

Reduzindo Reprovação e Evasão
no IFPB através de Escolhas
Guiadas de Disciplinas

Edital nº 11/2018 Programa Institucional
de Bolsas de Iniciação Científica -
PIBIC/CNPq - Edital de Pesquisa

Em
execução

Selecionado em
18/07/2018
23:00:00

Visualizar

Imprimir

Acompanhar Validação da Execução

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA

PRÓ-REITORIA DE PESQUISA, INOVAÇÃO E PÓS-GRADUAÇÃO

DIRETORIA DE PESQUISA

PROJETOS DE PESQUISA/INOVAÇÃO

Edital nº 11/2018 Programa Institucional de Bolsas de Iniciação Científica – PIBIC/CNPq

1 - UNIDADE PROPONENTE

Campus: CAMPUS-CG
Foco Tecnológico:

2 - IDENTIFICAÇÃO DO PROJETO

Título do Projeto: Reduzindo Reprovação e Evasão no IFPB através de Escolhas Guiadas de Disciplinas	
Grande Área de Conhecimento: CIÊNCIAS EXATAS E DA TERRA	Área de Conhecimento: CIÊNCIA DA COMPUTAÇÃO
Área Temática: Tecnologia da Informação e Comunicação	Tema: None
Período de Execução: Início: 01/08/2018 Término: 31/07/2019	

3 - CARACTERIZAÇÃO DOS BENEFICIÁRIOS

Público Alvo	Quantidade
--------------	------------

4 - EQUIPE PARTICIPANTE

PROFESSORES E/OU TÉCNICOS ADMINISTRATIVOS DO IFPB			
Membro	Contatos	Vínculo	Titulação
Nome: Anderson Fabiano Batista Ferreira da Costa Matrícula: 1526148	Tel.: (83) 8832-3373 E-mail: anderson@ifpb.edu.br	Voluntário	DOCTORADO
Nome: Gustavo Wagner Diniz Mendes Matrícula: 2254024	Tel.: E-mail: gustavo.mendes@ifpb.edu.br	Voluntário	MESTRE+RSC-III (LEI 12772/12 ART 18)

5 - DISCRIMINAÇÃO DO PROJETO

<p>Introdução</p> <p>Reprovação e evasão estudantis estão intimamente relacionados. De acordo com estudo feito por Ruben Klein da Fundação Cesgranrio, com dados do Censo Escolar 2016, 80% dos estudantes que evadiram repetiram uma série pelo menos uma vez (G1 2018). Essa repetência custa, aos cofres públicos, R\$ 16 bilhões, o que representa 8% do gasto em educação no Brasil</p>

em 2016, e é três vezes maior do que ocorre em países desenvolvidos (Censo Escolar 2016). De acordo com o PISA (Programa Internacional de Avaliação de Alunos, 2015), o Brasil apresenta os maiores índices do mundo de reprovação, perdendo, na América Latina, apenas para a Colômbia. O principal efeito da repetência é a evasão, e como, infelizmente, tal situação não vem equipada de políticas que auxiliem os alunos a enfrentarem as dificuldades que estão vivendo, a reprovação normalmente é associada apenas à punição, e não a um momento de reflexão e de aprendizagem.

No estudo que realizamos no ano passado, no IFPB Campus Campina Grande (IFPB-CG), no projeto “Previsão Automática de evasão estudantil nos cursos do IFPB”, detectamos que, levando em consideração todos os cursos do IFPB-CG, 35% de todos os alunos que realizaram alguma matrícula evadiram, e nos cursos de Informática o percentual foi de 40%.

Esses números corroboram com o estudo realizado por Gigioli (2016) sobre evasão em instituições federais de ensino superior no Brasil, que mostra que a evasão média varia, dependendo do curso e da instituição, de 30% a 40%.

O resultado prático dessa alta taxa de evasão é que milhares de alunos deixam de se formar em seus cursos, interrompem seus sonhos e de seus familiares, dificultam as possibilidades de escolha de uma carreira promissora, além de deixarem o mercado de trabalho mais escasso.

As instituições de ensino que perdem esses alunos precisam manter uma estrutura física montada, tais como laboratórios e salas de aula, para uma quantidade específica de alunos, e a evasão causa, além do custo social, um dano financeiro enorme, já que alunos evadidos representa ociosidade de espaços e recursos não utilizados.

Como exemplo dos altos custos da evasão, considerando uma turma de 40 alunos, se 10 deles evadirem do curso no primeiro semestre, o IFPB-CG deixará de arrecadar anualmente R\$ 444.000 e, durante os três anos de um curso de Informática, em torno de 1.5 milhões de reais, levando em consideração que um aluno, em média, custa R\$ 3.700 mensais ao IFPB-CG.

Diversas medidas podem ser tomadas na tentativa de mitigar a evasão, tais como: recepção adequada dos alunos; implementação de projetos de tutoria; realização de eventos; e pesquisas científicas, inserindo os alunos em um contexto mais amplo da instituição. Porém, independentemente dessas medidas, é importante conhecermos melhor os possíveis evasores, para aumentar a possibilidade de agirmos antes que a evasão ocorra.

Foi o que fizemos no projeto “Previsão Automática de evasão estudantil nos cursos do IFPB”, do Interconecta 2017, cujo objetivo principal foi detectar automaticamente evasores, com o intuito de sugerir medidas a serem consideradas pelos gestores, pelos coordenadores de curso e pela direção de ensino, antes que a evasão ocorresse. Enquanto os discentes se encontrarem nas instalações da instituição de ensino, políticas ainda podem ser adotadas para tentar reverter uma possível evasão deste estudante. Utilizando algoritmos de aprendizagem de máquina (Geitgey 2016), conseguimos detectar alunos que iriam evadir com acurácia acima de 83%.

Essa abordagem de detecção automática representa um grande diferencial à abordagem manual de análise de evasores, pois os algoritmos já detectam, no meio de milhares de dados, qual a parte da população à qual devemos despender esforços, e sinaliza aos gestores com quais alunos devemos trabalhar. Se fôssemos trabalhar com esses dados de forma manual, provavelmente não conseguiríamos detectar padrões na velocidade que o tema exige.

Apesar dessa abordagem automática já ser um grande avanço para o tratamento de evasões, ainda não estamos lidando com as suas causas. Como visto no Censo Escolar 2016, mais de 80% dos alunos que evadem são repetentes. O ideal seria termos acesso a todos os dados de todos os discentes que já cursaram disciplinas num determinado curso e, dado um conjunto de disciplinas possíveis do discente cursar num semestre, alguém informar qual a melhor configuração de disciplinas que maximize a probabilidade de conclusão com a maior nota possível, evitando ou diminuindo, assim, a repetência e, conseqüentemente, a evasão. Porém, mesmo que tenhamos acesso a esses dados, analisarmos manualmente não daria tempo hábil de dar as respostas necessárias aos alunos.

Justificativa

Atualmente, os sistemas acadêmicos geram uma grande quantidade de registros, tais como notas de disciplinas por aluno, disciplinas já cursadas, coeficiente de rendimento acadêmico (CRA) entre outros. Visando identificar e diminuir a evasão estudantil, diversas técnicas vêm sendo aplicadas com resultados bastante motivadores. Tais técnicas se baseiam em análise automatizada desses registros acadêmicos, diminuindo, assim, o problema temporal da análise manual.

Utilizando estatística e técnicas de inteligência artificial, os algoritmos de Aprendizagem de Máquina (Machine Learning) vêm sendo utilizados em diversos campos para análise de dados. Desde sugestões de compras na Amazon.com até escolha de filmes para assistir na Netflix, algoritmos de Aprendizagem de Máquina têm se mostrados ideais para análise de uma grande quantidade de dados a fim de identificar padrões de comportamento e informações que, a priori, não poderiam ser deduzidas pelo ser humano.

Melo et al. (2016) aplicaram algoritmos de Aprendizagem de Máquina nos registros de histórico acadêmico de todos os cursos da UFCG e conseguiram classificar, com uma acurácia de mais de 80%, se um aluno iria ou não evadir.

Com abordagem semelhante, um dos autores deste trabalho desenvolveu o projeto

“Previsão Automática de evasão estudantil nos cursos do IFPB” 2 durante o Interconecta 2017 e conseguiu, com acurácia maior que 83%, detectar possíveis evasores nos cursos de Informática do IFPB-CG.

Nazareno (2015) utilizou o processo KDD (do inglês Knowledge Discovery in Databases), ou Descoberta de Conhecimento em Base de Dados, para identificar, dado o conjunto de disciplinas escolhidas por um aluno num determinado semestre no curso de Ciência da Computação da UFCG, se a escolha foi boa ou se há uma grande probabilidade do aluno repetir pelo menos uma disciplina. A abordagem desta proposta de projeto é semelhante à aplicada por Nazareno (2015).

O problema de escolha de disciplina é, basicamente, um problema de agrupamento: identificar as disciplinas que, ao serem escolhidas, aumentam a chance de o aluno cursá-las com notas altas, sem reprovação. O uso de algoritmos de Aprendizagem de Máquina no

contexto de evasão tem o intuito de não somente identificar tais disciplinas, mas, principalmente, reduzir o custo de tal evasão, que, em pesquisa realizada em 2016, na Universidade de São Paulo (USP), resultou em 2 mil reais por aluno, por mês (Gigioli 2016).

Este trabalho tem o intuito de dar continuidade ao trabalho já iniciado no projeto “Previsão Automática de evasão estudantil nos cursos do IFPB”, porém com um viés diferente,

qual seja, auxiliar os alunos e gestores na escolha da melhor configuração de disciplinas a ser cursada num determinado semestre, aplicando algoritmos de aprendizagem de máquina dentro do contexto do processo de KDD. A proposta deste projeto é disponibilizar, ao discente, uma ferramenta a ser usada no momento das matrículas (por exemplo, no futuro, dentro do SUAP),

e, ao se escolher um conjunto de disciplinas, os algoritmos sugerem se é uma boa escolha, ou se o aluno deve revê-la, apresentando um cenário de disciplinas alternativo.

Em um segundo momento, os gestores dos cursos podem usar esses mesmos resultados para adaptarem, eventualmente, suas grades curriculares para prover uma melhor configuração de disciplinas. Além disso, visamos expandir, no futuro, os resultados aplicados para os demais campi do IFPB.

De maneira sumarizada, as principais contribuições oriundas deste trabalho são:

- Análise automática da melhor configuração de disciplinas para um aluno cursar em cursos semestrais do IFPB-CG, a partir da aplicação do processo KDD e de algoritmos de Aprendizagem de Máquina, visando diminuir o tempo para conclusão do curso e aumentar o CRA;
- Definição de uma metodologia sistemática para escolha de disciplinas a serem cursadas por alunos; e
- Indiretamente, este trabalho poderá reduzir a quantidade de evasores a partir das escolhas guiadas de disciplinas ao reduzir reprovações;

Fundamentação Teórica

Neste trabalho, visamos aplicar o processo de descoberta de conhecimento em banco de dados (KDD) com algoritmos de Aprendizagem de Máquina na fase de mineração de dados.

KDD – Knowledge Discovery in Databases

Cada vez mais os sistemas armazenam uma grande quantidade de dados, tornando difícil ao ser humano a tarefa de revisar e obter informações a partir desses dados. Como exemplo, a base de dados do sistema QAcadêmico, principal sistema de controle acadêmico do IFPB até 2016, tem mais de 27GB de dados. Quando somado à base do sistema SUAP, a dificuldade aumenta ainda mais, passando dos 30GB.

Essa dificuldade de se extrair informações úteis desses dados fez surgir ferramentas e técnicas na área de mineração de dados. KDD, ou descoberta de conhecimento em banco de dados, é um processo para transformar dados em informação útil. O produto final desse processo é a obtenção de conhecimento útil, a partir de um extenso conjunto de dados. As principais etapas do processo KDD são citadas abaixo (Fayyad 1996):

1. Compreensão do domínio da aplicação: entendimento dos objetivos que se deseja com o uso de mineração de dados. No nosso projeto, a redução da reprovação e consequente diminuição de evasão;
2. Pré-processamento:
 - a. Seleção de dados: recuperação de dados relevantes para a análise. Nesta etapa, serão extraídos dados do QAcadêmico e do SUAP, bem como escolheremos os dados mais relevantes a serem usados;
 - b. Limpeza de dados: remoção de ruídos e de dados inconsistentes. Como serão obtidas informações de duas bases de dados distintas, essa etapa será de suma importância;
 - c. Integração de dados: combinação de múltiplas fontes de dados. Nesta etapa, os dados obtidos das duas bases serão integrados em uma única base de dados.
3. Transformação de dados: conversão dos dados para um formato apropriado, facilitando a mineração;
4. Mineração de dados: fase essencial, em que algoritmos específicos são aplicados com o intuito de extrair padrões nos dados. Nesta fase, aplicaremos os algoritmos de Aprendizagem de máquina para encontrarmos padrões e sugerirmos a melhor configuração de disciplinas para um determinado semestre, dada a atual situação do aluno;
5. Interpretação e avaliação dos resultados: visualização, análise e consolidação do conhecimento.

Esse processo é aplicado de forma interativa, ou seja, passamos por essas etapas várias

vezes até obtermos o conhecimento pretendido. Na etapa da Mineração de Dados, são usados algoritmos e técnicas conhecidas do Aprendizado de Máquina para reconhecimento de padrões e estatísticas.

Aprendizado de Máquina

Aprendizado de Máquina se utiliza de métodos estatísticos, de Inteligência Artificial e do poder computacional disponível atualmente para encontrar padrões nos dados analisados. Ele é a base de aplicações amplamente utilizadas atualmente. Por exemplo, quando se faz uma pesquisa no Google pelo termo X, é informado que pessoas que pesquisaram por X também pesquisaram por Y e Z. Esse padrão encontrado só foi possível através da utilização de algoritmos de Aprendizado de Máquina.

Para explicar como Aprendizado de Máquina funciona na prática, daremos um exemplo adaptado de (Geitgey 2016). Imagine o problema de avaliar o preço de um imóvel.

Provavelmente, para saber o valor de um imóvel, precisaríamos chamar um corretor especialista, que iria especificar alguns atributos que representem o imóvel, tais como tamanho

total, quantidade de cômodos, quantidade de quartos e suítes, entre outros. O valor dado pelo especialista normalmente vem da sua experiência de vendas passadas, de imóveis similares previamente vendidos. Suponha, agora, a necessidade de se avaliar o preço de um apartamento, sem ajuda de um especialista, apenas com valores dos atributos dos imóveis já vendidos, bem como o valor da venda desses imóveis. Seria necessário encontrar um padrão, talvez encontrar um ou mais atributos que são mais relevantes para o preço do imóvel.

Perceba que o espaço de possibilidades é elevado, ou seja, a quantidade de atributos e suas variações é grande demais, pois demoraria uma considerável quantidade de tempo para testar todos os pesos dos atributos com todos os seus possíveis valores, que são infinitos. Para evitar esse problema do tempo de resposta, os matemáticos encontraram uma forma de achar rapidamente bons valores para os pesos, sem precisar testar todos. Simplificando, dada uma equação de custo, que representa o quão errada a nossa função de estimativa de valor de um imóvel está, procuramos achar o menor custo, aquele cujos pesos dos atributos sejam os mais próximos do preço real de venda, reduzindo assim o erro de nossa função de preço de imóveis.

O gráfico aproximado dessa equação é apresentado na Figura 1. Na prática, o que buscamos é descer o máximo o vale desta figura para chegar na área azul, que é onde se encontram os menores erros, ou seja, onde tem os valores dos pesos ideais para a fórmula matemática que procuramos. E é isso que os algoritmos de Aprendizado de máquina fazem.

Figura 1 - Gráfico da função de custo. O eixo y representa o custo. Quanto mais alto no gráfico, maior o custo.

Os algoritmos de Aprendizado de Máquina podem ser divididos em dois tipos:

supervisionados ou não supervisionados. Os algoritmos supervisionados são aqueles que aprendem a partir de um conjunto de dados de treinamento que já carregam respostas para o problema proposto. Por exemplo, se dermos a um algoritmo os dados de vendas de imóveis, com os atributos dos imóveis e o preço de venda, esse algoritmo certamente é supervisionado, pois ele foi treinado a partir de dados reais. Já os algoritmos não supervisionados são aqueles em que são passados dados, e se quer saber quais informações podem ser tiradas do conjunto de dados. Neste trabalho, utilizaremos algoritmos supervisionados.

Manhães et al. utilizaram uma técnica chamada Mineração de Dados Educacionais (Educational Data Mining), similar à utilizada na UFCG, para prever quais alunos estavam em risco de evasão (Manhães et al. 2014). Eles construíram uma arquitetura chamada WAVE, analisaram três cursos de graduação e usaram o classificador de Aprendizado de Máquina Naive Bayes, já que apresentou melhor taxa de verdadeiros positivos em relação a outros.

Utilizaremos técnicas similares à utilizada por Manhães et al. com enfoque à escolha de disciplinas.

Yadin estudou qual a melhor linguagem de programação para ser usada nos primeiros semestres em disciplinas de programação, com o intuito de reduzir a evasão (Yadin 2011).

Perceberam que linguagens orientadas a objetos apresentam grande complexidade para o primeiro contato de estudantes, utilizaram Python com a ferramenta visual Guido Van Robot e conseguiram reduzir de 45.8% de repetentes para 16.7%, reduzindo, por consequência, a quantidade de evasão. Essa é uma das medidas que podemos aplicar para reduzir a evasão após estudarmos seus efeitos no contexto do IFPB.

Neste trabalho, aplicaremos o processo de KDD com algoritmos de Aprendizado de Máquina com o intuito de guiar os alunos nas escolhas de disciplinas de seus cursos visando um menor tempo possível para conclusão do curso, com maior CRA. Para isso, queremos responder as seguintes perguntas:

-

Se duas disciplinas A e B estão correlacionadas, e o aluno passou em A com média X, qual a probabilidade de passar em B por média?

-

Qual a melhor seleção de disciplinas, num semestre específico, para que o aluno conclua seu curso no tempo mínimo esperado, com um bom CRA?

- Quais disciplinas que geram mais atrasos nos cursos?
- Qual a estimativa do CRA, dado um conjunto de notas já obtidas e disciplinas cursadas?
- Dado um perfil de disciplinas e notas de um aluno, qual a probabilidade desse aluno concluir o curso?

Para isso, utilizaremos os dados de registros acadêmicos de períodos passados para

servirem como entrada para os algoritmos de Aprendizagem de Máquina. O resultado final deste trabalho será um modelo de escolha de disciplinas a ser utilizado por alunos e pela alta administração, bem como por professores e coordenadores de cursos, para que possam realizar atividades preventivas com os alunos mais propensos à evasão, bem como entender melhor o efeito das escolhas de disciplinas num determinado semestre.

Objetivo Geral

Este trabalho tem como objetivo geral analisar e aplicar algoritmos de Aprendizagem de Máquina com o intuito de ajudar na escolha de disciplinas e reduzir reprovações e evasões discentes no contexto dos cursos do IFPB Campus Campina Grande na visão de alunos, administradores, professores e coordenadores de cursos.

Metas

- 1 - Estudar e realizar a seleção dos melhores atributos preditores de escolha de disciplinas.
- 2 - Utilizar diferentes algoritmos de classificação e medir a acurácia do resultado a fim de usar na previsão de evasão.
- 3 - Executar algoritmos supervisionados de Aprendizado de Máquina para prever disciplinas a serem cursadas.
- 4 - Apresentar os resultados aos administradores do IFPB para propor ações visando redução de evasão.

Metodologia da Execução do Projeto

Estudar e realizar a seleção dos melhores atributos que predizem a melhor escolha de disciplinas para um determinado semestre;

- Utilizar diferentes algoritmos de classificação e agrupamento e medir a acurácia do resultado a fim de usar na escolha de disciplinas;
- Executar algoritmos supervisionados de Aprendizagem de Máquina para auxiliar na escolha de disciplinas.
- Estudar e desenvolver projeto para inserção da análise automática no sistema de controle acadêmico do IFPB (SUAP)

Disseminação dos Resultados

1. Levantamento bibliográfico

Nesta etapa, realizaremos um estudo bibliográfico para levantar os principais trabalhos que aplicaram as técnicas de Aprendizagem de Máquina para termos uma base sólida para aplicação no projeto.

2. Análise de Algoritmos de Seleção de Atributos

Alguns atributos são melhores preditores que outros na escolha de disciplinas. Apesar de termos indícios de quais podemos escolher, precisaremos analisar, no âmbito do IFPB, quais melhor se aplicam para escolha de disciplinas (ex.: perfil do aluno, semestre sendo cursado, o CRA do aluno, quantidade de disciplinas reprovadas, entre outros).

3. Levantamento dos dados de registros acadêmicos do SUAP/QAcadêmico

Nesta etapa, serão recuperados os dados do SUAP/QAcadêmico referentes aos registros estudantis de semestres passados. Precisaremos definir a partir de qual ano pegaremos dados confiáveis para análise. Esta etapa é crítica, e será realizada com a ajuda da equipe de TI do IFPB que lida com os dados do SUAP/QAcadêmico.

4. Limpeza da base de dados recuperada

Nesta etapa, precisaremos transformar os dados recuperados na etapa anterior em um formato que os algoritmos de Aprendizagem de Máquina possam extrair informações úteis.

5. Confecção e Apresentação do Relatório Parcial

6. Execução de algoritmos de Aprendizagem de Máquina nos dados tratados

Nesta etapa, executaremos diferentes algoritmos de Aprendizagem de Máquina com o intuito de saber qual melhor se adequa à seleção de disciplinas. Precisaremos analisar a acurácia dada pelo algoritmo e, a partir desta análise, usaremos o algoritmo escolhido para análise de dados.

7. Análise dos dados obtidos

Nesta etapa, analisaremos os resultados obtidos e estudaremos alunos específicos para comprovar a acurácia obtida. O resultado desta atividade será discutido com coordenadores de cursos e administradores do IFPB-CG para levantarmos ações possíveis a serem aplicadas.

8. Estudo e projeto para inserir o resultado deste trabalho no controle acadêmico

Nesta etapa, analisaremos e faremos um projeto para inserir a análise automática de disciplinas no sistema de controle acadêmico do IFPB, o SUAP, para uso dos alunos durante o processo de matrícula.

9. Apresentação dos Resultados

Nesta etapa, apresentaremos os resultados obtidos em eventos e/ou semanas temáticas do IFPB. Também será desenvolvido um artigo técnico.

10. Confecção e Apresentação do Relatório Final

Referências Bibliográficas

1. Censo Escolar, 2016. INEP - Censo Escolar. <http://inep.gov.br/censo-escolar>. Acessado em fevereiro/2018.

2. Fayyad, Usama M. Piatetksky-Shapiro, Gregory. Smith, Padhraic. From data mining to knowledge discovery: an overview. 1996. Em *Advances in knowledge discovery and data mining*. American Association for Artificial Intelligence, Menlo Park, CA, USA 1-34
3. G1. Brasil gasta R\$16 milhões com reprovação de 3 milhões de alunos em 2016, aponta levantamento. 2018. <https://g1.globo.com/educacao/noticia/brasil-gasta-r-16-bilhoes-com-reprovacao-de-3-milhoes-de-alunos-em-2016-aponta-levantamento.ghtml>. Acessado em 03/2018.
4. Geitgey, Adam. 2016. Machine Learning is Fun! <https://medium.com/@ageitgey/machine-learning-is-fun-80ea3ec3c471#nwsccff5sb>, acessado em 02/2018.
5. Gigioli, Renato de Sousa Porto. 2016. Evasão em instituições federais de ensino superior no Brasil: Expansão da rede, SISU e desafios. Consultoria Legislativa: Câmara dos Deputados, Brasília.
6. Manhães, Laci, Sérgio. Zimbrão, Geraldo. 2014. WAVE: an Architecture for Predicting Dropout in Undergraduate Courses using EDM. *Anais da 29 o Symposium on Applied Computing*, ACM. Pág. 243-247. República da Korea.
7. Melo, Allan Sales da Costa. Balby, Leandro. Cajueiro, Adalberto. 2016. Previsão automática de evasão estudantil: um estudo de caso na UFCG. Campina Grande: UFCG. Dissertação de Mestrado. Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina.
8. Nazareno. 2015. Mineração de dados para apoio à gestão acadêmica na UFCG. Projeto PIBITI 2015. UFCG.
9. PISA-Programa Internacional de Avaliação de Alunos. 2015. <https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf>. Acessado em 02/2018.
10. Yadin, Aharon. 2011. Reducing the Dropout rate in introductory programming course. *ACM Inroads*, Volume 2, Issue 4. Pág. 71-76, Nova York.
11. Witten, Ian H. Frank, Eibe. Hall, Mark A. *Data Mining: Practical Machine Learning Tools and Techniques*. 3a ed. Burlington: Morgan Kaufmann, 2011.

6 - CRONOGRAMA DE EXECUÇÃO

Meta	Atividade	Especificação	Indicador(es) Qualitativo(s)	Indicador Físico		Período de Execução	
				Unid.de Medida	Qtd.	Início	Término
1	1	Levantamento bibliográfico: Nesta etapa, realizaremos um estudo bibliográfico para levantar os principais trabalhos que aplicaram as técnicas de Aprendizagem de Máquina para termos uma base sólida para aplicação no projeto.	Obter conhecimento sobre os principais trabalhos ligados à aplicação de Aprendizado de Máquina no contexto de evasão estudantil.		0	01/08/2018	30/09/2018
1	2	Análise de Algoritmos de Seleção de Atributos: Alguns atributos são melhores preditores que outros na escolha de disciplinas. Apesar de termos indícios de quais podemos escolher, precisaremos analisar, no âmbito do IFPB, quais melhor se aplicam para escolha de disciplinas. Já há alguns atributos que consideramos bons para a análise do perfil do aluno (situação atual dele no curso), tais como semestre em que um determinado aluno está cursando, o CRA do aluno, quantidade de disciplinas reprovadas, entre outros. Porém, executaremos diferentes algoritmos para analisar quais os que mais impactam a escolha de disciplinas.	Ter sido escolhido um algoritmo que melhor preveja as disciplinas que melhor se adequam ao semestre a ser cursado e perfil atual do aluno, analisando os dados do QAcadêmico em conjunto com os do SUAP.		0	03/09/2018	30/11/2018
2	1	Levantamento dos dados de registros acadêmicos do SUAP: Nesta etapa, serão recuperados os dados do SUAP/QAcadêmico referentes aos registros estudantis de semestres passados. Precisaremos definir a partir de qual ano pegaremos dados confiáveis para análise. Esta etapa é crítica, e será realizada com a ajuda da equipe de TI do IFPB que lida com os dados do SUAP/QAcadêmico.	Obter os dados de registros acadêmicos de anos anteriores do SUAP e QAcadêmico.		0	01/10/2018	30/11/2018
2	2	Limpeza da base de dados recuperada: Nesta etapa, precisaremos transformar os dados recuperados na etapa anterior em um formato que os algoritmos de Aprendizagem de Máquina possam extrair informações para prever evasão. Por exemplo, se não houver informação de qual grade curricular um determinado aluno se encontra, precisaremos analisar outros atributos para deduzir qual seu perfil no curso.	Base recuperada limpa, com os atributos a serem analisados.		0	03/12/2018	28/02/2019

Meta	Atividade	Especificação	Indicador(es) Qualitativo(s)	Indicador Físico		Período de Execução	
				Unid.de Medida	Qtd.	Início	Término
2	3	Confecção do Relatório Parcial: Nesta etapa, confeccionaremos um relatório parcial deste trabalho para avaliar os resultados obtidos.	Relatório confeccionado e enviado à Coordenação de Pesquisa.		0	01/03/2019	31/03/2019
3	1	Execução de algoritmos de Aprendizado de Máquina nos dados tratados: Nesta etapa, executaremos diferentes algoritmos de Aprendizagem de Máquina com o intuito de saber qual melhor se adequa à seleção de disciplinas. Precisaremos analisar a acurácia dada pelo algoritmo e a partir desta análise usaremos o algoritmo escolhido para análise de dados. Para tal, utilizaremos a linguagem Python com a biblioteca científica SciPy.	Base de dados de treinamento; Base de dados de análise; Scripts em Python dos algoritmos de Aprendizado de Máquina.		0	01/01/2019	28/06/2019
3	2	Análise dos dados obtidos: Nesta etapa, analisaremos os resultados obtidos e analisaremos alunos específicos para comprovar a acurácia obtida. O resultado desta atividade será discutido com coordenadores de cursos e administradores do IFPB-CG para levantarmos ações possíveis a serem aplicadas.	Análise dos resultados pelos administradores do IFPB; Lista de ações a serem tomadas.		0	01/03/2019	31/07/2019
4	1	Estudo e projeto para inserir o resultado deste trabalho no controle acadêmico: Nesta etapa, analisaremos e faremos um projeto para inserir a análise automática de disciplinas no sistema de controle acadêmico do IFPB, o SUAP, para uso dos alunos durante o processo de matrícula.	Projeto para inserção da metodologia e sistema construído no SUAP.		0	01/07/2019	31/07/2019
4	2	Apresentação dos Resultados: Nesta etapa, apresentaremos os resultados obtidos em eventos e/ou semanas temáticas do IFPB. Também será desenvolvido um artigo técnico.	Retorno dos resultados apresentados à administração; Apresentação realizada na semana de Teleinformática do IFPB-CG; Apresentação realizada no SIMPIF e/ou em outro evento; Rascunho de artigo científico mostrando os resultados obtidos.		0	03/06/2019	28/06/2019
4	3	Confecção do Relatório Final: Nesta fase, confeccionaremos o relatório final deste trabalho.	Relatório final confeccionado e entregue.		0	01/08/2019	30/08/2019

7 - PLANO DE APLICAÇÃO

Classificação da Despesa	Especificação	PROEX (R\$)	DIGAE (R\$)	Campus Proponente (R\$)	Total (R\$)
333018	Auxílio Financeiro a Estudantes	0	0	0	0
TOTAIS		0	0	0	0

8 - CRONOGRAMA DE DESEMBOLSO

Despesa	Mês 1	Mês 2	Mês 3	Mês 4	Mês 5	Mês 6	Mês 7	Mês 8	Mês 9	Mês 10	Mês 11	Mês 12
333018 - Auxílio Financeiro a Estudantes	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00	400.00

Anexo A

MEMÓRIA DE CÁLCULO

CLASSIFICAÇÃO DE DESPESA	ESPECIFICAÇÃO	UNIDADE DE MEDIDA	QUANT.	VALOR UNITÁRIO	VALOR TOTAL	ATIVO
TOTAL GERAL					4.800,00	

CLASSIFICAÇÃO DE DESPESA	ESPECIFICAÇÃO	UNIDADE DE MEDIDA	QUANT.	VALOR UNITÁRIO	VALOR TOTAL	ATIVO
333018 - Auxílio Financeiro a Estudantes	Bolsa para estudante.	bolsa	12	400.00	4800.00	Sim
TOTAL GERAL					4.800,00	