

Análise Baseada em Dados de Relações do Desempenho entre Disciplinas: Um Estudo de Caso na UFRN

Andressa Stéfany Silva de Oliveira¹, Rute Souza de Abreu¹, Luiz Affonso Guedes¹

¹Departamento de Engenharia de Computação e Automação
Universidade Federal do Rio Grande do Norte (UFRN) – Natal, RN – Brasil

{astefanysoliveira,rute.s.abreu}@gmail.com, affonso@dca.ufrn.br

Abstract. *In this paper we use of scientific data visualization techniques to analyze relations of performance of students in courses that compose their curricular grade. In this case, the correlation mapping and parallel coordinate graphs techniques were used in the academic data of students of the Bachelor of Science and Technology graduate course at the Federal University of Rio Grande do Norte. The analysis was encouraged because of the significant number of student dropouts in this course and because the database is only available in tables.*

Resumo. *Neste trabalho utilizamos técnicas de visualização de dados científicos para analisar as relações de desempenho dos alunos nos cursos que compõem a grade curricular. Neste caso, as técnicas de mapeamento de correlação e de gráficos de coordenadas paralelas foram utilizadas nos dados acadêmicos dos alunos do curso de bacharelado em Ciência e Tecnologia da Universidade Federal do Rio Grande do Norte. A análise foi incentivada por causa do número significativo de desistentes de alunos neste curso e porque o banco de dados está apenas disponível na forma de tabelas.*

1. Introdução

Com o advento da tecnologia digital, dados têm sido gerados em grande quantidade no contexto escolar. Uma grande fonte de dados são os sistemas de gerenciamento acadêmico, que armazenam informações institucionais produzidas a partir de sistemas utilizados no gerenciamento de cursos, professores, aulas e alunos. Por outro lado, é justamente no contexto de existência de dados em quantidade e qualidade em níveis estatisticamente representativos que a área de ciência de dados está inserida.

A ciência de dados objetiva transformar dados em estado bruto com baixo valor ao negócio em questão em informação útil com alto valor, para isto utiliza diversas técnicas e algoritmos de extração, processamento, exibição e interpretação. Assim, a ciência de dados se caracteriza como área essencialmente interdisciplinar que se baseia em técnicas e teorias vindas de campos básicos da engenharia e ciências básicas [Porto and Zaviani 2015].

Ressaltando-se que atividade de visualização de dados é uma atividade essencial na ciência de dados, pois auxilia, via visualização adequada de resultados, na compreensão e descoberta de padrões [Santos et al. 2015, Nascimento and Ferreira 2011].

No mesmo âmbito da educação, há diversos trabalhos que utilizam técnicas de visualização de dados para análise de resultados e descoberta de padrões. Em

[WEBBER 2008, Costa et al. 2017b] são utilizadas técnicas de visualização de dados para relacionar a influência que as disciplinas pré-requisitos exercem sobre o sucesso ou insucesso de um discente, ao cursar uma disciplina dependente desse pré-requisito. Em [Barros et al. 2017] utiliza-se técnicas estatísticas em conjunto com técnicas de visualização, como a técnica de agrupamento *K-means*, a análise de correspondência, o *boxplot*, o gráfico violino e o mapa percentual, para a potencialização da interpretação de notas de alunos. Com o intuito de auxiliar alunos na escolha das matérias a serem cursadas, a UFCG disponibiliza um programa web que simula a matrícula [Costa et al. 2017a] e exibe em formato visual amigável o risco de reprovação, a quantidade de pessoas que também cursaram as mesmas disciplinas selecionadas e a porcentagem da conclusão do curso a partir do número de disciplinas obrigatórias concluídas pelo aluno. Assim, pode-se observar que utilização de técnicas de visualização de dados agregadas com informação estatísticas (visualização científica de dados) pode ser de grande valor para análise, interpretação e descoberta de conhecimento no âmbito de ciência de dados aplicada no contexto escolar (*learning analytics*).

Então, assumindo-se a premissa da relevância da utilização de técnicas de visualização científica de dados no contexto educacional, o presente artigo apresenta os resultados obtidos com utilização de duas técnicas: mapa da correlação e gráfico de coordenadas paralelas. A primeira apresenta as correlações de pares de variáveis em forma de uma matriz em cores. A segunda descreve de forma visual em série as relações entre diversas variáveis. Ambas as técnicas apresentam de forma visual conceitos estatísticos bem fundamentados.

Como estudo de caso para analisar a pertinência do uso dessas técnicas se utilizou dados de desempenho escolar de alunos do curso de Bacharelado em Ciências e Tecnologia (BCT) da Universidade Federal do Rio Grande do Norte (UFRN) entre 2015 e 2017. Essa escolha de estudo de caso se deveu ao fato do panorama presente no BCT, em que há um número bastante significativo de discentes desistentes, ou seja, que trancaram o curso ou abandonaram, tendo-se uma problemática na qual torna interessante aliar técnicas da Ciência de Dados com os materiais fornecidos pelo Portal de Dados da UFRN para ser feita uma análise da aprendizagem baseada em dados, buscando relações existentes entre as disciplinas através do desempenho dos alunos. Assim, tem-se como objetivo principal da análise tornar possível discussões e conclusões a partir da visualização dos dados em forma de gráficos.

O restante deste artigo está organizado da seguinte forma. A Seção 2 apresenta Fundamentação Teórica, abordados a Ciência de Dados, Análise de Aprendizado, Mapa de Correlação e Gráfico de Coordenadas Paralelas. A Seção 3 especifica o Estudo de Caso realizado com dados acadêmicos de alunos do curso Bacharelado em Ciência e Tecnologia da UFRN. A Seção 4 apresenta os resultados obtidos com respectivas discussões. E, por fim, a Seção 6 apresenta as principais conclusões e indicações de trabalhos futuros.

2. Fundamentação Teórica

2.1. Ciência de Dados

A geração de novos dados é constante e o mundo se encontra em um cenário em que a maioria dos serviços prestados e processos de trabalho são informatizados. Por exemplo, o armazenamento de informações institucionais, a medição de grandezas físicas por meio

de sensores com a internet das coisas, sistemas de informação da área da saúde que geram dados de seus pacientes, e até redes sociais, as quais podem englobar dados pessoais dos usuários, postagens compartilhadas, dentre outros dados. Porém, gerenciar, manipular e analisar toda a massa de dados tem sido um desafio. Para isso, existe a Ciência de Dados, a qual atua realizando análises preditivas dos dados, e a Engenharia de Dados que está ligada à infraestrutura do armazenamento dos dados.

A ciência de dados emergiu das disciplinas de estatística, computação e informática, mas não se restringe apenas a isso, se caracterizando como ciência interdisciplinar, pois é possível envolver também as áreas da gestão, comunicação e sociologia, podendo operar em pesquisas relacionadas com o serviço social, áreas da saúde, esporte, comunicação, entre outras. A partir das informações adquiridas, independentemente da área, essa ciência auxilia na interpretação e extração de informações desses dados brutos, os quais por si só às vezes não são significativos e por isso se faz necessário a análise dos dados [Cao 2017, Porto and Zaviani 2015].

Dessa forma, "a análise de dados corresponde a um conjunto de atividades que devem ser desempenhadas, desde a seleção dos dados até a produção do conhecimento, que é o principal produto da análise"[Porto and Zaviani 2015], onde primeiro ocorre a seleção dos dados, esses dados são posteriormente pré-processados e, então, são aplicados métodos para análise e avaliação. Anteriormente, era utilizado apenas a estatística para a análise de dados, porém, agora também se inclui a Mineração de Dados e a Aprendizagem de Máquina, do inglês, *Data Mining* e *Machine Learning*, respectivamente.

Na busca de destrinchar a complexidade dos dados e obter resultados, são utilizados sistemas de associação, correlação, dependência e causalidade existentes, os quais mostram o grau de associabilidade entre os dados, como por exemplo, no ambiente educacional, como o desempenho de um aluno na disciplina de matemática pode impactar no desempenho da disciplina de física? A resposta para esse questionamento envolve a correlação entre as disciplinas [Cao 2017, Costa et al. 2017b].

2.2. Análise de Aprendizado

Na educação, com o surgimento da interação do aluno com aplicativos que gerenciam cursos e aulas, a quantidade de dados produzida tem crescido bastando, assim, proporcionando uma nova oportunidade para obter *insights* sobre o aprendizado dos alunos, além do surgimento de um aliado para compreensão do processo de aprendizagem [Gašević et al. 2015, Siemens and Baker 2012].

A Análise de Aprendizado, ou em inglês *Learning Analytics*, compreende em medir, coletar, analisar e relatar os dados sobre aprendizes e seus contextos, possuindo o objetivo de compreender e otimizar a aprendizagem e os ambientes em que ela ocorre [Gašević et al. 2015]. Nesse processo da análise de aprendizagem, é utilizado a visualização dos dados, também conhecido como mapeamento visual, a qual é responsável pela representação visual dos dados que inicialmente estão apenas na estrutura de tabelas [Nascimento and Ferreira 2011].

2.3. Mapa de Correlação

Diante do cenário da Ciência de Dados, faz-se necessário extrair conclusões das informações contidas nos dados. Desta forma, a estatística é de extrema importância,

pois serve de auxílio, disponibilizando métodos para organizar e resumir tais informações. Especificamente, a estatística descritiva busca resumir e descrever as características importantes dos dados, alguns dos métodos disponíveis são gráficos, como a construção de histogramas, *boxplots* e gráficos de dispersão, enquanto outros envolvem cálculos, como a média, desvio padrão e coeficientes de correlação [Devore 2011].

Antes de explicar o coeficiente de correlação, será abordada a covariância, pois o primeiro depende do segundo. A definição de covariância é dada pela Equação 1.

$$Cov(X, Y) = \sum_x \sum_y (x - \mu_x)(y - \mu_y)p(x, y) \quad (1)$$

Sendo X e Y as variáveis aleatórias discretas da distribuição conjunta com função massa de probabilidade (fmp) $p(x, y)$, onde essa fmp conjunta descreve a quantidade de massa de probabilidade existente em cada par de valores possíveis de x e y , e μ é a média da população.

A covariância representa a variabilidade presente na distribuição das variáveis aleatórias. Se o valor resultante for positivo, significa que quanto maior o valor de X , maior será o valor de Y e quanto menor o valor de X , menor será o de Y , ou seja, possuem um comportamento semelhante. Sendo o resultado negativo, o comportamento das variáveis se dará de maneira oposta, quanto menor o valor de um, maior será o valor do outro. No caso em que não há padrão no comportamento, o valor da covariância tenderá a zero [Devore 2011]. A correlação é a relação existente entre as variáveis X e Y , sendo a sua definição dada pela Equação 2.

$$Corr(X, Y) = \rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad (2)$$

Como mostrado na Equação 2, o coeficiente de correlação é a covariância das duas variáveis aleatórias sobre o produto do desvio padrão da população das variáveis. A correlação é uma medida de grau da relação linear entre X e Y , se $\rho = 0$ essas variáveis são não-correlacionadas linearmente, pois ρ representa o quão fortemente é a relação entre X e Y . Porém, não se pode afirmar que $\rho = 0$ prova que as variáveis são independentes, pois é possível que exista uma relação não linear entre elas [Devore 2011].

2.4. Gráfico de Coordenadas Paralelas

O gráfico de coordenadas paralelas mapeia dados n -dimensionais em uma estrutura bi-dimensional utilizando n eixos paralelos verticais que são denominados coordenadas, as quais representam as dimensões ou atributos dos dados. Como mostrado na Figura 1(a), o dado é representado pela linha que passa por $p_1, p_2, p_3 \dots p_n$ possuindo n dimensões. Essa técnica foi inicialmente proposta por Inselberg [Inselberg and Dimsdale 1990, Nascimento and Ferreira 2011].

Esse tipo de gráfico permite a observação de padrões pelo fato de cada item de dado estar associado a uma linha gráfica que se conecta aos seus respectivos valores nos eixos, demonstrando assim, o fluxo de mudanças de valores. Outra característica interessante é a seleção de um intervalo de interesse em um eixo. Na Figura 1(b) as linhas destacadas são as que passam pelo intervalo selecionado, possibilitando a observação do comportamento desses dados em específico caracterizando uma filtragem simples dos dados [Nascimento and Ferreira 2011].

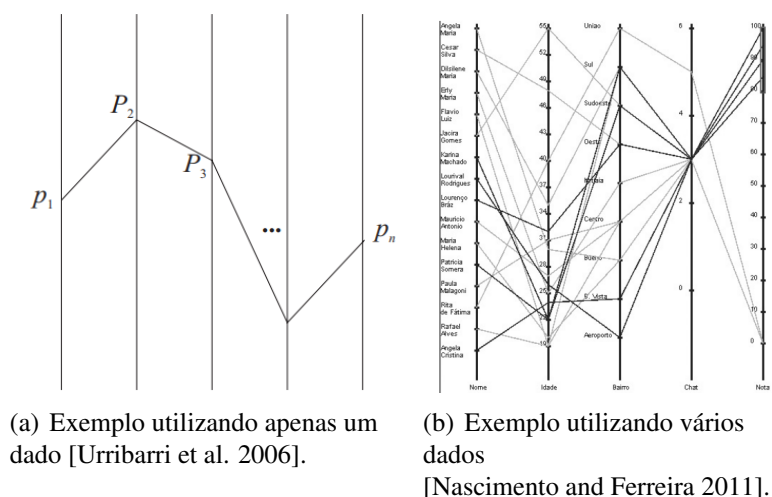


Figura 1. Coordenadas paralelas.

3. Estudo de Caso

O estudo de caso foi realizado com dados dos alunos do curso de Bacharelado em Ciências e Tecnologia (BCT) da Universidade Federal do Rio Grande do Norte (UFRN) no período de 2015 a 2017, totalizando dados de 6946 alunos. As disciplinas consideradas serão as cursadas nos primeiros quatro períodos do curso, obrigatórias na grade curricular.

Com o intuito de obter informações sobre um padrão de comportamento dos alunos, foram utilizados os dados disponíveis no Portal de Dados Abertos da UFRN, o qual disponibiliza dados e informações públicas com o propósito de dar acesso à informação e o acompanhamento do funcionamento da universidade para toda a população. A partir desse portal foram obtidas a relação de alunos do Bacharelado em Ciências e Tecnologia, como também, a relação das matrículas em componentes dos discentes, a relação das turmas, os componentes curriculares oferecidos no curso, entre outros dados.

Para o tratamento e manipulação de tais dados, foi usada a linguagem de programação Python, as bibliotecas Pandas, NumPy, Matplotlib, Plotly para Javascript, e os *frameworks* React e Django. Essas ferramentas auxiliaram tanto no tratamento dos dados, como também para a visualização dos dados.

3.1. Portal de Dados Abertos da UFRN

O Portal de Dados Aberto da UFRN fornece informações a respeito da universidade, por exemplo: comunicados e documentos, contratos e convênios, despesas e orçamento, turmas dos cursos técnicos, turmas da graduação e pós graduação, processos seletivos realizados na UFRN, programas de extensão, relação dos docentes da instituição, empréstimos dos acervos da academia, a avaliação a docência a qual é realizada todo semestre e telefones do estabelecimento de ensino, são alguns exemplos de dados disponíveis no site do Portal no formato *Comma Separated Values* - CSV.

Alguns dados de ensino obtidos do portal foram manipulados para se chegar ao objetivo do presente trabalho. As tabelas utilizadas foram as de matrículas dos discentes nas disciplinas, a relação das turmas semestrais, componentes curriculares da universidade e cursos de graduação oferecidos.

3.2. Tecnologias Utilizadas

Para a realização deste trabalho, foi utilizada linguagem de programação Python 3.6 juntamente com algumas de suas bibliotecas científicas para a ciência de dados, sendo elas: NumPy, Pandas, Matplotlib e Seaborn. O NumPy possui suporte para arrays multidimensionais com operações básicas e funções de álgebra linear, enquanto que o Matplotlib e o Seaborn são bibliotecas para a visualização dos dados, ou seja, usadas na plotagem de gráficos [Iguai and Seguí 2017].

A interface do *software* foi feita com o a biblioteca do JavaScript denominada React, ela é baseada em componentes que devem ser códigos pequenos e isolados, a ideia é a construção de interfaces de usuário complexas a partir da junção de tais componentes. Juntamente com o React, se faz o uso do Plotly, uma biblioteca de gráficos que possui um componente específico para se trabalhar com o React, além desse modo, essa biblioteca pode ser usada no JavaScript, Python e R.

Os dados utilizados para exibir os gráficos são consumidos de um *web service* desenvolvido com o *framework* Django. A interface do usuário faz requisições ao *web service* tendo como estrutura a URL: o endereço que irá receber a requisição, o método: comando GET usado para trazer informações, e o corpo: que informa dados necessários para requisição, por exemplo, se é necessário saber as disciplinas do primeiro período, é preciso que seja informado no corpo o período e como resultado dessa requisição serão retornadas as disciplinas.

3.3. Tratamento dos Dados

O *web service* foi criado para que os dados fossem tratados antes de serem usados na interface. A interface possui páginas que exibem gráficos, e antes dos dados dos gráficos chegarem na aplicação, as tabelas adquiridas no Portal de Dados são modificadas. As subseções a seguir descrevem os serviços disponíveis no *web service*.

3.3.1. Disciplinas

O serviço de disciplinas recebe o número referente ao período e retorna dados de todas as disciplinas do respectivo período, sendo eles o código da disciplina, nome e o pré-requisito. Para se chegar nesse retorno, foi utilizada a tabela da relação de componentes curriculares oferecidos pela UFRN. Inicialmente, foi filtrada pela unidade responsável que possuíam a *string* "ESCOLA DE CIÊNCIA E TECNOLOGIA", e filtrado o ano do programa da disciplina, utilizando-se apenas os anos igual ou maiores que 2014. Esse número foi escolhido pelo fato de querermos apenas disciplinas da grade vigente do curso, e as disciplinas do ano de programa de 2014 tiveram suas primeiras turmas em 2015, tendo sido isso diagnosticado através da própria base de dados do Portal. Em seguida, foram selecionadas apenas as colunas referente ao código, nome e pré-requisitos da disciplinas desejadas, a tabela foi serializada e, por conseguinte, o serviço disponibiliza a resposta através de uma requisição GET.

3.3.2. Correlação das disciplinas

O cálculo do serviço de correlação recebe a lista de disciplinas, tendo como retorno uma matriz de correlação. Por exemplo, pede-se a correlação entre cálculo I, II e III, será calculado a correlação de cálculo I e II, cálculo I e III, e cálculo II e III; são consideradas as notas que o discente obteve quando foi aprovado na primeira matéria e se faz a correlação com a primeira nota que o respectivo discente atingiu na segunda matéria, considerando apenas alunos que possuem notas nas duas matérias.

Para isso, são criadas duas tabelas, cada uma relacionada a uma das disciplinas, como por exemplo, uma será de cálculo I e a outra de cálculo II. Em seguida, são filtrados apenas os discentes aprovados e sua respectivas notas de aprovação da primeira tabela, ou seja, as notas dos aprovados em cálculo I. Há uma filtragem no segundo conjunto de dados selecionando apenas os alunos pré-estabelecidos no primeiro conjunto de dados, porém, para os casos em que há repetência das notas, é escolhido apenas a primeira nota que o aluno obteve na disciplina. Após isso, é criada uma tabela com as colunas cálculo I e cálculo II, onde as linhas representam a média final de um discente. Com essa última tabela, é calculado o coeficiente de correlação linear com a função *corr()*, já existente no Pandas. Nela pode ser escolhido um de três métodos para esse cálculo: Pearson, Kendall Tau e Spearman, o utilizado neste trabalho foi o padrão, o de Pearson. Da mesma forma, esse cálculo é feito para Cálculo I e Cálculo III; e Cálculo II e Cálculo III, obtendo assim uma matriz resultante.

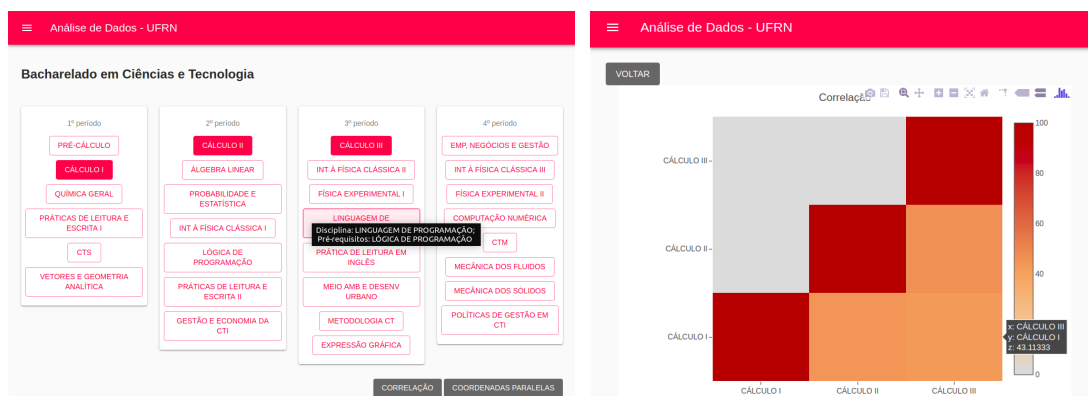
4. Resultados e Discussões

Para o auxílio da análise de dados, os dados coletados foram plotados em gráficos dinâmicos, ou seja, o usuário pode interagir com o gráfico selecionando apenas a área que deseja analisar, por exemplo.

Na aplicação *web*, são exibidas matérias do primeiro ao quarto período, as quais são matérias obrigatórias à todos os alunos do curso. Os pré-requisitos também podem ser visualizados, basta posicionar o cursor em cima do botão referente à matéria que se deseja saber o pré-requisito, observe a Figura 2(a). Além de mostrar as disciplinas, é possível selecioná-las e posteriormente calcular a matriz de correlação ou mostrar o gráfico de coordenadas paralelas das disciplinas escolhidas.

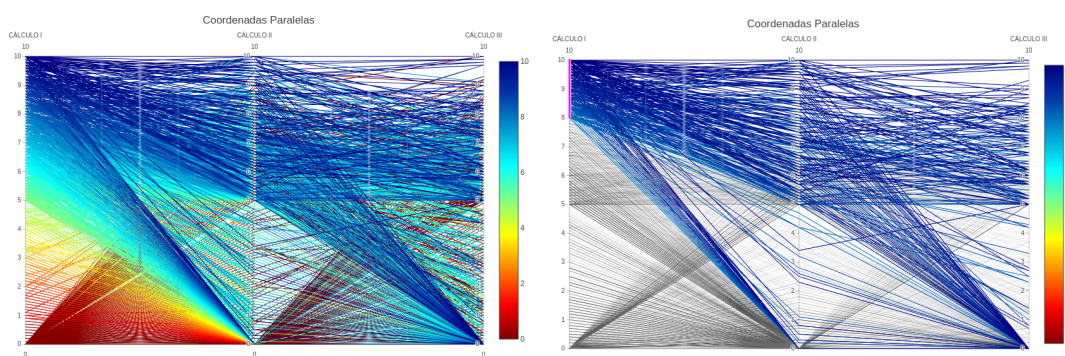
Na Figura 2(a), mostra-e a seleção das matérias de Cálculo I, II e III para o cálculo da matriz de correlação mostrada na Figura 2(b). É possível constatar que a correlação entre essas três disciplinas é em torno de 45%, dessa forma, por ser uma correlação positiva, conclui-se que um aluno que obtém nota boa em cálculo I, possui grandes chances de também conseguir nota boa em cálculo II, logo, é interessante que cálculo I seja pré-requisito de cálculo II devido a essa relação de influência, da mesma forma é a relação entre cálculo II e III.

Outra forma de ver essa relação entre disciplinas e verificar se o resultado da correlação faz sentido, é através do gráfico de coordenadas paralelas, as Figuras 2(c) e 2(d) mostram o gráfico de coordenadas paralelas para as disciplinas de cálculo já mencionadas, onde as linhas correspondem aos alunos, cada eixo vertical representa uma matéria e seu valor corresponde ao da média final de cada aluno variando de 0 a 10. Observe que os alunos que obtiveram média final entre 10 e 8 tenderam a continuar com notas acima de



(a) Tela da grade curricular do curso BCT.

(b) Matriz de correlação resultante das matérias selecionadas.



(c) Coordenadas paralelas com todos os alunos selecionados.

(d) Coordenadas paralelas dos alunos que tiveram média final entre 10 e 8 em cálculo I.

Figura 2. Telas da aplicação web.

5, logo, a correlação se mostra de acordo com os dados.

Outro exemplo de pré-requisitos, são as matérias, Vetores e Geometria Analítica (VGA), Álgebra Linear e Computação numérica, onde a primeira é pré-requisito da segunda e a segunda é pré-requisito da terceira. Veja na Figura 3(a) que as correlações entre si são em torno de 40%, havendo uma maior correlação em VGA e Álgebra e Linear, e entre Álgebra Linear e Computação numérica. Na Figura 3(b) é possível ver o fluxo das notas dos alunos, apesar da ocorrência de alunos que tiraram de 10 a 8 em VGA tirem nota baixa em Álgebra, a maioria continua no nível de notas de 5 pra cima.

Além disso, a aplicação *web* dispõe de uma tela que exibe os coeficientes de correlação entre todas as disciplinas do primeiro ao quarto período, como mostrado na Figura 4, quanto mais escuro a cor do quadrado no mapa de calor, maior é o índice de correlação, podendo assim, se ter uma visão geral de todas as correlações.

5. Conclusão

A partir dos gráficos apresentados, pode-se inferir que os cálculos da correlação mostram mais coeficientes positivos do que negativos, informando que de fato a disciplina pré-requisito influencia no desenvolvimento do aluno na disciplina posterior. Além disso, com o cálculo dos coeficientes de todas as matérias, vemos que existe correlação entre disciplinas que não possuem relação de pré-requisito entre si, como por exemplo, as dis-

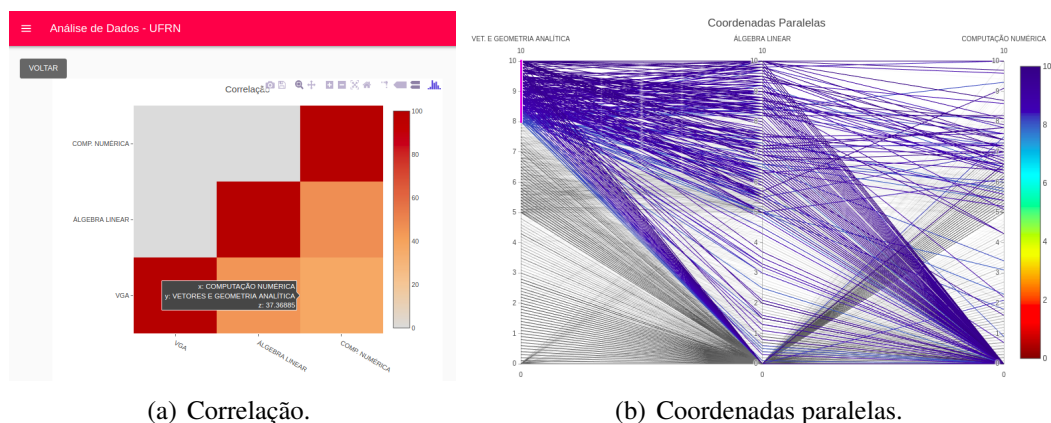


Figura 3. Gráficos envolvendo as disciplinas: Vetores e Geometria Analítica, Álgebra Linear e Computação Numérica.

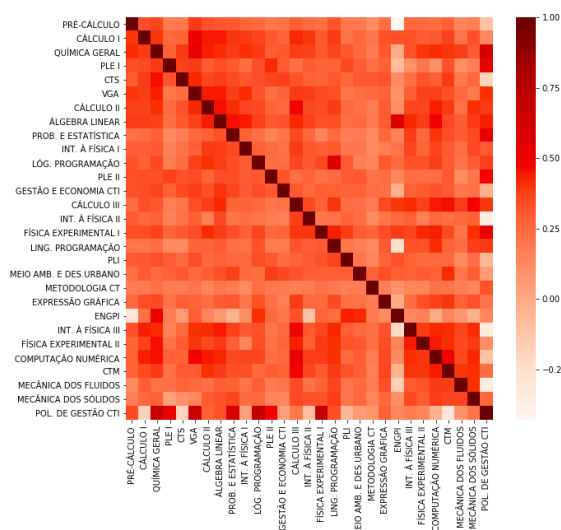


Figura 4. Matriz de correlação entre todas as disciplinas.

ciplinas de Gestão e Economia da Ciência, Tecnologia e Inovação e Cálculo III as quais possuem uma correlação na faixa de 20%.

É notório que essa aplicação *web* contribuiu com a facilitação da visualização dos dados sobre alunos do Bacharelado de Ciências e Tecnologia, ressaltando que essas informações estão presentes em tabelas gigantescas no Portal de Dados da UFRN, as quais em sua forma bruta dificultam e, por vezes tornam impossível, a interpretação e análise dos dados.

A apresentação das médias, comparação entre as notas das disciplinas e a correlação foram úteis, mas ainda há assuntos a serem explorados, como por exemplo, mostrar estatísticas com relação apenas às turmas individualmente, o que seria um ferramenta útil para o professor, uma vez que ele iria acompanhar o desempenho da turma. Uma outra abordagem seria procurar relações entre os dados socio-econômicos e o desempenho acadêmico dos alunos, como também, validar o resultado obtido nas correlações calculadas entre as disciplinas.

Referências

- Barros, T., Silva, I., and Guedes, L. (2017). Modelagem e visualização científica de dados educacionais: Estudo de caso sobre o desempenho em componentes curriculares. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 6, page 654.
- Cao, L. (2017). Data science: challenges and directions. *Communications of the ACM*, 60(8):59–68.
- Costa, A., Nascimento, M., and Santos, W. (2017a). O Uso do Cursos UFCG.
- Costa, E. J., Rodrigues, E. C., Silva, M. V. D., Gomes, R. C. S., and Assis, C. C. B. (2017b). Um estudo estatístico sobre o aproveitamento em Física de alunos de ensino médio e seus desempenhos em outras disciplinas. *Revista Brasileira de Ensino de Física*, 39(1):1–6.
- Devore, J. L. (2011). *Probability and Statistics for Engineering and the Sciences*. Cengage learning.
- Gašević, D., Dawson, S., and Siemens, G. (2015). Let’s not forget: Learning analytics are about learning. *TechTrends*, 59(1):64–71.
- Igual, L. and Seguí, S. (2017). A python approach to concepts, techniques and applications. In *Introduction to Data Science*, pages 1–4. Springer.
- Inselberg, A. and Dimsdale, B. (1990). Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the 1st conference on Visualization’90*, pages 361–378. IEEE Computer Society Press.
- Nascimento, H. A. and Ferreira, C. B. (2011). Uma introdução à visualização de informações. *Visualidades*, 9(2).
- Porto, F. and Zaviani, A. (2015). Ciência de Dados. *Lncc.Br*, pages 1–7.
- Santos, M. C. Q., Moraes, D. S. M. W., Silveira, D. M. S., Träsel, M., Tietzmann, D. R., and Manssour, D. I. H. (2015). A aplicação de visualizações gráficas para a análise jornalística de dados do Twitter.
- Siemens, G. and Baker, R. S. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 252–254. ACM.
- Urribarri, D. K., Castro, S. M., and Martig, S. R. (2006). Escalabilidad visual en coordenadas paralelas. In *VIII Workshop de Investigadores en Ciencias de la Computación*.
- WEBBER, E. G. (2008). *Análise da Relação entre as Notas das Disciplinas de Matemática e Física de Estudantes do Ensino Médio*. Trabalho de conclusão de curso, Universidade Regional Integrada do Alto Uruguais e das Missões.