

Análise de Dados de Percursos Curriculares dos Alunos de Ciência da Computação da Universidade Federal da Paraíba

Luiz Fernando Fonseca Pinheiro de Lima, CI - UFPB, luizfernando@cc.ci.ufpb.br

Ingrid Luana Almeida da Silva, CI - UFPB, ingrid@cc.ci.ufpb.br

Danielle Rousy Dias da Silva, CI - UFPB, danielle@ci.ufpb.br

Resumo. Diferentemente das pesquisas que analisam dados educacionais de cursos superiores, que visam, sobretudo, o estudo de evasão, este trabalho se propõe a analisar dados acerca de percursos curriculares realizados por alunos que concluíram o curso de Ciência da Computação em uma instituição superior pública. O estudo tem como objetivo identificar melhorias no projeto pedagógico do curso e para o seu desenvolvimento foram definidas questões a responder, a base de dados utilizada para respondê-las e as ferramentas que deram suporte a construção das análises dos dados. Assim foram desenvolvidas análises que responderam as questões e trouxeram informação efetiva para o curso, como os índices de alunos que cursam as disciplinas no semestre recomendado pela matriz curricular, as distribuições das disciplinas por semestres diferentes do recomendado e as taxas de situações obtidas pelos alunos em suas matrículas nas disciplinas. Com essas análises foi possível identificar possíveis pontos de melhoria no projeto pedagógico do curso.

Palavras-chave: Análise de Dados Acadêmicos, Ciência de Dados Educacionais, Matriz Curricular, Percursos Curriculares.

Data Analysis of Curricular Pathways of Computer Science Students of the Federal University of Paraíba

Abstract. Unlike researches that analyzes educational data from higher education courses, which mainly aims at the study of dropout, this paper aims to analyze data about curricular pathways taken by students who completed the Computer Science course in a public higher education institution. The study aims to identify improvements in the pedagogical project of the course and for its development were defined questions to answer, the dataset used to answer them and the tools that supported the construction of the data analysis. Thus, analyzes were developed that answered the questions and brought effective information to the course, such as the rates of students who attend the subjects in the semester recommended by the curriculum, the distributions of subjects by semesters different from the recommended and the situation rates obtained by students in their enrollment in the subjects. With these analyzes it was possible to identify possible points of improvement in the pedagogical project of the course.

Keywords: Academic Analytics, Educational Data Science, Curriculum Matrix, Curricular Pathways.

1. Introdução

A chegada da internet e da grande evolução tecnológica tem proporcionado que, cada vez mais, empresas, governos e usuários produzam um massivo número de dados com a utilização de sistemas *web*, aplicativos para *smartphones* e dispositivos inteligentes como *smartwatches* e *smartbands*, criando, assim, o conceito de *Big Data* (CHEN et al., 2014).

Neste contexto, a Ciência de Dados surge exatamente como a área responsável por tratar e analisar os dados gerados pelas aplicações. Essa área é caracterizada por sua inter e multidisciplinaridade e tem como base áreas tradicionais como estatística, probabilidade e ciência da computação, mas pode ser aplicada a diversas áreas como indústria, saúde, jogos, política e educação (PORTO e ZIVIANI, 2014).

Segundo SILVA et al. (2017) a aplicação de técnicas de Ciência de Dados sobre dados educacionais tem se tornado usual, originando uma subárea própria, a Análise ou Ciência de Dados Educacionais, e ela pode ser entendida em três temáticas, a Mineração de Dados Educacionais (ou *Educational Data Mining*), a Análise de Aprendizado (ou *Learning Analytics*) e a Análise de Dados Acadêmicos (ou *Academic Analytics*).

A Análise de Dados Acadêmicos utiliza dados originados de sistemas educacionais administrativos ou de gestão acadêmica e aplica, sobre eles, técnicas de mineração de dados, como algoritmos de Aprendizado de Máquina, com objetivo de levantar conhecimento inerente aos dados, bem como técnicas de análise exploratória dos dados com objetivo de confirmar hipóteses ou questões levantadas.

Na Ciência de Dados Educacionais diversos trabalhos têm sido desenvolvidos com objetivo de prever, por exemplo, risco de evasão de alunos (BRITO et al., 2015) e desempenho de alunos (BRITO et al., 2014). Mas, além desses aspectos, diversos outros estudos podem ser realizados a partir de dados gerados pelos alunos em sua graduação. Um desses estudos é análise de efetividade das matrizes curriculares em relação à formação profissional dos estudantes, nesta análise pode-se identificar, por exemplo, a sequência e ordem de disciplinas mais comumente realizada pelos alunos.

Diante disto, esta pesquisa buscou analisar uma base de dados educacionais proveniente de discentes graduados no Bacharelado em Ciência da Computação (BCC) da Universidade Federal da Paraíba (UFPB) gerando, assim, informação que auxilie na tomada de decisão do problema real, enfrentado atualmente no curso, sobre a atualização de sua matriz curricular. Este trabalho pode ser classificado como exploratório, uma vez que esse tipo de estudo ainda não foi realizado e proporcionará a uma maior familiaridade ao problema em questão, e ainda como descritivo, já que tem como objetivo analisar e descrever as características dos dados dos alunos de BCC (KAUARK; MANHÃES; MEDEIROS, 2010). Este trabalho ainda é categorizado como um projeto de Ciência de Dados Educacionais, mais especificamente, na temática de Análise de Dados Acadêmicos, uma vez que os dados utilizados foram coletados a partir do sistema de gestão acadêmica da UFPB.

O artigo foi organizado em 5 seções. Na Seção 2 são apresentados os trabalhos relacionados. Na Seção 3 é descrito o processo de desenvolvimento do trabalho. Na Seção 4 são apresentados os resultados das análises realizadas. E na Seção 5 é apresentada a conclusão e trabalhos futuros.

2. Trabalhos Relacionados

BAEPLER e MURDOCH (2010) apresentam trabalhos desenvolvidos na área de Análise de Dados Acadêmicos pelo Sistema Universitário de Georgia e pela Universidade de Purdue. No trabalho produzido na Universidade de Georgia foi desenvolvido um modelo preditivo para taxas de conclusão e afastamento de alunos em ambientes *online* que conseguiu prever com 74% de acurácia a conclusão do curso *online* pelo estudante. Já nos trabalhos desenvolvidos na universidade de Purdue os pesquisadores desenvolveram uma ferramenta de alerta sobre desempenho de alunos, levando um rápido *feedback* que auxilia os alunos a entenderem como estão se saindo no curso e quando eles precisam dedicar mais atenção a suas atividades.

OLMOS e CORRIN (2012), os autores tiveram como objetivo estudar a excelência do currículo da Faculdade de Medicina da Universidade de Wollongong, visando extrair informações acerca dos dois principais desafios da graduação, a cobertura apropriada dos assuntos necessários no currículo e o engajamento dos alunos durante a formação nas experiências clínicas. Os autores reportaram que os resultados obtidos na análise ajudaram a universidade responder seus questionamentos sobre a excelência de seu currículo e que a aplicação de técnicas de visualização ajudaram a tornar os dados entendíveis para uma ampla audiência.

Também é possível observar os esforços para se desenvolver projetos de Análise de Dados Educacionais no Brasil, alguns desses trabalhos não se categorizam na temática de Análise de Dados Acadêmicos. Em (BRITO et al., 2015) os autores aplicaram técnicas de Mineração de Dados para identificação de alunos de BCC da UFPB com riscos de evasão, os modelos construídos classificaram corretamente os dados com acurácia a partir de 84%. SCHMITT, BERNARDI e KANTORSKI (2019) também trabalharam com gestão de evasão, os autores desenvolveram uma ferramenta com objetivo de facilitar o acesso de gestores de cursos de Educação a Distância às informações geradas a partir da mineração dos dados desse curso.

Já em (BRITO et al., 2014) foram desenvolvidos modelos preditivos sobre desempenho dos alunos, também, do curso de BCC da UFPB, todos os modelos criados no trabalho atingiram acurácia maior que 70% para o problema. Por fim, em (CARNEIRO et al., 2017) os autores executaram uma revisão sistemática sobre Aprendizagem Adaptativa em ambientes virtuais de aprendizagem, no trabalho foi possível identificar que se tem utilizado como estratégia para criação de personificações em ambientes virtuais de aprendizado modelos de agentes inteligentes e ontologias a partir dos dados gerados pelos alunos como interesses, competências e habilidades.

3. Estudo de Caso: Curso de BCC da UFPB

A Universidade Federal da Paraíba é uma instituição de ensino, pesquisa e extensão vinculada ao Ministério da Educação. O seu regimento geral apresenta sua estrutura organizacional e descreve o seu funcionamento (SODS, 2016).

Nesse regimento é descrita, por exemplo, a setorização da universidade em centros. A UFPB, atualmente, possui dezesseis centros, sendo cada um deles responsável por departamentos e cursos de uma área de conhecimento. O Centro de Informática é um dos centros da UFPB e é responsável pelos cursos de Bacharelado em Ciência da Computação, Engenharia da Computação e Matemática Computacional.

O curso de BCC da UFPB foi criado no ano de 1985. Após 34 anos, no semestre corrente da universidade, o curso conta com um total de 375 alunos distribuídos em diversos semestres (SIGAA, 2019a). Desde sua criação, houve uma única atualização da estrutura curricular do curso, ocorrida em 2006. Nessa estrutura os alunos precisam cursar uma carga horária de 3270 horas, o que equivale a 51 disciplinas (SIGAA, 2019b) em um prazo máximo de 12 semestres, mas é recomendado aos alunos um percurso de 8 semestres letivos para que se graduem.

O percurso recomendado deveria indicar aos alunos um melhor conjunto de disciplinas a serem cursadas em cada um dos semestres, contudo essa recomendação de fluxo nem sempre é seguida pelos discentes, que estabelecem múltiplos percursos diferentes do recomendado, por diversos motivos, como a necessidade de conciliar trabalho com a vida acadêmica, participação em laboratórios dentro da própria instituição, disciplinas com maior dificuldade, reprovações etc.

Analisando os percursos seguidos pelos alunos, diferentes do recomendado, pode-se identificar, por exemplo, conjuntos disciplinas inadequados e disciplinas pouco cursadas no semestre definido na matriz curricular, gerando evidências de um sequenciamento de disciplinas que pode ser mais adequado para os alunos seguirem no curso.

4. Processo de Desenvolvimento

Como processo e etapas de desenvolvimento para a produção desta pesquisa foi adotado o ciclo definido na proposta de AGARWAL (2018). Essa proposta descreve as sete etapas de projetos em Ciência de Dados. A primeira etapa tem como foco o entendimento do problema e o levantamento de questões a se responder. As segunda e terceira etapas são responsáveis pelas coleta e pré-processamento dos dados. A quarta etapa busca analisar de forma exploratória os dados a fim de identificar padrões entre eles. Na quinta etapa uma nova preparação dos dados é realizada para a etapa seguinte. A sexta etapa do ciclo está relacionada com a criação de modelos de Aprendizado de Máquina. A última etapa do processo tem como objetivo estudar as melhores formas de exibir os resultados obtidos. Ao final do ciclo as etapas podem se repetir, gerando iterações do processo.

O desenvolvimento deste trabalho focou na execução das quatro primeiras etapas do ciclo de vida de um projeto de Ciência de Dados. Dessa forma foram definidas questões que norteiam esta pesquisa (Seção 4.1), bem como a base de dados e o pré-processamento necessário (Seção 4.2) e as análises realizadas com objetivo de responder às questões levantadas (Seção 4.3). Também foram definidas quais as tecnologias foram utilizadas para desenvolver as análises (Seção 4.4).

4.1 Levantamento de Questões

Como mencionado na Seção 4, para a primeira etapa do projeto foi necessário entender o problema e levantar questões de investigação, essas são apresentadas no Quadro 1. As questões nortearam quais análises seriam desenvolvidas e quais dados seriam necessários para respondê-las.

Quadro 1 - Questões levantadas

Identificador	Questão
Q1	Quais disciplinas apresentam maior e menor índice de alunos que as cursaram no recomendado?
Q2	Como os alunos distribuem as disciplinas que não cursaram no recomendado?
Q3	Quais são as proporções das situações dos alunos na primeira tentativa de cursar as disciplinas?
Q4	Existe algum padrão de fluxo diferente do recomendado pela matriz curricular?

4.2 Base de Dados e Pré-Processamento

A base de dados utilizada neste trabalho foi originada a partir de dados cedidos pela Superintendência de Tecnologia da Informação da UFPB. Esses dados estavam contidos em três arquivos no formato CSV e representavam informações dos alunos que ingressaram no curso de BCC da UFPB desde 1985 a 2015 e que concluíram o curso até 2018. A partir dos dados foi elaborado o modelo da base de dados para análise

representar os discentes, outra os componentes curriculares e a última entidade para representar as matrículas nos componentes curriculares realizadas pelos estudantes.

A entidade “Discentes” contém os dados dos alunos que se graduaram no curso de BCC até 2018, seu arquivo relacionado possui o registro de 703 discentes. A entidade “ComponentesCurriculares” apresenta os dados acerca das disciplinas cursadas por pelo menos um aluno durante sua graduação, o arquivo referente a essa entidade apresenta 291 registros. Por fim, a entidade “Matrículas” representa a relação entre as duas entidades anteriores, ou seja, a relação de cada disciplina cursada por cada aluno, sua tabela relacionada possui 37768 instâncias de matrículas.

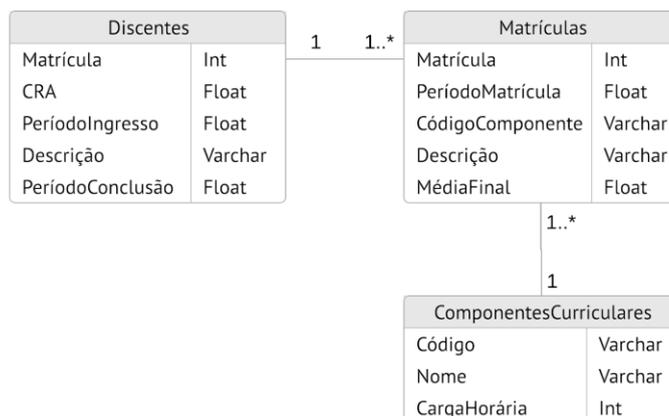


Figura 1. Diagrama da base de dados

Percebeu-se que para a execução das análises seguintes seria necessária a realização de três modificações na base de dados, especificamente, no arquivo da entidade “ComponentesCurriculares”, no qual foi notada a existência de três inconsistências: (i) disciplinas duplicadas com mesmo “Código”; (ii) disciplinas duplicadas com “Código” diferente; (iii) falta de campo de semestre recomendado.

As disciplinas duplicadas com mesmo “Código” apresentavam registros com valores diferentes apenas para o campo “Nome” (Figura 2) e para tratar essa inconsistência se levou em consideração apenas um dos registros como registro válido e os demais foram ignorados.

codigo	nome	ch_total
1107189	INTERACAO HOMEMMAQUINA	60
1107189	INTERAÇÃO HOMEM MAQUINA	60
codigo	nome	ch_total
1107190	IINTRODUÇÃO A COMPUTAÇÃO GRÁFICA	60
1107190	INTRODUÇÃO A COMPUTAÇÃO GRÁFICA	60

Figura 2. Exemplos de disciplinas duplicadas com mesmo “Código”

Alguns registros representavam a mesma disciplina, mas com o atributo “Código” diferente, devido à mudanças na instituição (Figura 3). Então, mais uma vez, foi considerado apenas um dos registros como registro válido, mas como essas instâncias de disciplinas possuíam códigos diferentes, foi necessário realizar a alteração dos códigos das disciplinas que foram substituídas pelo código da instância válida em

toda a base da entidade “Matrículas”, garantindo, assim, que nenhum registro de matrícula em qualquer disciplina fosse perdido.

Por fim, foi adicionada ao arquivo da entidade “ComponentesCurriculares” a coluna “Semestre”, que expressa o semestre indicado pela matriz curricular de BCC para se cursar a disciplina. Assim as disciplinas obrigatórias da matriz curricular de 2006 receberam, nessa coluna, um valor entre 1 e 8 que denota o semestre recomendado e as demais disciplinas não foram preenchidas, ficando com valor nulo nessa coluna para ser possível fazer a distinção entre essas disciplinas e as disciplinas obrigatórias.

codigo	nome	ch_total
1107149	LOGICA APLC A COMPUTACAO	60
codigo	nome	ch_total
1107200	LOGICA APLICADA A COMPUTACAO	60
codigo	nome	ch_total
1103176	CALCULO NUMERICO	60
codigo	nome	ch_total
1107114	CALCULO NUMERICO	60

Figura 3. Exemplos de disciplinas duplicadas com “Código” diferente

4.3 Análises Realizadas

Três análises foram desenvolvidas para gerar informação a partir dos dados obtidos. A primeira teve como objetivo analisar as proporções de alunos que cursaram as disciplinas no recomendado pela matriz curricular (Seção 4.3.1), após isso se quis observar como os alunos que não cursaram as disciplinas no recomendado distribuíram esses componentes pelos demais semestres (Seção 4.3.2), por fim se estudou as situações, como aprovação, reprovação e trancamento, dos alunos em suas primeiras tentativas de cursar as disciplinas (Seção 4.3.3).

Para a realização dessas análises foram utilizados apenas os dados dos alunos que ingressaram no curso a partir de 2006, uma vez que esse conjunto de discentes cursaram BCC com sua atual matriz curricular.

4.3.1 Índices de Alunos que Cursaram as Disciplinas no Recomendado (IACRs)

Para investigar Q1, foi necessário estudar quais as proporções de alunos que cursaram as disciplinas obrigatórias da matriz curricular seguindo sua recomendação e a proporção de alunos que não a seguiram.

Então, para a realização desse estudo, foi observado para cada aluno, cujo semestre de ingresso tenha sido igual ou superior a 2006.1, a ocorrência da matrícula em determinada disciplina, se esse aluno possuía o campo “Descrição” com valor igual à “APROVADO”, “DISPENSADO”, “APROVEITADO” ou “MATRICULADO”, já que os três primeiros indicam sucesso na disciplina e o último só poderia ser encontrado em alunos concluintes no ano de 2018.

Com esse registro identificado, foi calculado o semestre no qual esse aluno cursou a disciplina, a partir da diferença entre o campo “PeríodoMatrícula” desse registro e o campo “PeríodoIngresso” do registro do aluno. Então os resultados que apresentavam valor igual ao valor do campo “Semestre” da disciplina foram computados para a proporção de alunos que cursaram o componente no semestre recomendado, já os resultados com valores diferentes não foram computados no índice.

4.3.2 Distribuição dos Alunos Fora do Recomendado

Com o levantamento dos IACRs de cada disciplina foi possível identificar disciplinas nas quais os alunos de BCC tem maior disposição à cursá-las em semestres diferentes do que é recomendado pela matriz curricular do curso.

Então, após isso, para responder à Q2 e Q4, se buscou identificar a existência de padrões entre os alunos que não cursaram as disciplinas dentro do recomendado, ou seja, se buscou analisar como esses alunos distribuíram essas disciplinas pelos demais semestres. Desse modo, foram observados os registros de discentes que não cursaram determinada disciplina no recomendado e em qual semestre cada aluno a cursou, por fim foi calculada a proporção de alunos para cada semestre.

4.3.3 Situação dos Alunos na Primeira Tentativa de Cursar as Disciplinas

Outro estudo realizado com os dados obtidos foi a análise de situações, como aprovação, reprovação e trancamento, dos alunos na primeira matrícula nas disciplinas, seja no recomendado ou não. Essa análise teve como objetivo identificar algum padrão nessas situações e responder Q3, utilizando como referência o campo “Descrição” da tabela “Matrículas”.

4.4 Ferramentas Utilizadas

As análises foram desenvolvidas na linguagem de programação *Python*¹, em sua versão 3.6.8. Além de ser uma linguagem que proporciona um fácil entendimento e traz vantagens como o desenvolvimento rápido e a alta portabilidade entre sistemas operacionais, o *Python* possui ferramentas que suportam o desenvolvimento de projetos de Ciência de Dados (MCKINNEY, 2012).

Neste trabalho, além de bibliotecas padrão da linguagem (*statistics*, *math* e *os*), foram utilizadas as bibliotecas: *Numpy*², 1.16.4, para manipulação de arrays; *Pandas*³, 0.24.2, para manipulação de tabelas (ou *Dataframes*).

Todo o código fonte deste trabalho pode ser encontrado em repositório⁴ público.

5. Resultados e Discussões

Todas as análises descritas na Seção 4.4 foram realizadas para todos os registros da base de dados, mas serão apresentados os considerados mais relevantes, respondendo às questões da Quadro 1. Cada subseção seguinte é dedicada a uma dessas questões.

5.1 Quais disciplinas apresentam maior e menor índice de alunos que as cursaram no recomendado?

Os IACRs de todas as disciplinas foram obtidos por meio da análise descrita na Seção 4.4.2. Nessa análise pôde-se observar que as disciplinas recomendadas para o primeiro semestre são aquelas com maiores IACRs e as disciplinas com menores valores para o índice, menor que 20%, são Língua Inglesa I (ING), Modelos Probabilísticos para Computação (MOD), Séries e Equações Diferenciais Ordinárias (EDO) e Análise e Projeto de Algoritmos (APA).

Apesar de que os alunos egressos em BCC necessariamente devem se matricular nas disciplinas recomendadas para o semestre inicial, pode-se considerar que, para a base de dados dos alunos que concluíram o curso, essas disciplinas compõem um bom

¹ <https://www.python.org/>

² <https://numpy.org/>

³ <https://pandas.pydata.org/>

⁴ <https://github.com/limafernando/ADA-CC/>

Dentre as disciplinas com menores IACRs, ING, MOD e EDO são disciplinas recomendadas para o quarto semestre, o que pode indicar uma falha do conjunto de disciplinas nesse semestre.

Com a análise da Seção 4.4.2, ainda foi possível identificar o processo de diminuição em cadeia dos IACRs das disciplinas com pré-requisitos. Então, uma disciplina pode apresentar como IACR máximo o valor do IACR de sua disciplina definida como pré-requisito, já que os alunos aptos a cursarem a disciplina no recomendado são aqueles que cursaram o pré-requisito, também, no recomendado. Por exemplo, a disciplina Física Aplicada à Computação II (FAC II) tem como pré-requisito Física Aplicada à Computação I (FAC I), FAC I apresenta IACR de 78,5% de alunos, que seria o valor máximo para o IACR de FAC II, entretanto o índice para essa disciplina não passa de 56,9% dos alunos.

A disciplina Circuitos Lógicos (CL), que tem como pré-requisito definido FAC II, apresenta IACR de 68,9% dos alunos, maior que seu valor máximo de acordo com o que foi exposto no parágrafo anterior. Isso se deve à uma falha de inserção dos dados no sistema de gestão acadêmica, que permitiu que alguns alunos cursassem CL sem ter cursado anteriormente FAC II. Dessa forma deve-se avaliar se esse pré-requisito realmente se faz necessário, uma vez que uma parte dos discentes cursaram CL, com sucesso e no semestre recomendado, sem ter cursado FAC II.

5.2 Como os alunos distribuem as disciplinas que não cursaram no recomendado?

Com a análise descrita na Seção 4.4.3 foi observado em quais semestres os alunos, que não cursaram as disciplinas no recomendado, acabaram cursando os componentes curriculares.

Assim foram identificadas disciplinas com grande distribuição por semestres pelos alunos, como FAC I, Cálculo Diferencial e Integral II, ING, Especificação de Requisitos de Software (ERS), Administração de Empresas (ADM) e Interação Homem Máquina (IHM).

A disciplina FAC I, que é recomendada para o primeiro semestre, acaba sendo cursada até o décimo primeiro semestre, isso pode ser motivado pelo fato dos alunos considerarem essa disciplina como mais complexa e ela não interferir no aprendizado de disciplinas com conteúdo mais focado na área de computação.

Outras disciplinas como FAC II, Introdução à Computação Gráfica (CG), Cálculo Diferencial e Integral III (CALC III), Introdução ao Processamento Digital de Imagens (PDI), Cálculo Numérico (CALC NUM), MOD, EDO, APA, Construção de Compiladores (COMP), Redes de Computadores (RC), Introdução à Inteligência Artificial (IA), Computadores e Sociedade (C&S) e Engenharia de Software (ES) são cursadas, em média, até o décimo quarto semestre, valor que ultrapassa o limite permitido de semestres para que alunos de BCC concluam o curso, sendo necessária a abertura de processo de dilatação de tempo para conclusão.

5.3 Quais são as proporções das situações dos alunos na primeira tentativa de cursar as disciplinas?

Os resultados obtidos a partir da análise descrita na Seção 4.4.4 respondem a esta questão. As disciplinas com maior taxa de aprovação são as recomendadas para o sétimo semestre, C&S, ES e Sistemas de Informação e Decisão, todas elas apresentam taxa de aprovação maior que 90%. A disciplina com maior índice de reprovação é CG, na qual 19,3% dos alunos são reprovados em sua primeira tentativa de cursá-la. Para a

taxa de trancamento, CALC III é a disciplina que apresenta maior valor, na qual 12,3% dos alunos a trancam.

É importante notar que ambas as disciplinas com maiores taxas de reprovação e trancamento são disciplinas recomendadas para o terceiro semestre, o que pode indicar alguma falha nesse conjunto de disciplinas.

5.4 Existe algum padrão de fluxo diferente do recomendado?

Com as análises realizadas não foi possível identificar um único fluxo, diferente do recomendado, dentre os alunos. Mas pôde-se observar padrões por disciplina ou semestre, que indicam potenciais melhoramentos no projeto pedagógico do curso.

Por exemplo, as disciplinas que são cursadas em semestres além do permitido para os alunos de BCC, como FAC II, CG, CALC III, PDI, CALC NUM, MOD, EDO, APA, COMP, RC, IA, C&S e ES. Essa característica deve ser investigada pela coordenação de BCC, já que, para que elas possam ser cursadas além do limite, é necessária a abertura processo para dilatar o prazo máximo para conclusão do curso.

Outro padrão foi identificado em disciplinas com grande distribuição por semestre pelos alunos que não as cursaram no recomendado, mas que possuem boas taxas de aprovação, esse conjunto de disciplinas é composto por ING, ERS, ADM, IHM e C&S. Desse modo, é possível dizer que mesmo que os semestres recomendados para cada uma dessas disciplinas seja alterado, o desempenho dos alunos nelas pode não ser prejudicado.

Por fim, foi possível identificar que em um mesmo semestre, terceiro semestre, são recomendadas as disciplinas com maior taxa de reprovação e trancamento. Além disso, três das disciplinas com menores IACRs, abaixo de 20%, são recomendadas para o quarto semestre. Desse modo as análises apontam oportunidades de melhoria no projeto pedagógico do curso considerando as disciplinas planejadas para esses semestres.

6. Conclusões

Este trabalho teve como objetivo obter informações efetivas acerca de percursos curriculares traçados, diferentes do recomendado pela matriz curricular, a partir dos dados dos alunos do Bacharelado em Ciência da Computação da Universidade Federal da Paraíba que se graduaram até 2018.

Focando no conjunto de alunos que concluíram BCC com a atual matriz curricular, o trabalho se propôs a realizar análises que identificaram os índices de alunos que cursaram as disciplinas no semestre recomendado, como os alunos que não cursaram os componentes curriculares no recomendado distribuíram essas disciplinas pelos demais semestres do curso. Além disso, também foram analisadas quais as situações obtidas pelos alunos nas primeiras matrículas nas disciplinas.

Com essas análises foi possível responder as questões que haviam sido definidas para nortear o desenvolvimento deste trabalho. Pôde-se identificar quais disciplinas possuem maiores e menores IACRs, quais são mais e menos postergadas pelos alunos e quais apresentam maiores taxas de aprovação, reprovação e trancamento, levantando informações relevantes para a reformulação do projeto pedagógico do curso de BCC.

Contudo, há ainda outras análises a serem feitas. Desse modo são definidos como trabalhos futuros: (i) realizar análises comparativas entre a matriz curricular inicial de BCC e a atual; (ii) coletar e analisar os dados dos alunos que evadiram o curso; (iii) desenvolver uma ferramenta de recomendação de disciplinas; (iv) replicar as análises realizadas neste trabalho para as matrizes curriculares de outros cursos.

que possam auxiliar o acompanhamento da qualidade das propostas de projeto pedagógico em execução no curso, ponto este ainda pouco explorado pelas pesquisas atuais. Todas as análises realizadas podem ser replicadas para qualquer curso que deseje conhecer melhor quais os percursos reais de seus alunos e identificar oportunidades de adaptabilidade dos projetos pedagógicos.

Referências

- AGARWAL, S. **Understanding the Data Science Lifecycle**. 2018. Disponível em: <<http://sudeep.co/data-science/Understanding-the-Data-Science-Lifecycle>>. Acesso em: 11 set. 2019.
- BAEPLER, P.; MURDOCH, C.J. Academic analytics and data mining in higher education. **International Journal for the Scholarship of Teaching and Learning**, v. 4, n. 2, p. 17, 2010.
- BRITO, D.M. et al. Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2014. p. 882.
- BRITO, D.M. et al. Identificação de estudantes do primeiro semestre com risco de evasão através de técnicas de Data Mining. **Nuevas Ideas en Informática Educativa TISE**, p. 459-463, 2015.
- CARNEIRO, J.; SILVA, D.; SIEBRA, C. An analysis on student profiles in learning virtual environments. In: **2017 Twelfth Latin American Conference on Learning Technologies (LACLO)**. IEEE, 2017. p. 1-8.
- CHEN, M.; MAO, S.; LIU, Y. Big data: A survey. **Mobile networks and applications**, v. 19, n. 2, p. 171-209, 2014.
- KAUARK, F.S.; MANHÃES, F.C.; MEDEIROS, C.H. **Metodologia da pesquisa: um guia prático**. 2010.
- MCKINNEY, W. **Python for data analysis: Data wrangling with Pandas, NumPy, and IPython**. " O'Reilly Media, Inc.", 2012.
- OLMOS, M.; CORRIN, L. Academic analytics in a medical curriculum: Enabling educational excellence. **Australasian Journal of Educational Technology**, v. 28, n. 1, 2012.
- PORTO, F.; ZIVIANI, A. Ciência de dados. In: **III Seminário de Grandes Desafios da Computação no Brasil**, Rio de Janeiro, RJ, 2014.
- SILVA, L.A. et al. Ciência de Dados Educacionais: definições e convergências entre as áreas de pesquisa. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**. 2017. p. 764.
- SCHMITT, Jáder Adiel; BERNARDI, Giliane; KANTORSKI, Gustavo Zanini. iPermanency: uma ferramenta para gestão da evasão na EaD baseada em Mineração de Dados Educacionais. **RENOTE**, v. 17, n. 1, p. 306-315.
- SIGAA. **Alunos ativos no curso de BCC da UFPB**. 2019a. Disponível em: <https://sigaa.ufpb.br/sigaa/public/curso/alunos.jsf?lc=pt_BR&id=1626669>. Acessado em: 25 out. 2019.
- SIGAA. **Estrutura curricular do curso de BCC na UFPB**. 2019b. Disponível em: <<https://sigaa.ufpb.br/sigaa/link/public/curso/curriculo/940>>. Acessado em: 9 out. 2019.
- SODS. **Regimento geral da UFPB**. 11 mai 2016. Disponível em: <http://www.ufpb.br/sods/contents/menu/institucional/copy_of_regimentos/regimento-geral>. Acesso em: 6 set. 2019.