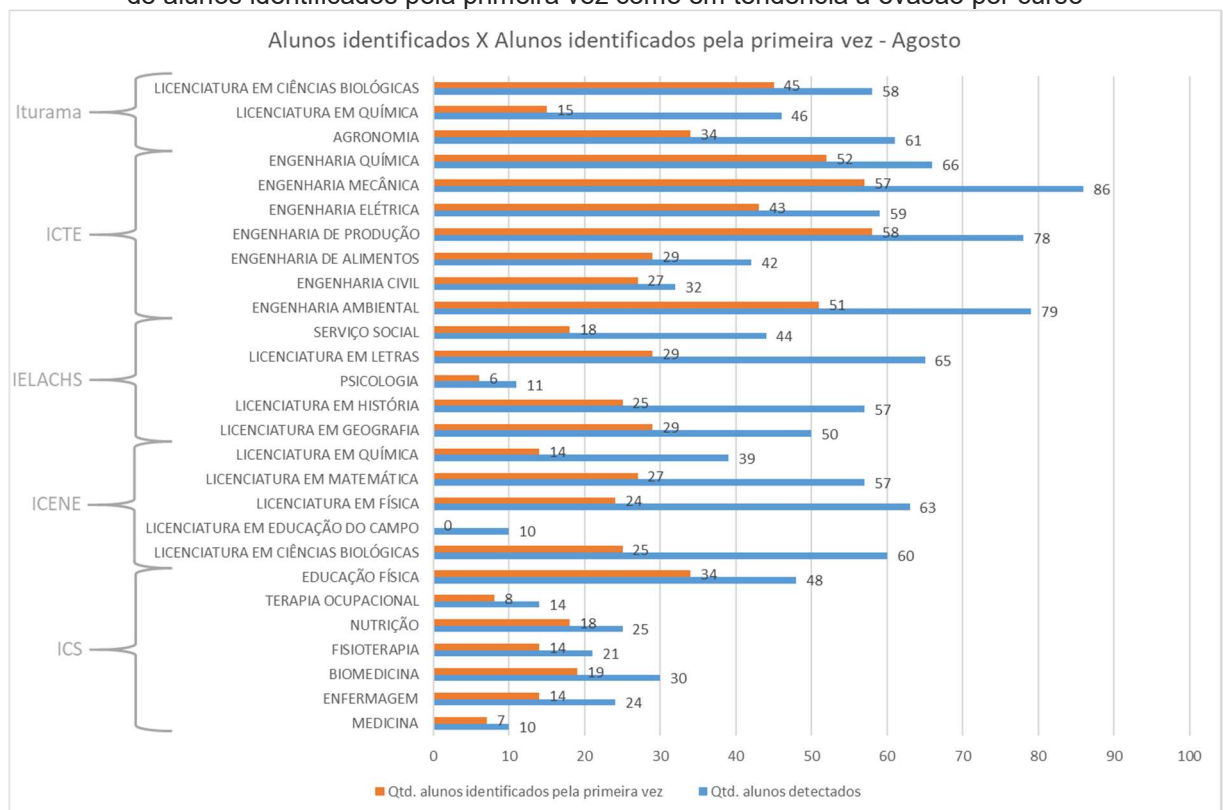
Fonte: O autor (2018).

Mesmo ao se fazer essa análise da quantidade de alunos identificados por curso, excluindo-se os alunos que já haviam sido identificados em meses anteriores, conforme pode ser visto no gráfico da Figura 14, o problema do número elevado de alunos identificados por curso ainda se mantém. Nos cursos de Engenharia de

Produção, Mecânica, Química e Ambiental, o modelo identifica mais de 50 alunos pela primeira vez por curso.

Todos esses dados apresentados (gráficos das Figuras 12, 13, 14 e 15) servem para demonstrar o problema do falso positivo, ou seja, de alunos identificados pelo modelo como em tendência à evasão, mas que não irão evadir. Esses falsos positivos podem causar uma necessidade de priorização do atendimento a alguns alunos em detrimento a outros, em especial ao se observar a figura do coordenador de curso que é uma pessoa única por curso, e que possui outras funções. Por isso é importante que a instituição disponibilize uma infraestrutura e suporte para otimizar o atendimento da maior quantidade possível de alunos.

Figura 15 – Detalhamento da quantidade de alunos identificados em comparação com a quantidade de alunos identificados pela primeira vez como em tendência à evasão por curso



Fonte: O autor (2018).

Tanto o problema dos falsos positivos, quanto o problema dos falsos negativos podem, possivelmente, ser melhorados com a utilização de outras ferramentas, utilização de outros parâmetros e até mesmo recriação do modelo quando houver maior disponibilidade de dados.

Em relação a trabalhos semelhantes, é interessante verificar Burgos et al. (2017), onde foi utilizada regressão logística em vez de RNA, que é uma outra técnica de *data mining*. Foi criado um modelo de regressão logística por disciplina e por semana de andamento do semestre, sendo um total de 20 semanas no semestre e cinco disciplinas analisadas (totalizando 100 modelos de regressão logística). Como parâmetros de entrada do modelo foram utilizados somente a nota do aluno nas atividades e o número da semana corrente em relação ao semestre. Os modelos gerados conseguiram identificar praticamente todos os alunos que se evadiram a partir da décima segunda semana do semestre (resultado próximo a 100% a partir da décima segunda semana). A partir da identificação dos alunos em risco de evasão, esses alunos foram abordados pelos tutores e eles obtiveram uma redução na evasão de 14%. Burgos et al. (2017) ajudam a demonstrar que é possível detectar a tendência à evasão dos alunos, mesmo utilizando-se outras técnicas de *data mining*, usando outros parâmetros de entrada e sendo em outro modelo de ensino (foi realizado em disciplinas à distância num ambiente virtual de aprendizado – Moodle). Mais do que isso, Burgos et al. (2018) ajudam a reforçar a importância da intervenção dos agentes da instituição, que sem a atuação junto aos alunos em tendência de evasão, a detecção dessa tendência perde o seu sentido.

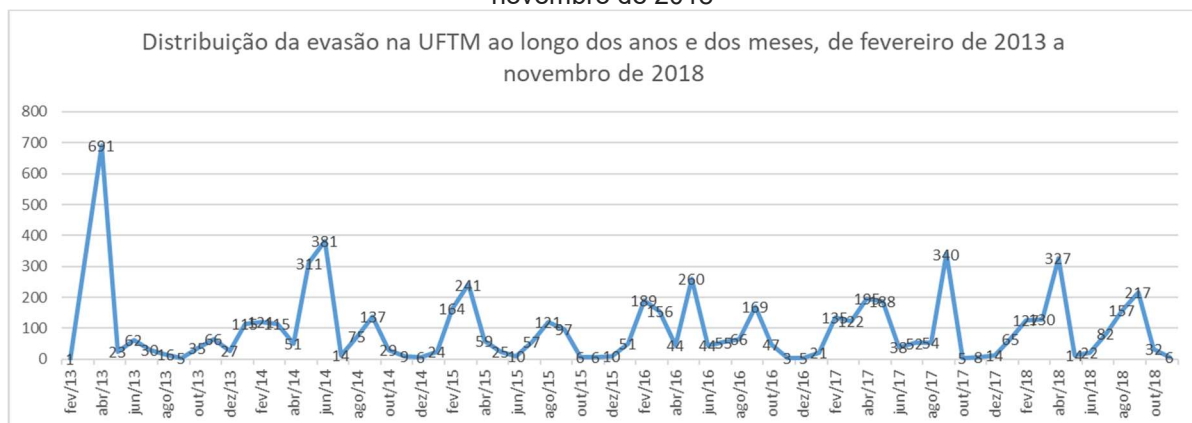
Uma outra análise que foi considerada importante para os resultados do modelo é a possibilidade de se determinar com quanto tempo de antecedência da ocorrência de evasão o modelo consegue identificar essa situação do aluno.

Para essa análise do tempo de antecedência, foi feito o processamento das cópias do banco de dados gravadas diariamente no período de setembro a novembro de 2018. Foram utilizados para validação dados referentes a 6.241 alunos. Desses, 245 alunos evadiram-se da Instituição. Ao se processar os dados gravados no dia 31 de agosto de 2018, foram encontrados 165 alunos marcados pelo modelo como em tendência de evasão, que realmente se evadiram em relação aos dados gravados de dezembro. Ao se processar os dados gravados no dia 01 setembro de 2018, foram encontrados 165 marcados pelo modelo como em tendência de evasão. Nos dados gravados diariamente no período de setembro a novembro, o modelo foi capaz de identificar corretamente pela primeira vez oito novos alunos. Esses oito alunos foram identificados pelo modelo na média com 8,7 dias, e desvio-padrão da média de 5,9 dias. Devido ao baixo número de amostras (somente 8 amostras), entende que o grau de confiança da encontrada média não está em bons patamares.

Durante a análise do processamento das cópias gravadas diariamente no período de setembro a novembro, percebeu-se que o número de alunos que se evadiram nesse período era pequeno, principalmente quando se comparado a outros meses do ano. Os dados detalhados dessa análise estão disponíveis em arquivo separado devido ao tamanho da tabela (tabela com 93 colunas e 21 linhas). Ao se analisar a quantidade de alunos que o modelo identificou como em tendência à evasão na Tabela 4, observou-se que havia uma grande diferença de alguns meses para outros.

Buscou-se então dados que indicassem se poderia existir ou não uma certa sazonalidade na evasão. Para a verificação da existência de sazonalidade na evasão foi gerado um gráfico da distribuição da evasão na UFTM ao longo dos anos e dos meses, de fevereiro de 2013 a novembro de 2018, apresentado na Figura 16.

Figura 16 – Distribuição da evasão na UFTM ao longo dos anos e dos meses, de fevereiro de 2013 a novembro de 2018



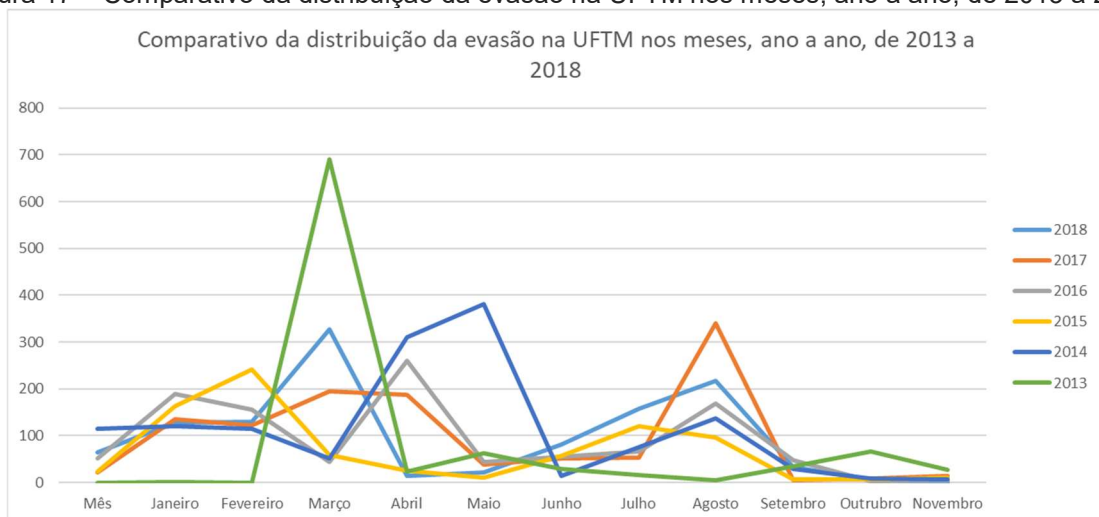
Fonte: O autor (2018) / Dados de UFTM (2018).

Com base no gráfico da Figura 16, há uma indicação de existência de sazonalidade, principalmente ao notar áreas de alta evasão próximas aos meses de abril, maio e agosto. Para tentar melhorar a análise dessa sazonalidade foram criados dois novos gráficos: o gráfico do comparativo da distribuição da evasão na UFTM nos meses, ano a ano, de 2013 a 2018, apresentado na Figura 17; o gráfico do comparativo da distribuição da evasão na UFTM em meses desde o início do semestre, ano/semestre a ano/semestre, de 2013 a 2018, apresentado na Figura 18.

Ao se observar o gráfico da Figura 17, entendeu-se que existe um certo coeficiente de sazonalidade na evasão, mais destacada nos meses de março, abril e agosto, mas que ele não é totalmente definido nesses meses ao longo de todos os

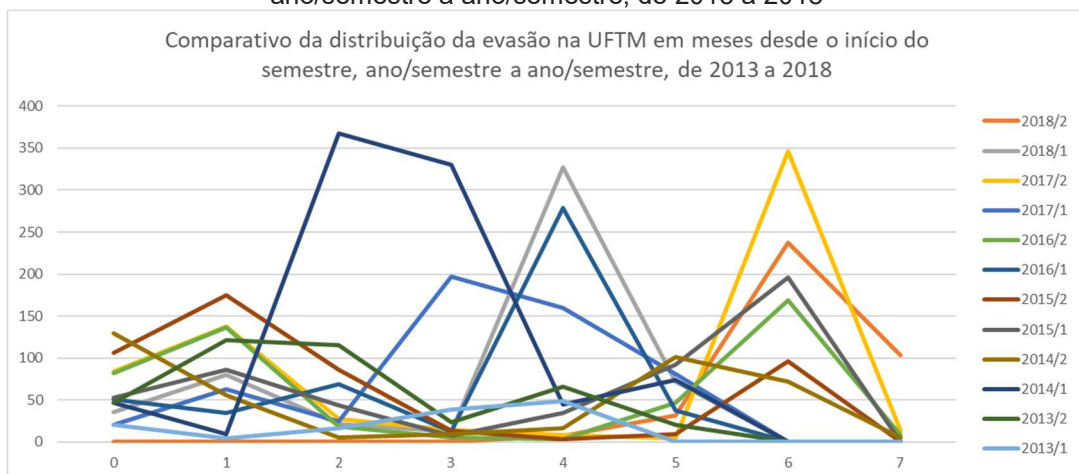
anos. É sugestiva a relação dessa sazonalidade na evasão em relação à dependência do calendário acadêmico. Com base nesse indício, foi construído o gráfico da Figura 18. Nesse gráfico, foi removida a variação do calendário acadêmico e feita a distribuição por meses desde o início do semestre, aumentando a evidência em relação a essa provável sazonalidade relacionada ao calendário escolar. Observa-se alguns pontos mais altos no gráfico da Figura 18, em alguns anos no segundo e terceiro mês, em alguns anos no quarto e em outros no sexto mês do semestre. Há um certo movimento de evasão não tão pronunciado no primeiro mês. Esse movimento no primeiro mês provavelmente possui forte influência dos alunos ingressantes que se matriculam e acabam evadindo da instituição por terem sido selecionados também em outra instituição.

Figura 17 – Comparativo da distribuição da evasão na UFTM nos meses, ano a ano, de 2013 a 2018



Fonte: O autor (2018) / Dados de UFTM (2018).

Figura 18 – Comparativo da distribuição da evasão na UFTM em meses desde o início do semestre, ano/semestre a ano/semestre, de 2013 a 2018



Fonte: O autor (2018) / Dados de UFTM (2018).

Buscou-se então outros estudos que tenham encontrado resultados similares a respeito da sazonalidade da evasão. Alguns trabalhos encontrados (TINTO, 1993 apud SILVA, 2015; MACHADO et al., 2015) estabelecem uma relação de sazonalidade da evasão, porém uma sazonalidade onde o aluno tem maior tendência à evasão no primeiro ano e no início do segundo ano. Foi encontrada indicação da sazonalidade anual, nos documentos entregues à Comissão de Valores Mobiliários (CVM) por Kroton Educacional (2017; 2018), onde é destacado que há uma ocorrência de maior evasão no segundo e quarto trimestres. Ao se comparar a informação de evasão de Kroton Educacional (2017; 2018) com os dados obtidos na Figura 13, pode-se notar uma certa similaridade, onde ocorre alto índice de evasão no segundo trimestre (março e abril) e no quarto trimestre (agosto). Possivelmente esse evento de sazonalidade deve ter sua ocorrência de forma generalizada, e deve ocorrer tanto em instituições de ensino públicas, quanto privadas. Mais estudos, em especial envolvendo uma maior abrangência de instituições, podem estabelecer melhores conclusões sobre o assunto da sazonalidade anual da evasão.

Como na análise das bases de dados gravadas diariamente de setembro a novembro não se obteve um bom conjunto amostral (somente oito novos alunos identificados e que evadiram), não foi possível determinar o tempo médio de antecedência à evasão que o modelo consegue indicar essa tendência. Decidiu-se então por utilizar os dados disponíveis das cópias do banco de dados gravadas em agosto. Foram utilizados dados referentes a 6.594 alunos, no processamento das cópias do banco de dados gravadas em agosto utilizados para validação. Desses, 415 alunos evadiram da instituição. Ao se processar os dados gravados no dia 30 de julho de 2018, foram encontrados 262 alunos marcados pelo modelo como em tendência de evasão, que realmente se evadiram em relação aos dados gravados de dezembro. Ao se processar os dados gravados no dia 1º agosto de 2018, foram encontrados 250 marcados pelo modelo como em tendência de evasão. Nos dados gravados no mês de agosto o modelo foi capaz de identificar corretamente pela primeira vez 25 novos alunos. Esses 25 alunos foram identificados pelo modelo, na média com 30 dias de antecedência da ocorrência de evasão, e desvio-padrão da média de 20,6 dias. Ao se fazer o cálculo da média com grau de confiança de 95% obteve-se uma variação da média de 21,5 a 38,5 dias ($30 \pm 8,5$ dias) de antecedência da ocorrência de evasão.

De conhecimento da possibilidade de existência de sazonalidade da evasão, entendeu-se que a análise somente dos dados gravados em agosto não seria

suficiente para se determinar um bom intervalo da média de antecedência da previsão da ocorrência de evasão. Embora os dados gravados mensalmente de abril a novembro não são a melhor opção para se fazer esse tipo de verificação, por terem sido gravados somente mensalmente, pode haver um espaço de tempo em que o modelo seria capaz de identificar alunos em tendência de evasão, mas que não é possível aferir por não haverem dados para a avaliação. Como não há outras cópias disponíveis do banco de dados do sistema, foi feita a análise da quantidade de dias de antecedência que o modelo consegue identificar a tendência à evasão com as cópias disponíveis, mesmo sabendo que não seria a melhor forma de fazer essa análise. Essa análise é importante pois serve como uma referência para a criação e organização de processos na Instituição relacionados ao uso do modelo. Por isso foi realizada a análise utilizando todos os dados disponibilizados pelo DTI, gravados no ano de 2018, no período de abril a novembro (UFTM, 2018). Para esse processamento foram utilizados para validação dados referentes a 6.987 alunos. Desses, 529 alunos evadiram-se da instituição. Ao se processar os dados gravados no dia 30 de abril de 2018, foram encontrados 157 alunos marcados pelo modelo como em tendência de evasão, que realmente se evadiram em relação aos dados gravados de dezembro. Ao se processar os dados gravados no dia 31 maio de 2018, foram encontrados 161 marcados pelo modelo como em tendência de evasão (alguns alunos identificados nos dados gravados em abril já haviam se evadido em maio, conforme apresentado na Tabela 4). Nos dados gravados no mês de maio o modelo foi capaz de identificar corretamente pela primeira vez quatro novos alunos. Nos dados disponíveis de maio a novembro foram encontrados 181 alunos identificados pela primeira vez pelo modelo como em tendência de evasão e que realmente se evadiram. Esses 181 alunos foram identificados pelo modelo, na média, com 36 dias de antecedência da ocorrência de evasão, e desvio-padrão da média de 22,4 dias. Ao se fazer o cálculo da média com grau de confiança de 95% obteve-se uma variação da média de 33 a 39 dias (36 ± 3 dias) de antecedência da ocorrência de evasão. Pode-se verificar que esse valor está coerente com o valor obtido no processamento dos dados gravados somente em agosto, descrito anteriormente.

Uma outra observação é em relação à baixa quantidade de amostras detectadas no processo de validação utilizando-se os bancos de dados gravados diariamente de setembro a novembro. Como o intervalo médio de detecção da tendência à evasão do modelo encontrado é de 33 a 39 dias, possivelmente muitos

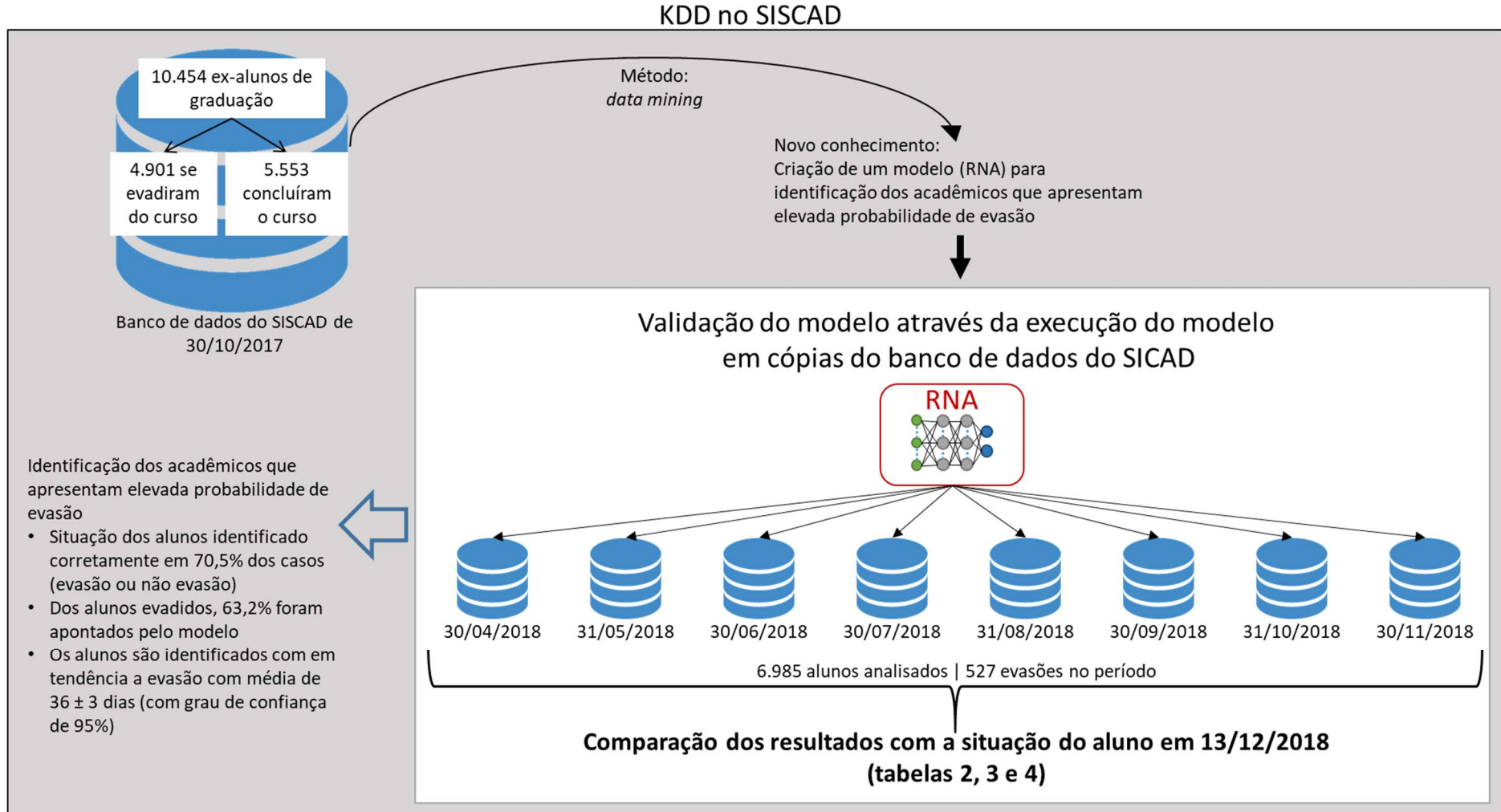
dos alunos que se evadiram em setembro teriam sido detectados pelo modelo em julho ou agosto, se ele tivesse sido executado diariamente. Nos meses de outubro e novembro, e até mesmo em setembro, não deveria haver muitos alunos em tendência à evasão, em decorrência da possível sazonalidade da evasão, conforme aparenta ocorrer, e é apresentado na Figura 18. De certa forma essa baixa amostragem obtida nesse processo de validação (setembro a novembro) está de acordo com os demais dados obtidos.

A Figura 19 apresenta um diagrama resumindo os resultados obtidos.

Para possibilitar que agentes da Instituição, como coordenadores de curso, servidores da PROACE da UFTM, ou outros agentes da Instituição, possam entrar em contato com os alunos, é necessário haver alguma forma de identificar esses alunos com tendência à evasão de forma prática. Para esse trabalho foi proposto um relatório, que poderá ficar disponível diretamente no SISCAD, que pode analisar a situação dos alunos em tempo real e gerar a lista dos alunos identificados como em tendência à evasão. Esse relatório seria tal como o apresentado na Figura 20. Propõe-se que além dos dados básicos do aluno, sejam apresentados também os dados utilizados pelo modelo para a análise da evasão, dessa forma permitindo esses agentes que estiverem acessando o relatório já tenham um resumo da situação geral do aluno. Nesse relatório serão listados somente os alunos com tendência à evasão.

O SISCAD, como um sistema de gestão acadêmica, permite que seja limitado o acesso de determinadas funcionalidades a determinados perfis. Dessa forma, esse relatório poderá ficar acessível somente aos perfis de agentes da Instituição que a Direção da Instituição julgar pertinente e permitir o acesso, dessa forma preservando a privacidade dos alunos.

Figura 19 – Diagrama representando um resumo dos resultados obtidos



Fonte: O autor (2018).

Figura 20 – Modelo de relatório proposto listando os alunos identificados com tendência à evasão

RGA 0000.0000.0	Nome Aaaaaaaa Aaaa Aaaaaaaa	Idade 27
Curso ENGENHARIA DE ALIMENTOS	Instituto INSTITUTO DE CIÊNCIAS TECNOLÓGICAS E EXATAS	Cidade Uberaba
Tipo de curso Bacharelado	Área do conhecimento Engenharias	Tipo de ingresso VESTIBULAR
Ação afirmativa AMPLA - Ampla Concorrência		
Qtd. de trancamentos 2	Qtd. reprovações 51	Máximo de reprovações na mesma disciplina 4
% médio de faltas 17.04	Nota média 4.56	Qtd. reprovações por falta 16
Quantidade de disciplinas do último semestre	Quantidade de reprovações no último semestre	Média da quantidade de disciplinas por semestre 7.2
Carga horária de disciplinas do último semestre	% médio de faltas no último semestre 0.00	Média da carga horária de disciplinas por semestre 469.6
% médio de falta nos últimos 60 dias 0.00	Nota mínima nos últimos 60 dias 10.00	Quantidade de reprovações por falta no último semestre
		Quantidade de reprovações por nota no último semestre
		Nota média no último semestre 0.00
		Nota média nos últimos 60 dias 10.00
		Nota máxima nos últimos 60 dias 10.00
RGA 1111.1111.1	Nome Bbbbbbbb Bbbb Bbbbbbbb	Idade 26
Curso ENGENHARIA QUÍMICA	Instituto INSTITUTO DE CIÊNCIAS TECNOLÓGICAS E EXATAS	Cidade Uberaba
Tipo de curso Bacharelado	Área do conhecimento Engenharias	Tipo de ingresso VESTIBULAR
Ação afirmativa AMPLA - Ampla Concorrência		
Qtd. de trancamentos 3	Qtd. reprovações 9	Máximo de reprovações na mesma disciplina 2
% médio de faltas 6.44	Nota média 7.06	Qtd. reprovações por falta 1
Quantidade de disciplinas do último semestre	Quantidade de reprovações no último semestre	Média da quantidade de disciplinas por semestre 7.3
Carga horária de disciplinas do último semestre	% médio de faltas no último semestre 0.00	Média da carga horária de disciplinas por semestre 450.0
% médio de falta nos últimos 60 dias 0.00	Nota mínima nos últimos 60 dias 10.00	Quantidade de reprovações por falta no último semestre
		Quantidade de reprovações por nota no último semestre
		Nota média no último semestre 0.00
		Nota média nos últimos 60 dias 10.00
		Nota máxima nos últimos 60 dias 10.00
RGA 2222.2222.2	Nome Cccccccc Cccc Cccccccc	Idade 49
Curso BIOMEDICINA	Instituto INSTITUTO DE CIÊNCIAS DA SAÚDE	Cidade Uberaba
Tipo de curso Bacharelado	Área do conhecimento Outros	Tipo de ingresso INGRESSO VIA SISU

5 CONCLUSÃO E SUGESTÃO DE TRABALHOS FUTUROS

5.1 CONCLUSÃO

Durante a realização do presente trabalho, foram analisados os dados já existentes de registro e controle acadêmico contidos no SISCAD. Com esses dados foi possível gerar um modelo, utilizando-se redes neurais artificiais, que é capaz de identificar quais os alunos estão em tendência à evasão. Baseado nos resultados obtidos, é possível concluir que:

- É possível definir quais os alunos estão em tendência à evasão com base nos dados já existentes de registro e controle acadêmico;
- O modelo foi capaz de identificar 63,8% dos alunos que evadiram;
- O modelo foi capaz de identificar corretamente a situação de 70,5% dos alunos (entre evadidos e não evadidos);
- Foi possível definir que o modelo consegue identificar os alunos em tendência à evasão com 36 dias de antecedência da ocorrência da evasão, na média (± 3 dias com grau de confiança de 95%);
- Foi demonstrado que os parâmetros de curto prazo utilizados tiveram uma influência positiva em determinar o momento provável em que o aluno vai se evadir;
- Esse trabalho também reforça a necessidade do preenchimento periódico dos dados de notas e frequências dos alunos no SISCAD;
- Foi proposto um mecanismo de geração de alertas, através da geração de um relatório que pode ser disponibilizado diretamente na ferramenta de gestão acadêmica (SISCAD);
- Esse relatório poderá ser exibido para qualquer agente da UFTM que tenha acesso, a fim de se identificar quais os alunos em tendência à evasão;
- Foi criada então uma ferramenta, através do modelo gerado e do relatório de alunos em tendência à evasão, para auxiliar na redução da evasão acadêmica do ensino superior na UFTM;

- O mesmo método aplicado nesse trabalho pode ser executado em outras Instituições de Ensino Superior, possivelmente com resultados similares.

5.2 SUGESTÕES DE TRABALHOS FUTUROS

Algumas sugestões de trabalhos futuros:

- Atualização do modelo quando da disponibilidade de mais dados a fim de se melhorar a taxa de erro;
- Estudo da importância de cada um dos parâmetros individuais, em especial dos parâmetros de curto prazo na definição da tendência à evasão e do momento da ocorrência da evasão;
- Criação de um modelo de rede neural somente com os dados dos alunos que possuírem todos os dados completos, sem execução da correção de dados citada no processo de limpeza dos dados descrito no item 3.4.
- Criação de um modelo de rede neural do tipo *Long short-term memory* a fim de se analisar os dados dos alunos semestre a semestre. Possivelmente uma rede neural desse tipo analisando o histórico completo do aluno do ingresso ao período atual poderá ter uma taxa de acerto melhor em relação a definição da tendência à evasão e do momento da ocorrência da evasão;

REFERÊNCIAS

AGGARWAL, Charu C. **Data mining**. Cham: Springer International Publishing, 2015.

BAGGI, Cristiane Aparecida dos Santos; LOPES, Doraci Alves. Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. **Avaliação**, Campinas; Sorocaba, SP, v. 16, n. 2, p. 355-374, 2011. Disponível em: <<http://www.scielo.br/pdf/aval/v16n2/a07v16n2.pdf>>. Acesso em: 14 jan. 2019.

BRAMER, Max. **Principles of data mining**. 3. ed. London: Springer London, 2016.

BRASIL. Decreto nº 6.096, de 24 de abril de 2007. Institui o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais – REUNI. **Diário Oficial da União**, Brasília, DF, 25 abr. 2007. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/decreto/d6096.htm>. Acesso em: 14 jan. 2019.

_____. Lei nº 12.711, de 29 de agosto de 2012. Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências. **Diário Oficial da União**, Brasília, DF, 3 ago. 2012. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12711.htm>. Acesso em: 14 jan. 2019.

_____. Ministério da Educação – MEC. Portaria nº 315, de 4 de abril de 2018. Dispõe sobre os procedimentos de supervisão e monitoramento de instituições de educação superior integrantes do sistema federal de ensino e de cursos superiores de graduação e de pós-graduação lato sensu, nas modalidades presencial e a distância. **Diário Oficial da União**, Brasília, DF, 5 abr. 2018. Disponível em: <http://www.in.gov.br/web/guest/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/9177560/do1-2018-04-05-portaria-n-315-de-4-de-abril-de-2018-9177556>. Acesso em: 15 dez. 2018a.

_____. Ministério da Educação – MEC. Portaria nº 1.095, de 25 de outubro de 2018. Dispõe sobre a expedição e o registro de diplomas de cursos superiores de graduação no âmbito do sistema federal de ensino. **Diário Oficial da União**, Brasília, DF, 26 out. 2018. Disponível em: <http://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/47330359>. Acesso em: 15 dez. 2018b.

BUENO, José Lino Oliveira. A evasão de alunos. **Paidéia**, Ribeirão Preto, SP, n. 5, p. 9-16 ago. 1993. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-863X1993000200002&lng=pt&nrm=iso>. Acesso em: 14 jan. 2019.

BURGOS, Concepción; CAMPANARIO, María L.; PEÑA, David de la; LARA, Juan A.; LIZCANO, David; MARTÍNEZ, María A. Data mining for modeling students' performance: a tutoring action plan to prevent academic dropout. **Computers & Electrical Engineering**, 18 mar. 2017. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0045790617305220>>. Acesso em: 14 jan. 2019.

CNPQ. **Tabela de áreas do conhecimento**. Disponível em: <<http://www.cnpq.br/documents/10157/186158/TabeladeAreasdoConhecimento.pdf>>. Acesso em: 14 jan. 2019.

COMISSÃO DE PLANEJAMENTO AVALIAÇÃO. **Grupo de trabalho indicadores**. 2015. Disponível em: <http://www.uff.br/sites/default/files/indicadores_do_forplad.pdf>. Acesso em: 14 jan. 2019.

COMISSÃO ESPECIAL DE ESTUDOS SOBRE A EVASÃO NAS UNIVERSIDADES PÚBLICAS BRASILEIRAS – ANDIFES/ABRUEM/SESu/MEC. **Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas**. 1996. Disponível em: <http://www.andifes.org.br/wp-content/files_flutter/Diplomacao_Retencao_Evasao_Graduacao_em_IES_Publicas-1996.pdf>. Acesso em: 14 jan. 2019.

DAHAL, Paras. **Classification and loss evaluation: softmax and cross entropy loss**. DeepNotes. Disponível em: <<https://deepnotes.io/softmax-crossentropy/>>. Acesso em: 14 jan. 2019.

DUCATI, Erves; ALBERTON, LUIZ. **Evasão, custos e avaliação de desempenho: caso da UFSC**. 2016. Disponível em: <https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/172074/OK-101_00519-OK.pdf?sequence=1&isAllowed=y>. Acesso em: 14 jan. 2019.

GILIOLI, Renato de Sousa Porto. **Evasão em instituições federais de ensino superior no Brasil: expansão da rede, SISU e desafios**. 2016. Disponível em: <http://www2.camara.leg.br/documentos-e-pesquisa/publicacoes/estnottec/areas-da-conle/tema11/2016_7371_evasao-em-instituicoes-de-ensino-superior_renato-gilioli>. Acesso em: 14 jan. 2019.

GÓMEZ, Magela Reny Fonticiella; TORRES, Julio Cesar. Discutindo o acesso e a permanência no ensino superior no contexto do SISU (Sistema de Seleção Unificada). **Organizações e Democracia**, Marília, SP, n. 1, v. 16, 2015. Disponível em: <<http://www2.marilia.unesp.br/revistas/index.php/orgdemo/article/view/5162>>. Acesso em: 14 jan. 2019.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA – INEP. **Sinopse estatística da educação superior**. Brasília: Inep, 2017. Disponível em: <<http://portal.inep.gov.br/web/guest/sinopses-estatisticas-da-educacao-superior>>. Acesso em: 14 jan. 2019.

KASSAI, José Roberto; Oliveira, Amanda Yamashiro Campos de; Silva, Carlos Eduardo Alves da; Carvalho, Luiz Nelson Guedes de. Reflexões sobre o nível de evasão e o custo anual per capita das unidades de ensino da USP com base no método *Inquired Balance Sheet*. **XXXIV Encontro da ANPAD**, Rio de Janeiro, RJ, 25 a 29 de setembro de 2010. Disponível em: <<http://www.anpad.org.br/admin/pdf/epq1533.pdf>>. Acesso em: 14 jan. 2019.

KATZ, Guy; BARRETT, Clark; DILL, David; JULIAN, Kyle; KOCHENDERFER, Mykel. **Reluplex**: an efficient SMT solver for verifying deep neural networks. 3 fev. 2017. Disponível em: <<https://arxiv.org/pdf/1702.01135.pdf>>. Acesso em: 14 jan. 2019.

KROTON EDUCACIONAL S.A. **Formulário de Referência 2017 – Documentos Entregues à CVM**. [S.l.: s.n.], 2017. Disponível em: <http://ri.kroton.com.br/wp-content/uploads/sites/44/2018/01/Kroton_FRE_2017_v11.pdf>. Acesso em: 14 jan. 2019.

_____. **Formulário de Referência 2018 – Documentos Entregues à CVM**. [S.l.: s.n.], 2018. Disponível em: <https://s3.amazonaws.com/mz-filemanager/e1110a12-6e58-4cb0-be24-ed1d5f18049a/289489cb-f726-4900-9ac1-ceb5b4585188_KROTON_FRE_2018_V9.pdf>. Acesso em: 14 jan. 2019.

KUMAR, Pardeep; SEHGAL, Vivek Kumar; CHAUHAN, Durg Singh. A benchmark to select data mining based classification algorithms for business intelligence and decision support systems. **International Journal of Data Mining & Knowledge Management Process (IJDKP)**, v. 22, n. 55, 2012. Disponível em: <<https://arxiv.org/ftp/arxiv/papers/1210/1210.3139.pdf>>. Acesso em: 14 jan. 2019.

MACHADO, Roger Douglas; NARA, Elpidio; SCHREIBER, Jacques Nelson Corleta; SCHWINGEL, Guilherme Augusto. **Estudo bibliométrico em mineração de dados e evasão escolar**. Research Gate, 2015. Disponível em: <https://www.researchgate.net/publication/281034791_ESTUDO_BIBLIOMETRICO_EM_MINERACAO_DE_DADOS_E_EVASAO_ESCOLAR>. Acesso em: 14 jan. 2019.

MCCAFFREY, James. Machine learning: exploring the microsoft CNTK machine learning tool. **Microsoft Magazine**, jan. 2017. Disponível em: <<https://msdn.microsoft.com/en-us/magazine/mt791798.aspx>>. Acesso em: 14 jan. 2019.

MCCULLOCH, Warren S.; PITTS, Walter H. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, p. 115-133, 1943. Disponível em: <<http://www.cse.chalmers.se/~coquand/AUTOMATA/mcp.pdf>>. Acesso em: 14 jan. 2019.

MICROSOFT. **Microsoft cognitive toolkit**. Disponível em: <<https://www.microsoft.com/en-us/cognitive-toolkit/>>. Acesso em: 14 jan. 2019a.

_____. **Loss functions and metrics**. Disponível em: <<https://docs.microsoft.com/en-us/cognitive-toolkit/Loss-Functions-and-Metrics>>. Acesso em: 14 jan. 2019b.

NG, Andrew. **The problem of overfitting**: regularization. Coursera. Disponível em: <<https://www.coursera.org/lecture/machine-learning/the-problem-of-overfitting-ACpTQ>>. Acesso em: 14 jan. 2019.

NIKAM, Sagar S. A comparative study of classification techniques in data mining algorithms. **Orient.J. Comp. Sci. and Technol**, v. 8, n. 1, p. 13-19, 2015. Disponível em: <<http://www.computerscijournal.org/?p=1592>>. Acesso em: 14 jan. 2019.

OPREA, Cristina; POPESCU, Delia Mioara; PETRESCU, Anca Gabriela; BARBU, Irina. A data mining based model to improve university management: EBSCOhost. **Journal of Science and Arts**, ano 17, v. 2, n. 39, p. 285-296, 2017. Disponível em: <http://www.icstm.ro/DOCS/josa/josa_2017_2/a_09_Oprea.pdf>. Acesso em: 14 jan. 2019.

ORACLE. **Data mining concepts**: contents. Disponível em: <https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/toc.htm>. Acesso em: 14 jan. 2019.

POPELKA, Ondřej; HŘEBÍČEK, Jiří; ŠTENCL, Michael; HODINKA, Michael; TRENZ, Oldřich. Comparison of different non-statistical classification methods. **30th International Conference Mathematical Methods in Economics**, p. 727-732, 2012. Disponível em: <http://mme2012.opf.slu.cz/proceedings/pdf/125_Popelka.pdf>. Acesso em: 14 jan. 2019.

RASCHKA, Sebastian. **What is the role of the activation function in a neural network? How does this function in a human neural network system?** Quora, 2017. Disponível em: <<https://www.quora.com/What-is-the-role-of-the-activation-function-in-a-neural-network-How-does-this-function-in-a-human-neural-network-system/answer/Sebastian-Raschka-1>>. Acesso em: 14 jan. 2019.

ROMERO, Cristóbal; VENTURA, Sebastián; ESPEJO, Pedro G.; MARTÍNEZ, César Hervás. Data mining algorithms to classify students. Educational data mining 2008. **The 1st International Conference on Educational Data Mining**, Montréal, Québec, Canada, June 20-21, 2008, Proceedings. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.217.4114&rep=rep1&type=pdf>>. Acesso em: 14 jan. 2019.

ROSENBLATT, Frank. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 19-8, 1958. Disponível em: <<http://www.ling.upenn.edu/courses/cogs501/Rosenblatt1958.pdf>>. Acesso em: 14 jan. 2019.

SILVA FILHO, Roberto Leal Lobo e; MOTEJUNAS, Paulo Roberto; HIPÓLITO, Oscar; LOBO, Maria Beatriz de Carvalho Melo. A evasão no ensino superior brasileiro. **Cadernos de Pesquisa**, v. 37, n. 132, p. 641-659, set./dez. 2007. Disponível em: <<http://www.scielo.br/pdf/cp/v37n132/a0737132.pdf>>. Acesso em: 14 jan. 2019.

SILVA FILHO, Roberto Leal Lobo e; LOBO, Maria Beatriz de Carvalho Melo. **Esclarecimentos metodológicos sobre os cálculos de evasão**. 2012. Disponível em: <http://www.institutolobo.org.br/imagens/pdf/artigos/art_078.pdf>. Acesso em: 14 jan. 2019.

SILVA, Ivan Nunes da; SPATTI, Danilo Hernane; FLAUZINO, Rogério Andrade. **Redes neurais artificiais para engenharia e ciências aplicadas**: curso prático. São Paulo: Artliber, 2010.

SILVA, Hércules Ferrari Domingues da. **Evasão na educação superior**: um estudo em uma ies privada do médio tietê. Dissertação (Mestrado em Educação) – Programa de Pós-Graduação em Educação, Pró-Reitoria Acadêmica, Universidade de Sorocaba, Sorocaba, SP, 2015. Disponível em: <http://educacao.uniso.br/prod_cientifica/alunos/2015/hercules-silva.pdf>. Acesso em: 14 jan. 2019.

SPENCE, A Michael. **Signaling in retrospect and the informational structure of markets**. 2001. Disponível em: <http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2001/spence-lecture.pdf>. Acesso em: 24 abr. 2017.

TONTINI, Géron; WALTER, Anita. Pode-se identificar a propensão e reduzir a evasão de alunos? Ações estratégicas e resultados táticos para instituições de ensino superior. **Avaliação: Revista da Avaliação da Educação Superior**, v. 19, n. 1, p. 89-110, 2014. Disponível em: <<http://www.redalyc.org/pdf/2191/219130127005.pdf>>. Acesso em: 14 jan. 2019.

UNIVERSIDADE FEDERAL DO TRIÂNGULO MINEIRO – UFTM. **Resolução nº 10**, de 27 de novembro de 2012, do Conselho Universitário da UFTM. Aprova o novo Regulamento dos Cursos de Graduação da UFTM. Uberaba/MG: [s.n.], 2012. Disponível em: <<http://sistemas.uftm.edu.br/integrado/?to=N29zTFVkdGh2bjcyeC9odGFISIRIRGthNjZ1VWY5Z1N1b1FtdTJLUnFmbDdkU0V1YzVvZEtjbkZhTyt2UFBaeXRFSnpFbEMweitJNWV6NXR3RWZBVGE2T2dYMityc3JqbVp5UitkT3Z4LzFiNFNtNHdwU2ZNRTQ0R3RCVURjenluR0hnVzE4Ynd2T0psYkdwZFUeHRpTXBUQmVDVFNyM1FZZFM1Mzd4VHpDcXp0bkppNEYyd0VjUkFKaS9TNllt&secret=uftm>>. Acesso em: 16 dez. 2018.

_____. **SISCAD** 30 out. 2017. Base de Dados em PostgreSQL. Acesso em: 4 nov. 2017.

_____. **SISCAD** abr./dez. 2018. Base de Dados em PostgreSQL. Acesso em: 15 dez. 2018.

VAN VEEN, Fjodor. **The neural network zoo**. The Asimov Institute, 2016. Disponível em: <<http://www.asimovinstitute.org/neural-network-zoo/>>. Acesso em: 14 jan. 2019.

ZENTUT. **Data mining processes**. Disponível em: <<http://www.zentut.com/data-mining/data-mining-processes/>>. Acesso em: 14 jan. 2019.

APÊNDICE A - DADOS DE INGRESSO, MATRÍCULAS, CONCLUINTES E EVASÃO DA GRADUAÇÃO PRESENCIAL DA EDUCAÇÃO SUPERIOR

Tabela 5 – Dados de ingresso, matrículas, concluintes e evasão da graduação presencial da educação superior de 2000 a 2017 no Brasil

Ano	Ingressantes Presencial	Matriculados Presencial	Concluintes Presencial	% Evasão Brasil
2000	897.557	2.694.245	324.734	-
2001	1.206.273	3.030.754	395.988	23%
2002	1.411.208	3.479.913	466.260	21%
2003	1.540.431	3.887.022	528.223	22%
2004	1.621.408	4.163.733	626.617	24%
2005	1.678.088	4.453.156	717.858	22%
2006	1.753.068	4.676.646	736.829	22%
2007	1.808.970	4.880.381	756.799	22%
2008	1.873.806	5.080.056	800.318	22%
2009	1.732.613	5.115.896	826.928	21%
2010	1.801.901	5.449.120	829.286	15%
2011	1.915.098	5.746.762	865.161	17%
2012	2.204.456	5.923.838	876.091	24%
2013	2.227.545	6.152.405	829.938	22%
2014	2.383.110	6.486.171	837.304	23%
2015	2.225.663	6.633.545	916.363	22%
2016	2.142.463	6.554.283	938.732	23%
2017	2.152.752	6.529.681	947.606	22%

Fonte: O autor (2018) / Dados INEP - Censo da Educação Superior

Tabela 6 – Dados de ingresso, matrículas, concluintes e evasão da graduação presencial pública da educação superior de 2000 a 2017 no Brasil

Ano	Ingressantes Presencial	Matriculados Presencial	Concluintes Presencial	% Evasão Brasil
2000	233.083	887.026	112.451	-
2001	274.816	939.225	132.616	14%
2002	320.354	1.051.655	151.101	9%
2003	321.689	1.136.370	169.159	10%
2004	357.979	1.178.328	202.262	15%
2005	331.365	1.192.189	195.554	12%
2006	335.767	1.209.304	183.085	12%
2007	336.223	1.240.968	193.531	12%
2008	352.615	1.273.965	187.758	12%
2009	379.134	1.351.168	187.804	11%
2010	435.710	1.461.696	178.407	12%
2011	456.635	1.595.391	194.666	11%
2012	499.370	1.715.752	202.394	13%
2013	494.940	1.777.974	206.261	15%
2014	504.627	1.821.629	225.714	16%
2015	504.038	1.823.752	224.196	17%
2016	505.002	1.867.477	231.572	15%
2017	502.621	1.879.784	238.061	16%

Fonte: O autor (2018) / Dados INEP - Censo da Educação Superior

Tabela 7 - Dados de ingresso, matrículas, concluintes e evasão da graduação presencial privada da educação superior de 2000 a 2017 no Brasil

Ano	Ingressantes Presencial	Matriculados Presencial	Concluintes Presencial	% Evasão Brasil
2000	664.474	1.807.219	212.283	-
2001	931.457	2.091.529	263.372	27%
2002	1.090.854	2.428.258	315.159	27%
2003	1.218.742	2.750.652	359.064	28%
2004	1.263.429	2.985.405	424.355	28%
2005	1.346.723	3.260.967	522.304	25%
2006	1.417.301	3.467.342	553.744	25%
2007	1.472.747	3.639.413	563.268	26%
2008	1.521.191	3.806.091	612.560	26%
2009	1.353.479	3.764.728	639.124	24%
2010	1.366.191	3.987.424	650.879	16%
2011	1.458.463	4.151.371	670.495	19%
2012	1.705.086	4.208.086	673.697	28%
2013	1.732.605	4.374.431	623.677	25%
2014	1.878.483	4.664.542	611.590	26%
2015	1.721.625	4.809.793	692.167	24%
2016	1.637.461	4.686.806	707.160	26%
2017	1.650.131	4.649.897	709.545	25%

Fonte: O autor (2018) / Dados INEP - Censo da Educação Superior

Tabela 8 - Dados de ingresso, matrículas, concluintes e evasão da graduação presencial pública da educação superior de 2000 a 2017 da região Sudeste

Ano	Ingressantes Presencial	Matriculados Presencial	Concluintes Presencial	% Evasão Sudeste
2011	155.996	533.355	68.263	-
2012	164.483	558.899	71.125	15%
2013	168.793	590.242	73.321	14%
2014	172.684	617.285	77.337	14%
2015	174.973	631.597	81.170	15%
2016	177.956	647.413	82.717	15%
2017	177.498	665.756	86.239	14%

Fonte: O autor (2018) / Dados INEP - Censo da Educação Superior

Tabela 9 - Dados de ingresso, matrículas, concluintes e evasão da graduação presencial pública da educação superior de 2000 a 2017 de Minas Gerais

Ano	Ingressantes Presencial	Matriculados Presencial	Concluintes Presencial	% Evasão Minas Gerais
2011	43.554	149.072	19.553	-
2012	50.992	157.331	17.542	18%
2013	47.952	167.300	19.806	15%
2014	49.822	182.162	22.926	10%
2015	52.597	190.076	23.674	14%
2016	52.295	193.441	23.602	15%
2017	53.131	200.299	24.561	13%

Fonte: O autor (2018) / Dados INEP - Censo da Educação Superior

Tabela 10 - Dados de ingresso, matrículas, concluintes e evasão da graduação presencial pública da educação superior de 2000 a 2017 da UFTM

Ano	Ingressantes Presencial	Matriculados Presencial	Concluintes Presencial	% Evasão Sudeste
2011	1.409	3.958	257	-
2012	1.321	4.610	334	29%
2013	1.357	5.113	499	24%
2014	1.325	5.203	492	14%
2015	1.811	5.776	595	25%
2016	2.042	6.166	653	22%
2017	2.096	6.557	753	21%

Fonte: O autor (2018).

APÊNDICE B - BREVE DESCRIÇÃO DO CURRÍCULO DO AUTOR

Possui graduação em Ciência da Computação em 2010 pelo Centro Universitário do Triângulo (UNITRI) e especialização em Gerenciamento de Projetos pela Universidade Cândido Mendes (UCAM).

Possui experiência de 16 anos no desenvolvimento de sistemas, já tendo trabalhado como programador, analista de sistemas, líder técnico em TI e consultor em TI. Já trabalhou com as seguintes linguagens de desenvolvimento: C# .Net, VB.Net e PHP. Tem experiência com os seguintes sistemas de banco de dados: MS SQL Server, Oracle, Sybase, DB2 e PostgreSQL.

Possui experiência em especial com sistemas do ramo financeiro (banco e cartões de crédito), atacadista, recursos humanos, transporte de cargas, aviação e construção.

Há três anos atua no setor de Serviço de Sistemas de Apoio ao Ensino do DTI da UFTM, sendo supervisor desse setor desde agosto de 2017. Esse setor tem como função a manutenção, evolução e criação de novos sistemas com base ao atendimento das atividades apoio ao ensino. Dentre esses sistemas, há sistemas para a gestão dos ingressantes, gestão de auxílios acadêmicos, gestão de eventos acadêmicos, gestão de projetos de pesquisa, ensino e extensão, e principalmente o sistema de gestão acadêmica (SISCAD). Dentre todos esses sistemas citados o SISCAD é o maior e mais importante deles, tendo registrado os dados dos alunos da graduação, pós-graduação e técnicos da UFTM.

Devido à destinação dos sistemas de responsabilidade do setor, tem como seus principais clientes internos a PROENS e PROPPG, porém por serem sistemas ligados a atividade fim da UFTM, acaba por atender solicitação de todas as pró-reitorias, de outros setores internos da UFTM, e até mesmo de órgãos externos a UFTM.

Maiores detalhes do currículo do autor podem ser conferidos através dos endereços:

<http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K8380167E5>

<https://www.linkedin.com/in/alan-lopes-melo-21378525/>

APÊNDICE C – DADOS COMPARATIVO DA DISTRIBUIÇÃO DA EVASÃO NA UFTM NOS MESES

Tabela 11 – Dados comparativos da distribuição da evasão na UFTM nos meses, ano a ano, de 2013 a 2018

Mês	2018	2017	2016	2015	2014	2013
Janeiro	65	21	51	24	115	0
Fevereiro	127	135	189	164	121	1
Março	130	122	156	241	115	0
Abril	327	195	44	59	51	691
Mai	14	188	260	25	311	23
Junho	22	38	44	10	381	62
Julho	82	52	55	57	14	30
Agosto	157	54	66	121	75	16
Setembro	217	340	169	97	137	5
Outubro	32	5	47	6	29	35
Novembro	6	8	3	6	9	66
Dezembro	0	14	5	10	6	27

Fonte: O autor (2018).

Tabela 12 – Dados comparativos da distribuição da evasão na UFTM em meses desde o início do semestre, ano/semestre a ano/semestre, de 2013 a 2018

Qtd. de meses desde o início do semestre	2018/2	2018/1	2017/2	2017/1	2016/2	2016/1	2015/2	2015/1	2014/2	2014/1	2013/2	2013/1
0	0	36	84	21	82	51	106	53	130	47	47	21
1	0	80	138	63	137	35	175	86	56	10	121	4
2	0	22	28	25	19	69	86	44	5	368	115	17
3	0	13	15	197	5	13	14	7	10	330	24	39
4	7	327	8	160	3	279	3	35	17	45	66	49
5	32	72	4	81	47	38	10	92	101	74	21	0
6	237	0	346	0	169	0	96	196	72	0	0	0
7	103	0	14	0	8	0	0	0	5	0	0	0

Fonte: O autor (2018).

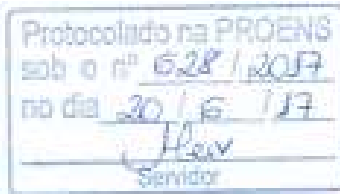
**APÊNDICE D – DADOS COMPARATIVO DA IDENTIFICAÇÃO DE ALUNOS EM
TENDÊNCIA À EVASÃO PELO MODELO, POR CURSOS NO MÊS DE
AGOSTO/2018**

Tabela 13 – Dados comparativos da identificação de alunos em tendência à evasão pelo modelo, por cursos no mês de agosto de 2018

Curso	Qtd. alunos	Qtd. alunos detectados	Qtd. alunos evadidos	Qtd. alunos identificados e evadidos	Qtd. alunos identificados pela primeira vez
MEDICINA	550	10	1	1	7
ENFERMAGEM	297	24	0	0	14
BIOMEDICINA	201	30	4	1	19
FISIOTERAPIA	320	21	1	0	14
NUTRIÇÃO	273	25	0	0	18
TERAPIA OCUPACIONAL	199	14	0	0	8
EDUCAÇÃO FÍSICA	249	48	1	0	34
LICENCIATURA EM CIÊNCIAS BIOLÓGICAS	265	60	16	12	25
LICENCIATURA EM EDUCAÇÃO DO CAMPO	83	10	1	1	0
LICENCIATURA EM FÍSICA	133	63	15	13	24
LICENCIATURA EM MATEMÁTICA	175	57	26	13	27
LICENCIATURA EM QUÍMICA	164	39	21	16	14
LICENCIATURA EM GEOGRAFIA	193	50	22	9	29
LICENCIATURA EM HISTÓRIA	271	57	16	8	25
PSICOLOGIA	307	11	1	0	6
LICENCIATURA EM LETRAS	244	65	23	15	29
SERVIÇO SOCIAL	266	44	23	15	18
ENGENHARIA AMBIENTAL	207	79	16	14	51
ENGENHARIA CIVIL	280	32	0	0	27
ENGENHARIA DE ALIMENTOS	174	42	1	1	29
ENGENHARIA DE PRODUÇÃO	286	78	12	12	58
ENGENHARIA ELÉTRICA	195	59	2	1	43
ENGENHARIA MECÂNICA	309	86	9	5	57
ENGENHARIA QUÍMICA	268	66	1	1	52
AGRONOMIA	189	61	2	2	34
LICENCIATURA EM QUÍMICA	85	46	17	13	15
LICENCIATURA EM CIÊNCIAS BIOLÓGICAS	132	58	13	12	45

Fonte: O autor (2018).

ANEXO A – PEDIDO DE AUTORIZAÇÃO DE ACESSO A DADOS




 MINISTÉRIO DA EDUCAÇÃO
 UNIVERSIDADE FEDERAL DO TRIÂNGULO MINEIRO - UFTM
 INSTITUTO DE CIÊNCIAS TECNOLÓGICAS E EXATAS - ICTE
 Programa de Mestrado Profissional em Inovação Tecnológica - PMPIT
 Av. Doutor Raulo de Souza Amor, 1291 - Universidade - 35064-290 UBERABA/MG (343331-200)

Memorando nº 93/2017/UFTM/ICTE - Coordenação do Programa de Mestrado Profissional em Inovação Tecnológica - PMPIT

Em 14 de junho de 2017.

À Pro-Reitoria de Ensino

C/Cópia: Diretoria de Tecnologia da Informação

Assunto: Pedido de autorização para acesso a dados

1 Solicitamos, respeitosamente, autorização para acessar os dados de alunos do banco de dados do SISCAD como parte da pesquisa: USO DA TÉCNICA DE MINERAÇÃO DE DADOS COMO FERRAMENTA DE GESTÃO DA EVASÃO NO ENSINO SUPERIOR, sob a responsabilidade do professor Dr. Julio Cesar de Souza Inácio Gonçalves. O objetivo dessa pesquisa é identificar padrões na base de dados do SISCAD, entre os alunos que evadem do curso, antes da ocorrência da evasão. Desta forma busca-se prever quais os alunos estão susceptíveis a evadir do curso e permitir que ações podem ser feitas junto a esses alunos para reduzir a evasão. Para atingirmos os resultados dessa pesquisa necessitamos das informações/dados de alunos de graduação registrados no SISCAD (instituto, curso, idade, período, notas, faltas, reprovações, local de residência da família, tipo do ingresso e ocorrências) como parte da metodologia proposta para alcance dos objetivos propostos no projeto da pesquisa.

2 Salientamos que esta autorização é indispensável para a submissão do projeto junto ao Comitê de Ética em Pesquisa da UFTM, órgão responsável pela apreciação ética em pesquisa com seres humanos. E conforme prevê a Resolução 466/12 CNS, a pesquisa somente será iniciada a partir da aprovação do referido comitê.

3 Sem mais para o momento, agradecemos a atenção e nos colocamos a disposição para eventuais esclarecimentos.

Atenciosamente,

De acordo com a realização da pesquisa:

deferido indeferido

Assinatura do Responsável da instituição

Prof. Dr. Geoffrey Roger Pointer Malpass
Coordenador do PMPIT da UFTM


 Programa de Mestrado Profissional
 em Inovação Tecnológica
 PMPIT - ICTE - UFTM


C. C. E. E. R. A. U. B. I. T. A. -

21.06.17


 Prof. Dr. Wagner Roberto Ruffini
 Diretor de Ensino - UFTM

Local e data

Prof. Dr. Julio Cesar de Souza Inácio Gonçalves
Docente do PMPIT


 Prof. Dr. JULIO CESAR DE SOUZA INÁCIO GONÇALVES
 Coordenador do Programa de Mestrado em
 Ciências e Tecnologia Ambiental
 UFTM
 Supp. 18-0081

ANEXO B – COMPROVANTE DE ENVIO DO PROJETO AO CEP



UFTM - UNIVERSIDADE
FEDERAL DO TRIÂNGULO
MINEIRO



COMPROVANTE DE ENVIO DO PROJETO

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: USO DA TÉCNICA DE MINERAÇÃO DE DADOS COMO UMA FERRAMENTA DE GESTÃO DA EVASÃO NO ENSINO SUPERIOR

Pesquisador: JULIO CESAR DE SOUZA INACIO GONCALVES

Versão: 1

CAAE: 71405317.0.0000.5154

Instituição Proponente: Universidade Federal do Triângulo Mineiro

DADOS DO COMPROVANTE

Número do Comprovante: 079604/2017

Patrocinador Principal: Universidade Federal do Triângulo Mineiro

Informamos que o projeto USO DA TÉCNICA DE MINERAÇÃO DE DADOS COMO UMA FERRAMENTA DE GESTÃO DA EVASÃO NO ENSINO SUPERIOR que tem como pesquisador responsável JULIO CESAR DE SOUZA INACIO GONCALVES, foi recebido para análise ética no CEP UFTM - Universidade Federal do Triângulo Mineiro em 18/07/2017 às 09:44.

Endereço: Rua Conde Prades, 191

Bairro: Nossa Sra. Abadia

UF: MG

Telefone: (34)3700-6963

CEP: 38.025-360

Município: UBERABA

E-mail: cep@uftm.edu.br

ANEXO C – PARECER CONSUBSTANCIADO DO CEP



UNIVERSIDADE FEDERAL DO
TRIÂNGULO MINEIRO - UFTM



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: USO DA TÉCNICA DE MINERAÇÃO DE DADOS COMO UMA FERRAMENTA DE GESTÃO DA EVASÃO NO ENSINO SUPERIOR

Pesquisador: JULIO CESAR DE SOUZA INACIO GONCALVES

Área Temática:

Versão: 1

CAAE: 71405317.0.0000.5154

Instituição Proponente: Universidade Federal do Triângulo Mineiro

Patrocinador Principal: Universidade Federal do Triângulo Mineiro

DADOS DO PARECER

Número do Parecer: 2.254.582

Apresentação do Projeto:

Segundo os pesquisadores, "Para esse estudo é proposta uma pesquisa aplicada ex-post-facto, do tipo quantitativa e descritiva com base nas informações existentes na base de dados do Sistema Acadêmico (SISCAD) da UFTM com o objetivo de identificar um padrão de informações relacionado à possibilidade de evasão dos alunos do ensino superior.

Com esse estudo pretende-se descobrir se antes de evadir do ensino superior o aluno apresenta sinais dessa tendência. Serão identificados e analisados esses sinais, a fim de se encontrar um padrão que permitirá a identificação, de forma sistêmica e automática, e que se utilize das informações já existentes no sistema, dos alunos com tendência a se evadirem do curso de graduação. Após a identificação dos alunos com tendência de evasão, possibilitar que agentes da instituição, como os coordenadores de curso e os servidores da Pró-Reitoria de Assuntos Comunitários e Estudantis (PROACE) da UFTM, tenham acesso a essa listagem a fim atuarem junto aos alunos com essa tendência de evasão e tentar revertê-la". As perguntas de pesquisa são "O aluno antes de evadir do curso apresenta algum sinal nos seus dados de registro e controle acadêmicos?

Quais são esses sinais?

Com base nesses sinais é possível prevenir/identificar quais alunos estão em tendência de evasão?

Endereço: Rua Madre Maria José, 122

Bairro: Nossa Sra. Abadia

UF: MG

Telefone: (34)3700-6778

CEP: 38.625-100

Município: LIBERABA

E-mail: cep@ceapq.uftm.edu.br



UNIVERSIDADE FEDERAL DO
TRIÂNGULO MINEIRO - UFTM



Continuação do Protocolo: 1.134.582

É possível gerar alertas dos alunos que estão em tendência de evasão de modo a se reverter essa situação?"

Objetivo da Pesquisa:

Segundo os pesquisadores, "Objetivo geral:

1. Criar uma ferramenta para reduzir a evasão acadêmica do ensino superior na UFTM.

Objetivos específicos:

1. Determinar os sinais apresentados pelo acadêmico antes de evadir, sinais estes presentes nos dados do aluno de registro e controle acadêmico.
2. Identificar quais os acadêmicos apresentam elevada probabilidade de evasão.
3. Criar um mecanismo de geração de alertas dos acadêmicos em tendência de evasão".

Avaliação dos Riscos e Benefícios:

Consta "O risco de perda de confidencialidade dos dados dos alunos que serão analisados é mínimo, pois os pesquisadores não repassarão, sob nenhum motivo, os dados dos alunos da instituição a terceiros e se comprometem em manter esses dados em local seguro.

Quanto aos benefícios da pesquisa entende-se que ao atingir seus objetivos, ou seja, ao apresentar alertas indicando os alunos com alta tendência a evadir, espera-se que a evasão seja reduzida. Reduzindo-se a evasão, possibilitará uma redução nos custos da educação superior, em especial nas instituições públicas pagas com dinheiro de impostos da sociedade, além de melhorar as perspectivas do aluno de conseguir obter o diploma, e com isso ter uma perspectiva melhor de salários e empregos".

Comentários e Considerações sobre a Pesquisa:

Pesquisa de relevância temática ao buscar a criação de uma ferramenta para reduzir a evasão acadêmica do ensino superior na UFTM.

Considerações sobre os Termos de apresentação obrigatória:

Foram apresentados os seguintes termos:

- Folha de rosto
- Projeto detalhado, conforme o protocolo do CEP/UFTM
- Autorização do local de coleta de dados
- Grupo de pesquisadores vinculado ao projeto na Plataforma Brasil

Endereço: Rua Madre Maria José, 122

Bairro: Nossa Sra. Abadia

CEP: 38.605-100

UF: MG

Município: UBERABA

Telefone: (34)3700-6776

E-mail: cep@pescpg.uftm.edu.br



UNIVERSIDADE FEDERAL DO
TRIÂNGULO MINEIRO - UFTM



Continuação do Parecer: 2.254.582

Conclusões ou Pendências e Lista de Inadequações:

De acordo com as atribuições definidas na Resolução CNS 466/12 e Norma Operacional 001/2013, o colegiado do CEP-UFTM manifesta-se pela aprovação do protocolo de pesquisa proposto. Situação definida em reunião do dia 01/09/2017.

Considerações Finais a critério do CEP:

A aprovação do protocolo de pesquisa pelo CEP/UFTM dá-se em decorrência do atendimento à Resolução CNS 466/12 e norma operacional 001/2013, não implicando na qualidade científica do mesmo.

Conforme prevê a legislação, são responsabilidades, indelegáveis e indeclináveis, do pesquisador responsável, dentre outras: comunicar o início da pesquisa ao CEP; elaborar e apresentar os relatórios parciais (semestralmente) e final. Para isso deverá ser utilizada a opção 'notificação' disponível na Plataforma Brasil.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB INFORMAÇÕES BÁSICAS DO PROJETO 962645.pdf	18/07/2017 09:44:26		Aceito
Projeto Detalhado / Brochura Investigador	ProjetoCEP.docx	18/07/2017 09:44:12	ALAN LOPES MELO	Aceito
Folha de Rosto	FolhaRosto.pdf	18/07/2017 09:41:32	ALAN LOPES MELO	Aceito
Declaração de Instituição e Infraestrutura	AcessoDados.pdf	18/07/2017 09:41:09	ALAN LOPES MELO	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Endereço: Rua Madre Maria José, 122

Bairro: Nossa Sra. Abadia

CEP: 38.029-100

UF: MG

Município: UBERABA

Telefone: (34)3700-6776

E-mail: cep@pescpq.uftm.edu.br



UNIVERSIDADE FEDERAL DO
TRIÂNGULO MINEIRO - UFTM



Continuação do Processo: 2.254.592

UBERABA, 01 de Setembro de 2017

Assinado por:

Marly Aparecida Spadotto Balarin
(Coordenador)

Endereço: Rua Madre Maria José, 122

Bairro: Nossa Sra. Abadia

CEP: 38.605-100

UF: MG

Município: UBERABA

Telefone: (34)3700-6778

E-mail: cep@paeopg.uftm.edu.br

ANEXO D – RELATÓRIO DE REPROVAÇÕES POR DISCIPLINA E ANO/SEMESTRE ORDENADAS PELO PERCENTUAL DE REPROVAÇÕES

Disciplina	Carga	Tipo	Ofertas	Matrículas	Aprovações	Reprovações	% Rep. Nota % Rep. Falta			
							Reprovações	% Rep. Nota	% Rep. Falta	
2504.000.016-3 - FUNDAMENTOS DE QUÍMICA II	105h/a	OBR	40	3	0	3	100	66,7	33,3	
2505.000.021-2 - ESTUDOS E DESENVOLVIMENTO DE PROJETOS IV	120h/a	OBR	54	3	0	3	100	-	100	
2504.000.013-9 - FUNDAMENTOS DE QUÍMICA I	60h/a	OBR	36	2	0	2	100	-	100	
3001.000.069-2 - PROJETO DE PROCESSOS AUXILIADOS POR COMPUTADOR	30h/a	ELE	32	1	0	1	100	100	-	
2505.000.079-4 - FUNÇÕES E SUAS APLICAÇÕES NO CAMPO AGRÁRIO	90h/a	OBR	60	9	1	8	88,9	44,4	44,4	
2505.000.078-6 - EDUCAÇÃO DO CAMPO, DIVERSIDADE LINGUÍSTICA E CULTURAL	30h/a	OBR	60	9	1	8	88,9	-	88,9	
2505.000.083-2 - PSICOLOGIA DA EDUCAÇÃO E DESENVOLVIMENTO	60h/a	OBR	60	9	1	8	88,9	-	88,9	
2505.000.086-7 - SEMINÁRIO INTEGRADOR II	30h/a	OBR	60	9	1	8	88,9	-	88,9	
2505.000.076-0 - CAMPOS NUMÉRICOS	90h/a	OBR	165	34	4	30	88,2	41,2	47,1	
2505.000.081-6 - LEITURA E ESCRITA DE TEXTOS	60h/a	OBR	165	34	7	27	79,4	2,9	76,5	
2004.000.010-3 - ATIVIDADES ACADÊMICO-CIENTÍFICO-CULTURAIS I	45h/a	OBR	240	46	10	36	78,3	-	23,9	
2505.000.088-3 - TEORIAS DA EDUCAÇÃO	60h/a	OBR	60	9	2	7	77,8	55,6	22,2	
2501.000.093-9 - MORFOFISIOLOGIA CELULAR E TECIDUAL	90h/a	OBR	84	53	12	41	77,4	62,3	15,1	
1001.000.071-0 - FUNDAMENTOS E ASPECTOS HISTÓRICOS DA EDUCAÇÃO FÍSICA	45h/a	OBR	13	12	3	9	75	8,3	66,7	
2004.000.012-0 - ATIVIDADES ACADÊMICO-CIENTÍFICO-CULTURAIS III	45h/a	OBR	175	23	7	16	69,6	-	26,1	
4001.000.030-4 - FUNDAMENTOS DE MATEMÁTICA	30h/a	OBR	215	164	50	114	69,5	40,9	28,7	
4001.000.121-1 - MINERALOGIA	60h/a	OBR	50	26	8	18	69,2	61,5	7,7	
2504.000.041-4 - FUNDAMENTOS EM QUÍMICA ORGÂNICA	30h/a	ELE	72	19	6	13	68,4	21,1	47,4	
4001.000.053-3 - QUÍMICA GERAL I	90h/a	OBR	330	110	32	74	67,3	42,7	24,5	
4001.000.055-0 - GEOMETRIA ANALÍTICA E ÁLGEBRA LINEAR	60h/a	OBR	123	92	31	61	66,3	34,8	31,5	
2501.000.028-9 - CÉLULA E TECIDOS VEGETAIS	60h/a	OBR	408	192	59	122	63,5	36,5	27,1	

3004.000.066-6 - ELETROMAGNETISMO	75h/a	OBR	116	32	5	20	62,5	28,1	34,4
4001.000.173-4 - BIOQUÍMICA APLICADA	60h/a	OBR	50	42	16	26	61,9	11,9	50
4001.000.064-9 - CÁLCULO DIFERENCIAL E INTEGRAL II	60h/a	OBR	110	42	16	26	61,9	19	42,9
3001.000.101-0 - FENÔMENOS DE TRANSPORTE I	45h/a	OBR	73	44	8	27	61,4	59,1	2,3
3004.000.022-4 - ELETROMAGNETISMO I	60h/a	OBR	375	191	74	117	61,3	41,4	19,9
2503.000.076-4 - PRÉ-CÁLCULO	60h/a	OBR, OPT	947	660	170	400	60,6	25,8	34,8
1001.000.007-8 - ATIVIDADES FÍSICA PARA GRUPOS ESPECIAIS	60h/a	OBR	279	249	95	150	60,2	36,5	23,7
3007.000.104-0 - ENGENHARIA DE PROCESSOS II	60h/a	OBR	74	15	6	9	60	33,3	26,7
2503.000.042-0 - CÁLCULO DIFERENCIAL E INTEGRAL III	60h/a	OBR	362	109	34	65	59,6	38,5	21,1
2503.000.035-7 - CÁLCULO DIFERENCIAL E INTEGRAL I	90h/a	OBR, OPT	376	238	62	141	59,2	31,1	28,2
2503.000.054-3 - FUNDAMENTOS DE MATEMÁTICA	60h/a	OBR	478	366	94	215	58,7	16,7	42,1
2503.000.018-7 - ÁLGEBRA LINEAR	60h/a	OBR, OPT	231	96	14	55	57,3	30,2	27,1
3001.000.097-8 - TERMODINÂMICA	60h/a	OBR, OPT	306	167	72	95	56,9	50,3	6,6
3010.000.028-4 - MODELAGEM MATEMÁTICA E SIMULAÇÃO COMPUTACIONAL	60h/a	ELE	130	16	7	9	56,3	56,3	-
3009.000.001-5 - CÁLCULO I	90h/a	OBR, OPT	3427	2930	1103	1637	55,9	48,5	7,4
3009.000.007-4 - ÁLGEBRA LINEAR	60h/a	OBR	3109	2609	866	1456	55,8	48,1	7,7
4001.000.029-0 - QUÍMICA GERAL E ORGÂNICA	30h/a	OBR	160	115	51	64	55,7	28,7	27
2004.000.011-1 - ATIVIDADES ACADÊMICO-CIENTÍFICO-CULTURAIS II	45h/a	OBR	200	29	13	16	55,2	-	6,9
2501.000.008-4 - APLICAÇÕES DO METABOLISMO ENERGÉTICO	30h/a	ELE	36	20	9	11	55	20	35
3009.000.010-4 - GEOMETRIA ANALÍTICA	60h/a	OBR, OPT ELE, OBR,	3318	2736	956	1486	54,3	45,3	9
3006.000.054-8 - PESQUISA OPERACIONAL I	60h/a	OPT	700	459	181	241	52,5	46,4	6,1
4001.000.062-2 - QUÍMICA ANALÍTICA I	75h/a	OBR	115	27	13	14	51,9	37	14,8
4001.000.039-8 - QUÍMICA ORGÂNICA I	60h/a	OBR	60	29	14	15	51,7	24,1	27,6
3004.000.014-3 - CIRCUITOS ELÉTRICOS I	75h/a	OBR	301	157	76	81	51,6	47,8	3,8
2503.000.080-2 - TEORIA DOS NÚMEROS	75h/a	OBR	324	78	33	40	51,3	21,8	29,5
2002.000.067-1 - TRABALHO DE CONCLUSÃO DE CURSO II	45h/a	OBR	364	218	80	111	50,9	37,6	13,3

3006.000.072-6 - TÓPICOS ESPECIAIS EM GERÊNCIA DA PRODUÇÃO	30h/a	ELE	92	55	23	28	50,9	20	30,9
1001.000.045-0 - FISIOLOGIA DO ESPORTE E DO EXERCÍCIO II	60h/a	OBR, OPT	363	242	98	123	50,8	33,1	17,8
3004.000.057-7 - TÓPICOS EM ENGENHARIA DE COMPUTAÇÃO	60h/a	ELE, OBR, OPT	320	192	95	97	50,5	32,3	18,2
1503.000.048-1 - FISIOLOGIA HUMANA	90h/a	OBR, OPT	372	274	104	137	50	27,7	22,3
2502.000.037-0 - FÍSICA I	75h/a	OBR, OPT	246	78	39	39	50	37,2	12,8
1008.000.017-4 - BIOESTATÍSTICA	60h/a	OBR	83	40	20	20	50	22,5	27,5
2004.000.013-8 - ATIVIDADES ACADÊMICO-CIENTÍFICO-CULTURAIS IV	45h/a	OBR	175	30	15	15	50	-	16,7
2003.000.057-7 - HISTÓRIA E NATUREZA	60h/a	ELE, OPT	46	10	5	5	50	10	40
4001.000.143-2 - FISIOLOGIA VEGETAL	90h/a	OBR	50	6	3	3	50	50	-
2505.000.154-5 - PESQUISA E ENSINO-APRENDIZAGEM DA MATEMÁTICA III	105h/a	OBR	72	6	1	3	50	50	-
1006.000.060-8 - FISIOTERAPIA DERMATOFUNCIONAL	45h/a	OBR	7	2	1	1	50	50	-
3010.000.040-3 - TÓPICOS ESPECIAIS EM HIGIENE INDUSTRIAL	30h/a	ELE	65	2	1	1	50	50	-
4001.000.001-0 - BIOLOGIA CELULAR	60h/a	OBR	269	169	80	84	49,7	29	20,7
2501.000.005-0 - ANATOMIA VEGETAL	60h/a	ELE, OBR, OPT	346	184	42	91	49,5	25	24,5
2503.000.083-7 - GEOMETRIA ANALÍTICA	60h/a	OBR, OPT	502	348	96	172	49,4	24,7	24,7
4001.000.007-0 - METODOLOGIA CIENTÍFICA	30h/a	OBR	167	152	77	75	49,3	24,3	25
4001.000.027-4 - FUNDAMENTOS DE SISTEMÁTICA	75h/a	OBR	156	118	53	58	49,2	31,4	17,8
2006.000.051-6 - TEORIA DA LITERATURA I	60h/a	OBR, OPT	555	391	142	192	49,1	26,1	23
2505.000.084-0 - SAÚDE, SEXO E REPRODUÇÃO	45h/a	OBR	163	51	26	25	49	15,7	33,3
1503.000.011-2 - BIOQUÍMICA	45h/a	OBR	383	341	137	166	48,7	32,3	16,4
2503.000.058-6 - GEOMETRIA ANALÍTICA	30h/a	OBR	362	140	48	68	48,6	23,6	25
2006.000.052-4 - TEORIA DA LITERATURA II	60h/a	OBR	502	270	98	131	48,5	26,7	21,9
4001.000.028-2 - MORFOLOGIA VEGETAL	30h/a	OBR	144	118	60	57	48,3	24,6	23,7
4001.000.038-0 - FUNDAMENTOS DE QUÍMICA II	120h/a	OBR	60	29	15	14	48,3	34,5	13,8
2503.000.021-7 - ANÁLISE REAL I	75h/a	OBR, OPT	288	52	20	25	48,1	32,7	15,4
2504.000.015-5 - FUNDAMENTOS DE QUÍMICA II	75h/a	OBR, OPT	373	215	93	103	47,9	30,7	17,2
2502.000.039-7 - FÍSICA I	75h/a	OBR, OPT	227	67	24	32	47,8	26,9	20,9
2503.000.050-0 - ESTRUTURAS ALGÉBRICAS I	75h/a	OBR	364	86	32	41	47,7	29,1	18,6

2505.000.087-5 - TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO APLICADAS À EDUCAÇÃO	45h/a	OBR	132	19	0	9	47,4	-	47,4
3005.000.059-6 - TERMODINÂMICA	75h/a	OBR, OPT	317	231	122	109	47,2	36,4	10,8
3004.000.001-1 - PROGRAMAÇÃO COMPUTACIONAL	60h/a	OBR, OPT	2761	2406	1203	1134	47,1	29,9	17,2
2501.000.024-6 - BIOLOGIA CELULAR	60h/a	OBR, OPT	370	296	116	139	47	34,8	11,8
1001.000.019-1 - CRESCIMENTO, DESENVOLVIMENTO HUMANO E ENVELHECIMENTO	60h/a	OBR	21	15	8	7	46,7	6,7	40
1504.000.006-9 - FUNDAMENTOS DE IMUNOLOGIA	45h/a	OBR	361	206	90	96	46,6	32	14,6
2004.000.028-6 - ESTUDOS ORIENTADOS COMPLEMENTARES I	30h/a	OBR	530	350	142	162	46,3	-	19,7
2504.000.022-8 - QUÍMICA ANALÍTICA QUALITATIVA	75h/a	OBR	137	52	28	24	46,2	26,9	19,2
1504.000.020-4 - IMUNOLOGIA BÁSICA	60h/a	OBR	361	347	151	160	46,1	31,4	14,7
2502.000.047-8 - FUNDAMENTOS DA FÍSICA	60h/a	OBR	361	183	75	83	45,4	18,6	26,8
3002.000.010-5 - HIDRÁULICA AMBIENTAL I	60h/a	OBR, OPT ELE, OBR,	287	172	94	78	45,3	45,3	-
1501.000.068-0 - NEUROANATOMIA	45h/a	OPT	413	341	149	154	45,2	33,4	11,7
2502.000.002-8 - TÓPICOS DE ENSINO DE FÍSICA III	30h/a	ELE	73	20	11	9	45	10	35
2009.000.121-9 - TRABALHO DE CONCLUSÃO DE CURSO II	120h/a	OBR	564	346	164	155	44,8	-	17,6
3006.000.031-9 - CONTROLE ESTATÍSTICO DE QUALIDADE	60h/a	ELE, OBR	382	302	167	135	44,7	30,5	14,2
2004.000.051-0 - LÍNGUA PORTUGUESA II	60h/a	OBR	459	276	104	123	44,6	19,9	24,6
3001.000.038-2 - FUNDAMENTOS DE FENÔMENOS DE TRANSPORTE	60h/a	OBR, OPT	2460	2069	973	916	44,3	37	7,2
4001.000.177-7 - QUÍMICA ANALÍTICA	60h/a	OBR	110	77	10	34	44,2	22,1	22,1
4001.000.058-4 - QUÍMICA ORGÂNICA I	75h/a	OBR	170	68	11	30	44,1	30,9	13,2
4001.000.170-0 - ECOLOGIA	60h/a	OBR	70	59	33	26	44,1	27,1	16,9
4001.000.066-5 - HISTÓRIA DA CIÊNCIA E ENSINO DE QUÍMICA	90h/a	OBR	111	25	14	11	44	20	24
4001.000.171-8 - QUÍMICA GERAL	60h/a	OBR	253	135	66	59	43,7	28,9	14,8
4001.000.014-2 - METODOLOGIA CIENTÍFICA	30h/a	OBR	158	126	71	55	43,7	19	24,6

1501.000.045-1 - EMBRIOLOGIA GERAL E APLICADA	45h/a	OBR, OPT	415	390	179	169	43,3	33,3	10
2501.000.108-0 - QUÍMICA MEDICINAL	30h/a	ELE, OBR	72	37	21	16	43,2	10,8	32,4
2505.000.033-6 - INTRODUÇÃO À CIÊNCIA	30h/a	ELE, OBR	393	234	115	101	43,2	10,3	32,9
2501.000.106-4 - PRINCÍPIOS QUÍMICOS E BIOQUÍMICOS	60h/a	OBR, OPT	377	290	136	125	43,1	21,4	21,7
3007.000.076-1 - TERMODINÂMICA PARA ENGENHARIA QUÍMICA I	60h/a	OBR	486	356	173	153	43	38,8	4,2
2505.000.074-3 - TÓPICOS DE ENSINO DE FÍSICA I	30h/a	ELE	72	28	16	12	42,9	3,6	39,3
4001.000.019-3 - FUNDAMENTOS DE QUÍMICA I	60h/a	OBR	46	40	23	17	42,5	37,5	5
4001.000.060-6 - PRÁTICA PEDAGÓGICA PARA O ENSINO DE QUÍMICA I	30h/a	OBR	194	118	68	50	42,4	11,9	30,5
4001.000.050-9 - ANATOMIA HUMANA	75h/a	OBR	120	64	33	27	42,2	28,1	14,1
3007.000.050-8 - FÍSICO-QUÍMICA I	60h/a	OBR	60	57	33	24	42,1	42,1	-
4001.000.183-1 - QUÍMICA ORGÂNICA	30h/a	OBR	123	57	33	24	42,1	26,3	15,8
2002.000.009-4 - CARTOGRAFIA I	75h/a	OBR	160	55	32	23	41,8	25,5	16,4
4001.000.168-8 - MATEMÁTICA APLICADA I	60h/a	OBR, OPT	233	156	48	65	41,7	31,4	10,3
1503.000.019-8 - BIOQUÍMICA E BIOFÍSICA	120h/a	OBR, OPT	25	24	14	10	41,7	37,5	4,2
2502.000.045-1 - FÍSICA MATEMÁTICA I	30h/a	OBR, OPT	367	82	36	34	41,5	22	19,5
3004.000.064-0 - CIRCUITOS ELÉTRICOS I	90h/a	OBR	109	82	19	34	41,5	31,7	9,8
2501.000.110-2 - REDAÇÃO CIENTÍFICA	30h/a	ELE	108	46	27	19	41,3	-	41,3
2503.000.077-2 - PROBABILIDADE E ESTATÍSTICA I	75h/a	OBR, OPT	325	83	46	34	41	14,5	26,5
2503.000.079-9 - RACIOCÍNIO LÓGICO-QUANTITATIVO	30h/a	ELE, OBR	837	528	239	216	40,9	11,9	29
3004.000.034-8 - INTELIGÊNCIA ARTIFICIAL	60h/a	OBR	260	135	80	55	40,7	31,9	8,9
1501.000.002-8 - ANATOMIA	75h/a	OBR, OPT	381	313	163	126	40,3	20,8	19,5
4001.000.169-6 - MORFOLOGIA E TAXONOMIA VEGETAL	60h/a	OBR	166	127	76	51	40,2	28,3	11,8
3009.000.005-8 - CÁLCULO III	90h/a	OBR	2020	1586	820	635	40	34,1	5,9

ANEXO E – RELATÓRIO DE RELAÇÃO DE REPROVAÇÕES POR DISCIPLINA

Disciplina	Matri. 1 ¹	Matri. 2 ²	Matri. 3 ³	Matri. 4 ⁴	Matri. 5 ⁵	Matri. 6 ⁶	Matri. 7 ⁷	Tot. Mat.	Vag. Ofer.
2009.000.114-6 - SUPERVISÃO DE ESTÁGIO CURRICULAR IV	180	9	4	1	-	-	1	195	380
2009.000.023-9 - ESTÁGIO SUPERVISIONADO EM SERVIÇO SOCIAL IV	182	7	4	1	-	-	1	195	385
3009.000.001-5 - CÁLCULO I	1172	510	174	36	12	2	-	1906	3434
3009.000.007-4 - ÁLGEBRA LINEAR	1037	411	165	56	20	2	-	1691	3281
3001.000.038-2 - FUNDAMENTOS DE FENÔMENOS DE TRANSPORTE	817	335	120	44	8	1	-	1325	2479
3008.000.001-2 - FÍSICA PARA ENGENHARIA I	875	116	42	8	-	1	-	1042	2127
1503.000.022-8 - BIOQUÍMICA E BIOFÍSICA	386	49	24	7	2	1	-	469	644
1503.000.033-3 - FARMACOLOGIA	333	40	18	8	2	1	-	402	733
1002.000.060-7 - PATOLOGIA ESPECIAL	355	1	-	-	-	1	-	357	433
2007.000.202-3 - TÉCNICAS PSICOTERÁPICAS IV	274	-	-	-	-	1	-	275	362
3006.000.054-8 - PESQUISA OPERACIONAL I	108	75	42	11	5	1	-	242	705
1503.000.011-2 - BIOQUÍMICA	160	45	20	5	1	1	-	232	383
2009.000.121-9 - TRABALHO DE CONCLUSÃO DE CURSO II	123	37	19	14	6	1	-	200	601
2004.000.029-4 - ESTUDOS ORIENTADOS COMPLEMENTARES II	151	16	9	4	1	1	-	182	461
1502.000.011-0 - GENÉTICA	130	24	5	2	-	1	-	162	375
1504.000.006-9 - FUNDAMENTOS DE IMUNOLOGIA	70	32	7	6	3	2	-	120	361
2002.000.043-4 - HIDROLOGIA	90	15	9	3	-	1	-	118	327
1001.000.007-8 - ATIVIDADE FÍSICA PARA GRUPOS ESPECIAIS	51	31	15	12	5	3	-	117	279
2002.000.067-1 - TRABALHO DE CONCLUSÃO DE CURSO II	66	24	12	5	6	3	-	116	364
1001.000.106-6 - PRODUÇÃO E VEICULAÇÃO DO CONHECIMENTO EM ESPORTES E SAÚDE II - TCC	86	12	7	2	-	1	-	108	288
2002.000.052-3 - ORIENTAÇÃO E ESTÁGIO CURRICULAR SUPERVISIONADO II	92	6	-	1	-	1	-	100	288
2005.000.041-6 - ORIENTAÇÃO E ESTÁGIO SUPERVISIONADO EM LÍNGUA INGLESA II	69	1	-	-	-	1	-	71	144
2503.000.050-0 - ESTRUTURAS ALGÉBRICAS I	32	9	4	2	2	1	-	50	364
2504.000.040-6 - TRABALHO DE CONCLUSÃO DE CURSO II	41	4	-	-	1	1	-	47	216
3009.000.010-4 - GEOMETRIA ANALÍTICA	1121	472	187	65	8	-	-	1853	3618
3007.000.001-0 - QUÍMICA PARA ENGENHARIA	1317	308	93	22	5	-	-	1745	2878
3002.000.003-2 - INTRODUÇÃO À CIÊNCIA DOS MATERIAIS	1457	225	48	10	4	-	-	1744	2652
3006.000.056-4 - PROBABILIDADE E ESTATÍSTICA PARA ENGENHARIA	1331	261	86	22	4	-	-	1704	2652

3004.000.001-1 - PROGRAMAÇÃO COMPUTACIONAL	1036	460	122	19	2	-	-	1639	2807
3002.000.002-4 - DESENVOLVIMENTO SUSTENTÁVEL E CIDADANIA	1482	63	15	3	1	-	-	1564	2214
2004.000.040-5 - LEITURA E PRODUÇÃO DE TEXTOS	1390	71	16	2	1	-	-	1480	2847
3008.000.004-7 - FÍSICA PARA ENGENHARIA II	891	161	30	13	2	-	-	1097	1721
3009.000.005-8 - CÁLCULO III	777	185	99	23	10	-	-	1094	2082
3009.000.003-1 - CÁLCULO II	877	130	37	11	2	-	-	1057	2090
1501.000.044-3 - EMBRIOLOGIA GERAL	828	113	36	12	1	-	-	990	1341
1504.000.035-2 - PARASITOLOGIA	692	72	17	10	2	-	-	793	1682
2001.000.003-2 - CONCEPÇÕES DE EDUCAÇÃO E NOVOS PARADIGMAS	625	30	6	-	1	-	-	662	1433
3009.000.006-6 - CÁLCULO IV	548	45	5	2	1	-	-	601	1379
1504.000.024-7 - MICROBIOLOGIA	543	46	7	3	1	-	-	600	927
1501.000.043-5 - EMBRIOLOGIA CLÍNICA	463	31	13	2	1	-	-	510	670
2503.000.076-4 - PRÉ-CÁLCULO	335	72	30	14	7	-	-	458	947
1501.000.068-0 - NEUROANATOMIA	321	41	12	3	1	-	-	378	563
1502.000.019-5 - GENÉTICA HUMANA	311	34	17	-	1	-	-	363	519
1503.000.020-1 - BIOQUÍMICA E BIOFÍSICA	267	42	12	4	1	-	-	326	440
1001.000.083-3 - INFORMÁTICA APLICADA À SAÚDE	290	17	2	-	1	-	-	310	405
1501.000.016-8 - ANATOMIA HUMANA I	258	36	11	1	1	-	-	307	445
1503.000.050-3 - FISILOGIA I	252	34	11	2	2	-	-	301	489
1001.000.108-2 - PSICOLOGIA DO ESPORTE	262	21	2	-	1	-	-	286	456
1006.000.009-8 - BIOFÍSICA	273	2	-	1	1	-	-	277	364
2002.000.035-3 - GEOLOGIA I	245	21	1	1	1	-	-	269	447
1501.000.002-8 - ANATOMIA	235	26	3	3	1	-	-	268	381
1503.000.010-4 - BIOQUÍMICA	196	48	9	2	1	-	-	256	362
1503.000.043-0 - FISILOGIA	199	38	11	1	6	-	-	255	403
1004.000.004-1 - ANTROPOLOGIA FILOSÓFICA	248	5	-	-	1	-	-	254	338
1504.000.020-4 - IMUNOLOGIA BÁSICA	183	46	15	3	3	-	-	250	361
1001.000.017-5 - CINESIOLOGIA	217	23	5	2	1	-	-	248	361
2004.000.028-6 - ESTUDOS ORIENTADOS COMPLEMENTARES I	182	34	21	8	1	-	-	246	530
2004.000.110-0 - TRABALHO DE CONCLUSÃO DE CURSO II	217	17	6	2	2	-	-	244	670
2501.000.024-6 - BIOLOGIA CELULAR	200	36	5	1	1	-	-	243	370
2501.000.106-4 - PRINCÍPIOS QUÍMICOS E BIOQUÍMICOS	203	25	8	2	1	-	-	239	377
1001.000.055-8 - FUNDAMENTOS DAS ATIVIDADES AQUÁTICAS I	187	26	7	3	1	-	-	224	360

1001.000.096-5 - OFICINA DE EXPERIÊNCIA PROFISSIONAL EM EDUCAÇÃO FÍSICA E ESPORTES	211	5	1	-	1	-	-	218	363
3006.000.066-1 - SISTEMAS DE INFORMAÇÃO	154	39	12	7	1	-	-	213	422
2501.000.037-8 - ECOLOGIA GERAL	173	26	6	2	1	-	-	208	386
1501.000.069-9 - NEUROANATOMIA	174	21	6	5	1	-	-	207	314
2002.000.015-9 - CLIMATOLOGIA I	195	6	2	-	1	-	-	204	281
2501.000.105-6 - PRINCÍPIOS FÍSICOS E BIOFÍSICOS	177	18	7	-	1	-	-	203	369
1501.000.053-2 - HISTOLOGIA	175	3	-	-	1	-	-	179	229
1010.000.023-9 - FUNDAMENTOS DE TERAPIA OCUPACIONAL I	160	15	-	1	1	-	-	177	360
1503.000.046-5 - FISIOLOGIA	132	17	9	2	1	-	-	161	222
2004.000.030-8 - ESTUDOS ORIENTADOS COMPLEMENTARES III	132	18	5	5	1	-	-	161	441
2004.000.077-4 - ORIENTAÇÃO E ESTÁGIO SUPERVISIONADO EM LÍNGUA PORTUGUESA I	126	12	-	2	1	-	-	141	377
2502.000.047-8 - FUNDAMENTOS DA FÍSICA	111	18	9	1	1	-	-	140	361
2009.000.028-0 - ESTUDOS E DESENVOLVIMENTO DE PROJETOS V	131	5	-	1	1	-	-	138	220
2503.000.035-7 - CÁLCULO DIFERENCIAL E INTEGRAL I	78	28	14	8	6	-	-	134	377
2004.000.032-4 - ESTUDOS ORIENTADOS COMPLEMENTARES V	111	16	1	2	1	-	-	131	388
2009.000.038-7 - FUNDAMENTOS HISTÓRICOS E TEÓRICO-METODOLÓGICOS DO SERVIÇO SOCIAL II	102	10	8	1	3	-	-	124	229
3004.000.022-4 - ELETROMAGNETISMO I	79	26	13	4	1	-	-	123	383
1001.000.105-8 - PRODUÇÃO E VEICULAÇÃO DO CONHECIMENTO EM ESPORTES E SAÚDE III - TCC	91	15	5	1	2	-	-	114	308
2002.000.066-3 - TRABALHO DE CONCLUSÃO DE CURSO I	92	10	8	2	2	-	-	114	325
1001.000.107-4 - PRODUÇÃO E VEICULAÇÃO DO CONHECIMENTO EM ESPORTES E SAÚDE I - TCC	70	19	6	3	1	-	-	99	250
2003.000.086-0 - SOCIEDADE COLONIAL AMERICANA	83	8	3	1	1	-	-	96	197
1001.000.095-7 - NUTRIÇÃO APLICADA AO EXERCÍCIO	63	25	3	2	1	-	-	94	230
2502.000.033-8 - FÍSICA EXPERIMENTAL I	77	4	-	1	1	-	-	83	303
2505.000.019-0 - ESTUDOS E DESENVOLVIMENTO DE PROJETOS III	71	9	-	-	1	-	-	81	360
2503.000.043-8 - CÁLCULO DIFERENCIAL E INTEGRAL III	53	12	2	-	1	-	-	68	426
2503.000.042-0 - CÁLCULO DIFERENCIAL E INTEGRAL III	30	15	6	4	3	-	-	58	363
2502.000.045-1 - FÍSICA MATEMÁTICA I	45	6	4	2	1	-	-	58	369
2502.000.039-7 - FÍSICA I	33	9	1	2	1	-	-	46	229
2004.000.071-5 - METODOLOGIA CIENTÍFICA	1996	64	8	1	-	-	-	2069	3173

3002.000.001-6 - CIÊNCIAS DO AMBIENTE	1642	93	16	7	-	-	-	1758	2449
1504.000.002-6 - BIOÉTICA	1356	43	5	2	-	-	-	1406	1794
3006.000.077-7 - ADMINISTRAÇÃO E ORGANIZAÇÃO	1230	66	9	1	-	-	-	1306	1814
1008.000.110-3 - SAÚDE E SOCIEDADE	1262	28	4	1	-	-	-	1295	1810
1501.000.026-5 - BIOLOGIA CELULAR	1231	46	7	2	-	-	-	1286	1662
3001.000.002-1 - LABORATÓRIO DE FUNDAMENTOS DE FENÔMENOS DE TRANSPORTE	1096	94	9	2	-	-	-	1201	1876
2004.000.046-4 - LÍNGUA BRASILEIRA DE SINAIS (LIBRAS)	1122	66	9	2	-	-	-	1199	3111
2008.000.010-4 - SOCIOLOGIA	1130	34	3	1	-	-	-	1168	1802
3004.000.020-8 - ELETRICIDADE APLICADA À ENGENHARIA	845	57	5	1	-	-	-	908	1590
3003.000.063-9 - MECÂNICA DOS MATERIAIS	586	149	26	4	-	-	-	765	1573
3008.000.009-8 - FÍSICA PARA ENGENHARIA III	692	57	8	1	-	-	-	758	1305
2001.000.005-9 - CURRÍCULO ESCOLAR	697	42	5	4	-	-	-	748	1704
1008.000.088-3 - POLÍTICAS DE SAÚDE	685	40	6	1	-	-	-	732	1099
2001.000.002-4 - AVALIAÇÃO EDUCACIONAL	657	31	7	3	-	-	-	698	1572
2001.000.014-8 - POLÍTICA E ORGANIZAÇÃO DA EDUCAÇÃO BÁSICA NO BRASIL	623	22	3	1	-	-	-	649	1415
2001.000.010-5 - HISTÓRIA DA EDUCAÇÃO NO BRASIL	605	28	4	1	-	-	-	638	1449
1503.000.039-2 - FÍSICO-QUÍMICA	481	104	9	2	-	-	-	596	849
1501.000.015-0 - ANATOMIA HUMANA I	556	23	1	1	-	-	-	581	762
1503.000.068-6 - QUÍMICA GERAL E ANALÍTICA	453	88	15	2	-	-	-	558	810
1504.000.036-0 - PARASITOLOGIA	504	12	6	1	-	-	-	523	741
1503.000.036-8 - FARMACOLOGIA I	477	9	1	1	-	-	-	488	649
2003.000.040-2 - FORMAÇÃO SÓCIO-HISTÓRICA DO BRASIL	455	20	4	2	-	-	-	481	868
1504.000.021-2 - IMUNOLOGIA BÁSICA E APLICADA	387	65	16	3	-	-	-	471	722
1002.000.072-0 - TÉCNICA CIRÚRGICA E CIRURGIA EXPERIMENTAL	413	-	-	1	-	-	-	414	487
1001.000.088-4 - METODOLOGIA CIENTÍFICA	385	17	3	2	-	-	-	407	575
1003.000.059-6 - PROPEDÊUTICA E CLÍNICA MÉDICA	395	-	-	1	-	-	-	396	475
3003.000.066-3 - MECÂNICA GERAL	331	31	6	1	-	-	-	369	645
2006.000.051-6 - TEORIA DA LITERATURA I	321	28	3	1	-	-	-	353	555
2004.000.050-2 - LÍNGUA PORTUGUESA I	327	13	2	1	-	-	-	343	494
2004.000.034-0 - FUNDAMENTOS DE EDUCAÇÃO I	333	6	-	1	-	-	-	340	487
1501.000.013-3 - ANATOMIA HUMANA	279	42	11	3	-	-	-	335	480
1503.000.052-0 - FISILOGIA II	293	23	6	1	-	-	-	323	481
1501.000.045-1 - EMBRIOLOGIA GERAL E APLICADA	267	40	13	1	-	-	-	321	415

2007.000.015-2 - AVALIAÇÃO PSICOLÓGICA I	310	8	1	1	-	-	-	320	392
1503.000.005-8 - BASES FUNCIONAIS DO COMPORTAMENTO - PSICOFISIOLOGIA	303	9	1	1	-	-	-	314	396
3003.000.051-5 - GEOLOGIA	208	78	13	2	-	-	-	301	481
2503.000.054-3 - FUNDAMENTOS DE MATEMÁTICA	249	34	15	1	-	-	-	299	478
2007.000.001-2 - ANÁLISE DO COMPORTAMENTO	287	9	1	1	-	-	-	298	366
1501.000.010-9 - ANATOMIA HUMANA	244	41	9	3	-	-	-	297	408
1001.000.049-3 - FUNDAMENTOS DA GINÁSTICA I	266	18	4	2	-	-	-	290	437
2004.000.109-6 - TRABALHO DE CONCLUSÃO DE CURSO I	270	16	2	2	-	-	-	290	782
2503.000.083-7 - GEOMETRIA ANALÍTICA	230	43	4	5	-	-	-	282	503
2007.000.017-9 - AVALIAÇÃO PSICOLÓGICA III	269	4	2	1	-	-	-	276	361
1006.000.026-8 - ENVOLVIMENTO PRÁTICO ORIENTADO III	269	2	1	1	-	-	-	273	360
1008.000.002-6 - ANTROPOLOGIA DA SAÚDE	253	15	2	2	-	-	-	272	378
1010.000.033-6 - PRÁTICAS E VIVÊNCIAS EM TERAPIA OCUPACIONAL I	261	8	1	1	-	-	-	271	371
1010.000.031-0 - INTRODUÇÃO E HISTÓRIA DA TERAPIA OCUPACIONAL	262	6	2	1	-	-	-	271	373
1501.000.048-6 - FUNDAMENTOS DE HISTOLOGIA	243	20	6	1	-	-	-	270	400
1006.000.012-8 - CINESIOLOGIA I	247	8	1	1	-	-	-	257	362
1001.000.052-3 - FUNDAMENTOS DA GINÁSTICA III	233	20	-	1	-	-	-	254	442
1501.000.022-2 - ANATOMIA PALPATÓRIA	242	5	-	1	-	-	-	248	360
1009.000.008-8 - ATENÇÃO NUTRICIONAL	239	5	-	2	-	-	-	246	375
1008.000.039-5 - ESTATÍSTICA APLICADA À PSICOLOGIA	235	8	1	1	-	-	-	245	322
2504.000.014-7 - FUNDAMENTOS DE QUÍMICA I	209	29	2	1	-	-	-	241	400
3003.000.076-0 - RESISTÊNCIA DOS MATERIAIS I	214	14	9	3	-	-	-	240	438
3003.000.033-7 - DESENHO ARQUITETÔNICO	227	11	-	1	-	-	-	239	276
1503.000.030-9 - FARMACOLOGIA	177	49	9	4	-	-	-	239	387
1501.000.027-3 - BIOLOGIA CELULAR	187	36	7	6	-	-	-	236	363
2004.000.051-0 - LÍNGUA PORTUGUESA II	190	36	2	2	-	-	-	230	459
1006.000.015-2 - CLÍNICA CARDIOLÓGICA	213	13	1	1	-	-	-	228	360
1006.000.045-4 - FISIOTERAPIA APLICADA À CARDIOLOGIA	217	8	2	1	-	-	-	228	362
1501.000.057-5 - HISTOLOGIA	216	7	3	1	-	-	-	227	335
3003.000.077-9 - RESISTÊNCIA DOS MATERIAIS II	186	28	10	2	-	-	-	226	352
1001.000.011-6 - BIOMECÂNICA	176	40	6	1	-	-	-	223	366
2501.000.031-9 - DIVERSIDADE DE INVERTEBRADOS	193	21	4	1	-	-	-	219	361

2501.000.062-9 - FUNDAMENTOS MATEMÁTICOS E BIOESTATÍSTICA	179	28	10	1	-	-	-	218	446
3007.000.076-1 - TERMODINÂMICA PARA ENGENHARIA QUÍMICA I	117	71	23	7	-	-	-	218	486
2501.000.104-8 - PRINCÍPIOS EVOLUTIVOS	203	8	2	1	-	-	-	214	336
1503.000.048-1 - FISILOGIA HUMANA	169	34	7	4	-	-	-	214	372
1001.000.060-4 - FUNDAMENTOS DO ATLETISMO II	203	6	2	1	-	-	-	212	322
2505.000.049-2 - ORIENTAÇÃO E ESTÁGIO CURRICULAR SUPERVISIONADO III	197	8	3	1	-	-	-	209	828
3006.000.031-9 - CONTROLE ESTATÍSTICO DE QUALIDADE	139	52	9	8	-	-	-	208	390
1008.000.095-6 - PSICOLOGIA DA PERSONALIDADE	188	17	-	1	-	-	-	206	312
2009.000.104-9 - SERVIÇO SOCIAL, NOVAS TECNOLOGIAS DE COMUNICAÇÃO E GESTÃO DO CONHECIMENTO	200	2	1	2	-	-	-	205	365
3007.000.071-0 - QUÍMICA ORGÂNICA I	166	28	8	2	-	-	-	204	309
2002.000.093-0 - GEOGRAFIA DA POPULAÇÃO	199	2	-	1	-	-	-	202	289
3001.000.066-8 - TERMODINÂMICA APLICADA	146	39	10	7	-	-	-	202	347
2009.000.096-4 - SEMINÁRIO DE TRABALHO DE CONCLUSÃO DE CURSO II	196	2	-	2	-	-	-	200	334
2503.000.033-0 - CÁLCULO DIFERENCIAL E INTEGRAL I	141	41	11	7	-	-	-	200	707
1503.000.054-6 - FUNDAMENTOS DA INTERAÇÃO NUTRIENTES-FÁRMACOS	183	12	2	2	-	-	-	199	372
2009.000.063-8 - OFICINA DE PRÁTICA PROFISSIONAL II	197	1	-	1	-	-	-	199	324
1503.000.012-0 - BIOQUÍMICA	171	19	5	2	-	-	-	197	238
3007.000.092-3 - OPERAÇÕES UNITÁRIAS I	167	21	6	2	-	-	-	196	383
3006.000.055-6 - PESQUISA OPERACIONAL II	125	58	11	1	-	-	-	195	581
2501.000.059-9 - FUNDAMENTOS EM ANATOMIA	182	11	-	1	-	-	-	194	311
3007.000.078-8 - TRABALHO DE CONCLUSÃO DE CURSO	184	8	-	1	-	-	-	193	448
2004.000.066-9 - LINGUÍSTICA III	169	20	2	2	-	-	-	193	409
3007.000.033-8 - CINÉTICA QUÍMICA	135	40	10	2	-	-	-	187	298
3005.000.028-6 - DESENHO DE MÁQUINAS	148	32	5	1	-	-	-	186	358
2006.000.037-0 - LITERATURA PORTUGUESA I	171	13	-	1	-	-	-	185	404
1001.000.114-7 - TÓPICOS ESPECIAIS EM EDUCAÇÃO FÍSICA E SAÚDE	177	3	-	1	-	-	-	181	370
2004.000.053-7 - LÍNGUA PORTUGUESA III	157	20	1	1	-	-	-	179	396
1001.000.036-1 - ESTÁGIO PROFISSIONAL SUPERVISIONADO II	173	3	2	1	-	-	-	179	269
3007.000.077-0 - TERMODINÂMICA PARA ENGENHARIA QUÍMICA II	139	33	3	2	-	-	-	177	407

2505.000.046-8 - ORIENTAÇÃO E ESTÁGIO CURRICULAR SUPERVISIONADO II	167	3	1	1	-	-	-	172	720
3001.000.097-8 - TERMODINÂMICA	128	29	14	1	-	-	-	172	544
2004.000.054-5 - LÍNGUA PORTUGUESA IV	144	18	7	2	-	-	-	171	372
2006.000.013-3 - LITERATURA BRASILEIRA I	141	24	4	1	-	-	-	170	398
1504.000.030-1 - MICROBIOLOGIA	137	23	5	2	-	-	-	167	218
3007.000.095-8 - QUÍMICA ORGÂNICA II	156	8	2	1	-	-	-	167	247
2505.000.043-3 - ORIENTAÇÃO E ESTÁGIO CURRICULAR SUPERVISIONADO I	150	14	1	1	-	-	-	166	684
2006.000.053-2 - TEORIA DA LITERATURA III	139	23	2	1	-	-	-	165	438
3005.000.059-6 - TERMODINÂMICA	108	49	7	1	-	-	-	165	317
2504.000.017-1 - HISTÓRIA DA QUÍMICA	150	5	6	1	-	-	-	162	367
2006.000.041-9 - LITERATURAS AFRICANAS DE LÍNGUA PORTUGUESA	155	5	-	1	-	-	-	161	382
1001.000.045-0 - FISILOGIA DO ESPORTE E DO EXERCÍCIO II	104	38	15	4	-	-	-	161	363
2006.000.014-1 - LITERATURA BRASILEIRA II	143	13	1	3	-	-	-	160	396
3002.000.007-5 - QUÍMICA ANALÍTICA APLICADA À ENGENHARIA AMBIENTAL	107	44	6	3	-	-	-	160	407
2003.000.078-0 - O SABER HISTÓRICO NA SALA DE AULA	155	2	-	1	-	-	-	158	306
2003.000.034-8 - ESTUDOS BRASILEIROS CONTEMPORÂNEOS	144	9	2	1	-	-	-	156	351
2009.000.013-1 - AVALIAÇÃO E MONITORAMENTO EM POLÍTICAS PÚBLICAS	151	2	1	1	-	-	-	155	252
2503.000.065-9 - GEOMETRIA PLANA	131	16	3	2	-	-	-	152	371
2004.000.069-3 - LINGÜÍSTICA V	102	32	9	1	-	-	-	144	379
2003.000.079-8 - O USO DE TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO NA APRENDIZAGEM DE HISTÓRIA	136	6	1	1	-	-	-	144	306
1010.000.043-3 - TECNOLOGIA ASSISTIVA II	135	8	-	1	-	-	-	144	360
3005.000.031-6 - DINÂMICA	112	29	2	1	-	-	-	144	265
2504.000.015-5 - FUNDAMENTOS DE QUÍMICA II	92	26	11	9	-	-	-	138	373
2504.000.019-8 - LABORATÓRIO DE QUÍMICA	118	15	3	1	-	-	-	137	369
2003.000.085-2 - SEMINÁRIOS DE HISTÓRIA	127	7	1	2	-	-	-	137	306
2009.000.017-4 - ESTÁGIO SUPERVISIONADO EM SERVIÇO SOCIAL I	123	11	2	1	-	-	-	137	234
2004.000.031-6 - ESTUDOS ORIENTADOS COMPLEMENTARES IV	124	9	2	2	-	-	-	137	471
1010.000.028-0 - INTERVENÇÕES EM TERAPIA OCUPACIONAL: FASE ADULTA E VELHICE II	132	2	-	1	-	-	-	135	354

2004.000.078-2 - ORIENTAÇÃO E ESTÁGIO SUPERVISIONADO EM LÍNGUA PORTUGUESA II	120	12	1	1	-	-	-	134	366
2002.000.032-9 - GEOGRAFIA POLÍTICA	121	8	-	1	-	-	-	130	386
2004.000.033-2 - ESTUDOS ORIENTADOS COMPLEMENTARES VI	105	16	5	3	-	-	-	129	408
2002.000.057-4 - REGIONALIZAÇÃO DO ESPAÇO BRASILEIRO: CENTRO-SUL	105	19	3	1	-	-	-	128	421
2002.000.012-4 - CARTOGRAFIA SISTEMÁTICA	119	6	-	1	-	-	-	126	249
2007.000.185-0 - SEMINÁRIOS DE PESQUISA III - TCC	124	-	1	1	-	-	-	126	180
2007.000.139-6 - PSICOLOGIA	121	3	1	1	-	-	-	126	241
2009.000.082-4 - POLÍTICA SOCIAL E SERVIÇO SOCIAL I	114	5	5	1	-	-	-	125	400
2009.000.026-3 - ESTUDOS E DESENVOLVIMENTO DE PROJETOS IV	117	4	-	1	-	-	-	122	194
1001.000.100-7 - PLANEJAMENTO E MARKETING ESPORTIVO	113	3	2	1	-	-	-	119	282
2002.000.044-2 - HISTÓRIA DO PENSAMENTO GEOGRÁFICO	105	13	-	1	-	-	-	119	242
2002.000.058-2 - REGIONALIZAÇÃO DO ESPAÇO BRASILEIRO: NORDESTE/NORTE	113	4	1	1	-	-	-	119	324
3004.000.014-3 - CIRCUITOS ELÉTRICOS I	71	38	2	1	-	-	-	112	302
2002.000.016-7 - CLIMATOLOGIA II	106	4	-	1	-	-	-	111	216
2504.000.031-7 - QUÍMICA ORGÂNICA I	94	11	3	1	-	-	-	109	382
2002.000.037-0 - GEOLOGIA II	97	9	1	1	-	-	-	108	181
2009.000.025-5 - ESTUDOS E DESENVOLVIMENTO DE PROJETOS III	103	2	1	1	-	-	-	107	148
2501.000.028-9 - CÉLULA E TECIDOS VEGETAIS	45	42	13	6	-	-	-	106	408
1001.000.050-7 - FUNDAMENTOS DA GINÁSTICA II	92	11	1	1	-	-	-	105	201
2504.000.023-6 - QUÍMICA ANALÍTICA QUALITATIVA	79	17	7	1	-	-	-	104	376
2002.000.014-0 - CIRCULAÇÃO E TRANSPORTES	87	9	3	1	-	-	-	100	324
3004.000.015-1 - CIRCUITOS ELÉTRICOS II	76	19	3	2	-	-	-	100	279
2502.000.037-0 - FÍSICA I	69	19	4	1	-	-	-	93	392
2505.000.024-7 - ESTUDOS E DESENVOLVIMENTO DE PROJETOS IV	84	5	2	1	-	-	-	92	472
2002.000.065-5 - TEORIAS DA REGIÃO	77	10	3	1	-	-	-	91	252
2002.000.050-7 - ORIENTAÇÃO E ESTÁGIO CURRICULAR SUPERVISIONADO I	83	3	2	1	-	-	-	89	252
2503.000.058-6 - GEOMETRIA ANALÍTICA	53	18	13	3	-	-	-	87	362
2505.000.104-9 - ESTUDOS E DESENVOLVIMENTO DE PROJETOS II	74	9	1	1	-	-	-	85	222
1001.000.117-1 - TRABALHO DE CONCLUSÃO DE CURSO - TCC I	70	9	4	1	-	-	-	84	180

2006.000.030-3 - LITERATURA EM LÍNGUA INGLESA III	73	8	-	1	-	-	-	82	182
2003.000.092-5 - TERRA BRASIL: COLONIZAÇÃO PORTUGUESA NA AMÉRICA	73	5	2	1	-	-	-	81	144
2505.000.071-9 - SABERES DO DOCENTE EM FÍSICA	70	7	3	1	-	-	-	81	363
2003.000.114-0 - TRABALHO DE CONCLUSÃO DE CURSO II	59	14	6	1	-	-	-	80	207
2504.000.024-4 - QUÍMICA ANALÍTICA QUANTITATIVA	55	11	5	2	-	-	-	73	288
2505.000.015-8 - ESTÁGIO CURRICULAR SUPERVISIONADO IV	67	5	-	1	-	-	-	73	216
2505.000.042-5 - ORIENTAÇÃO AO ESTÁGIO CURRICULAR SUPERVISIONADO IV	67	5	-	1	-	-	-	73	216
2005.000.018-1 - LÍNGUA ESPANHOLA IV	61	6	-	1	-	-	-	68	184
2505.000.011-5 - ESTÁGIO CURRICULAR SUPERVISIONADO II	54	4	2	2	-	-	-	62	216
2009.000.108-1 - SUPERVISÃO ACADÊMICA DE ESTÁGIO CURRICULAR I	60	1	-	1	-	-	-	62	147
2009.000.018-2 - ESTÁGIO SUPERVISIONADO EM SERVIÇO SOCIAL I	59	1	-	1	-	-	-	61	189
2505.000.038-7 - ORIENTAÇÃO AO ESTÁGIO CURRICULAR SUPERVISIONADO II	52	4	2	2	-	-	-	60	216
2503.000.080-2 - TEORIA DOS NÚMEROS	38	15	2	1	-	-	-	56	324
2503.000.078-0 - PROBABILIDADE E ESTATÍSTICA II	39	9	2	1	-	-	-	51	400
2503.000.082-9 - TRABALHO DE CONCLUSÃO DE CURSO II	36	5	1	1	-	-	-	43	360
2503.000.038-1 - CÁLCULO DIFERENCIAL E INTEGRAL II	28	9	-	2	-	-	-	39	221
2004.000.010-3 - ATIVIDADES ACADÊMICO-CIENTÍFICO-CULTURAIS I	26	3	2	2	-	-	-	33	240
2504.000.022-8 - QUÍMICA ANALÍTICA QUALITATIVA	17	10	1	3	-	-	-	31	137

¹Matrí. 1 = Quantidade de alunos que foram aprovados na primeira matrícula ou estão cursando a primeira matrícula.

²Matrí. 2 = Quantidade de alunos que foram aprovados na segunda matrícula (tiveram 1 reprovação e após isso foram aprovados) ou estão cursando a segunda matrícula.

³Matrí. 3 = Quantidade de alunos que foram aprovados na terceira matrícula (tiveram 2 reprovações e após isso foram aprovados) ou estão cursando a terceira matrícula.

⁴Matrí. 4 = Quantidade de alunos que foram aprovados na quarta matrícula (tiveram 3 reprovações e após isso foram aprovados) ou estão cursando a quarta matrícula.

⁵Matrí. 5 = Quantidade de alunos que foram aprovados na quinta matrícula (tiveram 4 reprovações e após isso foram aprovados) ou estão cursando a quinta matrícula.

⁶Matrí. 6 = Quantidade de alunos que foram aprovados na sexta matrícula (tiveram 5 reprovações e após isso foram aprovados) ou estão cursando a sexta matrícula.

⁷Matrí. 7 = Quantidade de alunos que foram aprovados na sétima matrícula (tiveram 6 reprovações e após isso foram aprovados) ou estão cursando a sétima matrícula.