

CONSTRUÇÃO DE UM SISTEMA ESTATÍSTICO PARA AUXILIAR O SETOR
PEDAGÓGICO DO IFCE – CAMPUS CRATO NO ACOMPANHAMENTO DOS
DISCENTES

Francisco Wilcley Lacerda de Lima¹, Rennan Rodrigues Isidio Teles¹, Francisca Alves de Souza¹, Cícero Carlos Félix de Oliveira¹, Guilherme Esmeraldo¹

¹ Instituto Federal de Educação, Ciência e Tecnologia do Ceará – *campus* Crato

RESUMO: Uma das atribuições do setor pedagógico no ensino médio é o acompanhamento do desempenho acadêmico dos discentes a cada bimestre. Para tanto, se faz necessário agilidade e clareza na análise dos resultados emitidos pelo corpo docente em cada disciplina, para que seja possível, quando necessário, intervir com instrumentos eficazes de recuperação da aprendizagem dos alunos na disciplina que não houve o êxito desejado. Portanto, esse artigo descreve os elementos utilizados na construção de um sistema computacional, sem custo financeiro para a instituição e de fácil compreensão para o usuário, que utiliza os conhecimentos estatísticos para realizar a descrição, a apresentação e análise dos dados coletados.

Palavras-chave: Acompanhamento Pedagógico. Ferramenta computacional. Sistema Web. Análise Estatística.

ABSTRACT: One of the attributions of the pedagogical sector in high school is the monitoring of the academic performance of students every two months. Therefore, it is necessary to have agility and clarity in the analysis of the results emitted by the faculty in each discipline, so that it is possible, when necessary, to intervene with effective instruments to recover students' learning in the discipline that did not have the desired success. Therefore, this paper describes the elements used to build a computer system, with no financial cost for the institution and easy to understand for the users, who uses the statistical knowledge to perform the description, presentation and analysis of the collected data.

Keywords: Pedagogical Accompaniment. Computational tool. Web System. Statistical Analysis.

1. INTRODUÇÃO

O acompanhamento do desempenho acadêmico, na grande maioria das instituições de ensino brasileiras, vem sendo realizado através de avaliações quantitativas por disciplina, e seus resultados têm sido utilizados como parâmetro base para o cálculo do rendimento escolar. No ensino médio, por exemplo, essa avaliação ocorre a cada bimestre, de forma que pode-se obter parâmetros periódicos e parciais de rendimento, os quais podem ser utilizados como referencial para que o estudante possa, em tempo, estabelecer ações para recuperar, manter ou aumentar seu desempenho no decorrer de um ano.

As instituições de ensino frequentemente têm buscado otimizar os processos de ensino-aprendizagem visando aumentar os desempenhos de seus estudantes, contudo muitos são os fatores que colaboram para a retenção escolar, como a Hiperatividade, Transtorno do Deficit de Atenção, Transtornos do Espectro Autista, Dislexia, *Bullying*, problemas domésticos, renda familiar, tempo de estudo, formação dos pais e índice de violência (FEITOSA; ESMERALDO; DE OLIVEIRA, 2013) (AMÂNCIO-VIEIRA et al., 2015). Avaliar cada um desses aspectos torna-se uma tarefa que pode demandar

tempo, recursos humanos e financeiros excessivos, muitas vezes indisponíveis para os setores pedagógicos das instituições de ensino, e o resultado pode não ser suficientemente eficiente ao ponto de se evitar uma retenção (FEITOSA; ESMERALDO; DE OLIVEIRA, 2013). Nesse sentido, a análise quantitativa torna-se um instrumento essencial e factível de resultados no acompanhamento de desempenho acadêmico.

A análise estatística é uma importante ferramenta para a pesquisa em diversos campos do saber, como na economia, engenharias, fisiologia, etc. (CASTRUCCI, 2005). A estatística consiste de um conjunto de métodos quantitativos que tratam de elementos de pesquisa e inclui uma série de etapas de análise, como coleta, exploração, descrição e interpretação de dados numéricos, para, ao final, suportar a tomada de decisões (SILVESTRE, 2007) (CARVALHO e CAMPOS, 2008). Dependendo das técnicas estatísticas a serem aplicadas e dos tamanhos das bases de dados a serem analisadas, serão necessários cálculos cada vez mais complexos para se completar cada uma das fases do método estatístico.

Na literatura científica, há diversos tipos de ferramentas estatísticas, onde sua aplicabilidade e opção são condicionados a uma série de fatores, como: custo,

procedimento de instalação, necessidade de treinamento, pacotes complementares, modelo de interação, suporte ao usuário, geração de gráficos atraentes, entre outros (OZGUR; KLECKNER; LI, 2015). No entanto, mesmo com as facilidades tecnológicas que elas oferecem, deve-se considerar que, para sua utilização efetiva, o usuário necessita ter conhecimentos de estatística e, em muitos casos, de lógica de programação. Esses fatores podem tornar impraticável o uso dessas ferramentas por usuários leigos (PRADHANANGA, Y.; KARANDE, S.; KARANDE, 2016).

Diante desse contexto, o presente trabalho visa contribuir no auxílio à análise do desempenho acadêmico dos estudantes do Instituto Federal do Ceará (IFCE) *campus* Crato, através de uma nova solução de análises estatísticas. Objetiva-se que a ferramenta proposta possa ser utilizada pelo setor pedagógico para descrição, apresentação e análise de dados quantitativos escolares, como, por exemplo, notas, número de faltas e número de estudantes evadidos, de forma a estabelecer ações preventivas de evasão e/ou retenção.

2. REVISÃO TEÓRICA

Na literatura científica, é possível encontrar variados trabalhos que utilizam métodos estatísticos para análise e diagnóstico de diferentes indicadores educacionais.

Um dos exemplos é o trabalho em (FEITOSA; ESMERALDO; DE OLIVEIRA, 2013), o qual apresenta uma abordagem de uso de modelos de regressão múltipla, que considera as notas obtidas nas avaliações bimestrais, para estimar o desempenho acadêmico de estudantes do ensino médio. A abordagem proposta objetiva oferecer um instrumento para apoio à tomada de decisões para a redução da retenção escolar.

Assim como no trabalho anterior, a pesquisa em (AMÂNCIO-VIEIRA et al., 2015) utiliza modelos de regressão múltipla para caracterizar o rendimento escolar. No entanto, utilizou-se uma gama maior de variáveis candidatas ao modelo, distribuídas entre três categorias. São elas: 1) variáveis técnicas, que envolveram, por exemplo, número de alunos, número de professores graduados e pós-graduados e relação aluno-professor; 2) variáveis de custo pedagógico, sociais e administrativos; e 3) variáveis de desempenho acadêmico, como nota no Ideb e Brasil/Saeb. Os resultados comprovaram que a atuação docente e os fatores orçamentários possuem grande impacto nos resultados acadêmicos.

Já o trabalho em (DA SILVA, 2013) realiza modelagem estatística com o objetivo de identificar as principais variáveis que aumentam a correlação com a evasão em comparação com a média dos alunos do ensino superior da instituição. O estudo mostrou que a instituição analisada pode utilizar os resultados obtidos para tomar medidas administrativas para sanar dificuldades

pedagógicas/financeiras e com isso reduzir os índices de evasão.

Por fim, o estudo em (CAPOVILLA; DIAS, 2007) buscou identificar o desenvolvimento das estratégias de leitura logográfica, alfabética e ortográfica em crianças da 1a. a 4a. Séries do ensino fundamental. A abordagem buscou correlacionar, através de cálculos do coeficiente de Pearson, o desempenho do uso dessas técnicas de leitura às notas escolares obtidas pelos estudantes, a fim de delinear medidas de intervenção para tratamento ou prevenção de dificuldades de leitura.

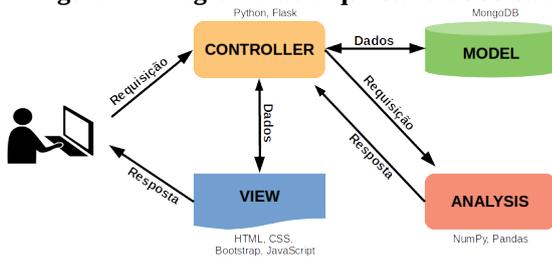
Nesse cerne, percebe-se que a estatística é um instrumento muito importante para compreensão dos diferentes contextos no mundo acadêmico e que pode ser utilizada para estabelecer indicadores para otimizar os resultados e aumentar a qualidade dos processos educacionais. O trabalho proposto neste artigo consiste de uma nova ferramenta que busca simplificar a realização de análises estatísticas, potencializando sua utilização na caracterização de indicadores para apoio pedagógico.

3. MATERIAIS E MÉTODOS

Na metodologia adotada para o desenvolvimento da solução aqui proposta, inicialmente, realizou-se um levantamento bibliográfico sobre as tecnologias de desenvolvimento de sistemas Web, seguido por discussões em equipe para elencar os requisitos iniciais de interface, arquitetura de software e principais funções.

Em seguida, após a codificação das funções em módulos e da interface do sistema proposto, realizou-se novo estudo visando aumentar o desempenho da camada de análises estatísticas e de armazenamento de dados. Do estudo, foram adicionadas novas bibliotecas de software para processamento de alto desempenho, além da substituição do sistema de gerenciamento de bancos de dados por outro com abordagem NoSQL, a qual tem se mostrado mais eficiente em relação à relacional, quando se trata de processamento paralelo distribuído (MONIRUZZAMAN; HOSSAIN, 2013) e de bancos de dados maiores (ALMEIDA, BERNARDINO; FURTADO, 2015). Ressalta-se que, nesta etapa, realizou-se diversos experimentos, que incluíram o uso de bases de dados com diferentes tamanhos e diferentes cálculos estatísticos. Além disso, a interface do sistema proposto sofreu diversos refinamentos para suportar os princípios de usabilidade, objetividade, amigabilidade, acessibilidade e responsividade (CYBIS; BETIOL; FAUST, 2015). A arquitetura de software resultante do sistema proposto pode ser vista na Figura 1.

Figura 1 - Diagrama de arquitetura de software.



Fonte: Elaborado pelo autor (2018).

A arquitetura de software proposta, ilustrada na Figura 1, é baseada no padrão MVC (*Model View Controller*) (GAMMA et al., 2015), e está dividida em quatro camadas: 1) **Model**: envolve o gerenciamento das bases de dados que serão analisadas e seu armazenamento em banco de dados; 2) **View**: disponibiliza uma interface para apresentação e interação com o sistema; 3) **Controller**: inclui a lógica de negócio do sistema e permite a comunicação entre as demais camadas; e 4) **Analysis**: inclui as rotinas para análises de alto desempenho e geração de gráficos estatísticos.

Para a codificação da arquitetura proposta, em cada camada utilizou-se as seguintes tecnologias: *Model*) sistema de gerenciamento de bancos de dados orientado a documentos MongoDB; *View*) HTML5, CSS, JavaScript e o framework Bootstrap; *Controller*) linguagem de programação Python e framework Web Flask; e *Analysis*) Pandas e Numpy.

As subseções a seguir trazem descrições de cada uma dessas tecnologias, de acordo com a respectiva camada.

3.1. Model

O MongoDB é um banco de dados escalável e flexível com abordagem NoSQL (Not Only SQL). A ferramenta armazena dados em documentos semelhantes a JSON, o que significa que os campos podem variar de documento para documento e a estrutura de dados pode ser alterada ao longo do tempo. O modelo de documento é mapeado para os objetos no código, facilitando a manipulação dos dados. Consultas *ad hoc*, indexação e agregação em tempo real fornecem formas poderosas de acessar e analisar seus dados. É um banco de dados distribuído em seu núcleo, de modo que a alta disponibilidade, o dimensionamento horizontal e a distribuição geográfica são integrados e fáceis de usar. Além do mais, é gratuito e de código aberto, publicado sob a licença GNU Affero General Public License.

3.2. View

HTML5 é a versão mais recente do HTML. Ela introduz tags simplificadas, nova semântica e elementos de mídia, e se baseia em um conjunto de bibliotecas JavaScript que permite aplicativos web. A nova versão introduz elementos que adicionam nova semântica a suas páginas, trazendo ao desenvolvedor mais opções para a criação da estrutura da página Web.

CSS, que é uma abreviação para o termo em inglês *Cascading Style Sheet*, ou traduzido para o português como “folhas de estilo em cascata”, consiste em “(..) um mecanismo simples para adicionar estilos (por exemplo: fontes, cores, espaçamentos) aos documentos web.” (SILVA, 2012. p.24).

JavaScript é a linguagem de programação dinâmica da Web. A ampla maioria dos sites modernos utiliza JavaScript e todos os navegadores modernos, presentes em computadores de mesa, consoles de jogos, tablets e smartphones, incluem interpretadores JavaScript, tornando-a a linguagem de programação mais onipresente da história. JavaScript faz parte da tríade de tecnologias que todos os desenvolvedores Web devem conhecer: HTML, para especificar o conteúdo de páginas Web; CSS para especificar a apresentação dessas páginas e JavaScript, para especificar o comportamento delas.

Bootstrap é um kit de ferramentas de código aberto para desenvolvimento com HTML, CSS e JS. Através dele é possível criar, de maneira rápida, interfaces responsivas, amigáveis, acessíveis e objetivas. O framework utiliza um sistema de “grids responsivos” que permitem que página se adapte ao tamanho da resolução do equipamento utilizado no acesso. O Bootstrap está licenciado sob a licença MIT.

3.3. Controller

Python é uma linguagem de altíssimo nível orientada a objetos, de tipagem dinâmica e forte, interpretada e interativa. O Python tem uma sintaxe clara e concisa que favorece a legibilidade do código-fonte, tornando a linguagem mais produtiva. Mantida pela *Python Software Foundation* (PSF), é um software de código aberto (com licença compatível com a *General Public License – GPL*, porém menos restritiva, permitindo que o Python seja inclusive incorporado em produtos proprietários). (BORGES, 2014. p.13)

Flask é um micro-framework (*framework* minimalista) desenvolvido em Python e baseado nas tecnologias:

1. **Werkzeug**: consiste de uma biblioteca para desenvolvimento de aplicações WSGI (*Web Server Gateway Interface*), ou, em outras palavras, inclui uma especificação universal de como deve ser a interface (ou modelo de interação) entre uma aplicação desenvolvida em Python e um servidor Web. Para tanto, Werkzeug

define um padrão para interceptar requisições Web, e lidar com as respectivas respostas, além de suportar o controle de *cache*, *cookies*, roteamento de *urls* e também conta com uma poderosa ferramenta de depuração de código.

2. Jinja2: É um *template engine* escrito em Python, onde, através dele, é possível gerar dinamicamente (renderizar) páginas Web estáticas a partir de marcações em um template.

Flask é mantido por uma comunidade bastante ativa, seu código é baseado em diferentes padrões de projeto de software, possuindo alta qualidade e legibilidade, e permite que desenvolvedores tenham liberdade em estruturar, de formas diferentes, novos aplicativos em desenvolvimento. Flask inclui ainda os *Blueprints* – mecanismos criados para centralizar o registro de novas operações provindas de uma extensão –, os quais tornam possível o reaproveitamento simplificado de extensões predefinidas para a adição de novos recursos ao projeto.

3.4. Analysis

Pandas é uma biblioteca de código aberto e fornece estruturas de dados e ferramentas de análise de dados de alto desempenho para a linguagem de programação Python. O Pandas é um projeto promovido pela NumFOCUS (*Numerical Foundation for Open Code and Useable Science*), que é uma fundação sem fins lucrativos que busca promover diversas tecnologias de desenvolvimento de software para aplicações científicas.

NumPy é a biblioteca fundamental para computação científica com a linguagem de programação Python. Ela inclui uma estrutura de dados de array N-dimensional, a qual permite representar dados em memória em baixo nível e, com isso, manipulá-los de forma mais eficiente, ferramentas para integração de código com C/C++ e Fortran, funções sofisticadas para análise de dados, como, por exemplo, para álgebra linear e transformada de Fourier. Além de seu uso científico, o NumPy também pode ser usado como um contêiner multidimensional eficiente de dados genéricos. Desta forma, tipos de dados arbitrários podem ser facilmente convertidos, permitindo que o NumPy integre-se de forma simplificada e eficiente a uma ampla variedade de bancos de dados.

4. RESULTADOS E DISCUSSÃO

O sistema Web proposto atualmente inclui os seguintes módulos funcionais:

1. **Login/Cadastro:** é o subsistema responsável por gerenciar os usuários do sistema. Ele permite a criação e edição das informações do usuário, autenticação, autorização bem como acesso aos recursos do sistema proposto;

2. **Dashboard:** consiste na área de trabalho do usuário. Este subsistema é o ponto central que orienta às atividades de cadastro de novas bases de dados, realização de análises estatísticas e visualização de resultados, através da geração de relatórios;
3. **Análises Estatísticas:** módulo de suporte à edição das bases de dados cadastradas e à realização de análises estatísticas sobre elas. Neste módulo, pode-se ainda criar perfis de análises, onde o usuário do sistema pode selecionar determinados cálculos estatísticos, que poderão ser executados conjuntamente em instantes predeterminados;
4. **Relatórios:** lista os relatórios gerados a partir de análises estatísticas previamente realizadas, permitindo visualizá-los e/ou exportá-los a qualquer momento.

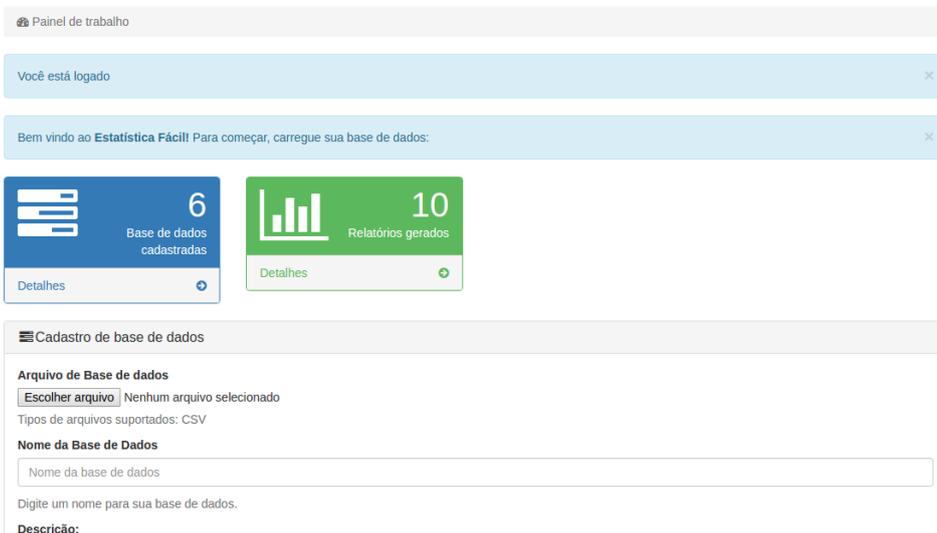
A Figura 2 mostra o *dashboard* de um determinado usuário, onde é possível visualizar um total de 6 bases de dados cadastradas e a disponibilidade de relatórios de 10 análises previamente realizadas.

Para realização de uma análise estatística, a partir do *dashboard*, são necessários apenas três passos básicos:

1. **Cadastro de uma base de dados:** esta etapa consiste em selecionar um arquivo em formato CSV, o qual inclui a base dados a ser analisada, definir um nome e uma descrição para a base e, por fim, realizar o upload dos dados para o sistema Web proposto. A Figura 2 mostra os controles para cadastro de uma nova base de dados;
2. **Análise de dados:** após o cadastro, a base de dados fica disponível para manipulação e/ou análise estatística. A análise consiste em selecionar os tipos de cálculos e gráficos estatísticos que serão gerados a partir da base dados. Na Figura 3, pode-se observar parte da base de dados intitulada “notas de todos os alunos”, a qual inclui os registros de notas dos 4 bimestres letivos de vários alunos. Ainda na figura, verifica-se que foram selecionados os cálculos de média, mediana, moda, somatório, variância, desvio padrão, desvio absoluto, amplitude e a geração de histogramas.
3. **Geração de Relatório:** por fim, após uma análise de dados, o relatório torna-se disponível no *dashboard* e pode ser visualizado sempre que desejado. Na Figura 4, pode-se observar o relatório da análise estatística configurada na Figura 3. Nesse relatório, é possível observar os resultados dos cálculos solicitados para cada bimestre, bem como a presença dos gráficos histograma.

Figura 2. Área de trabalho de um determinado usuário.

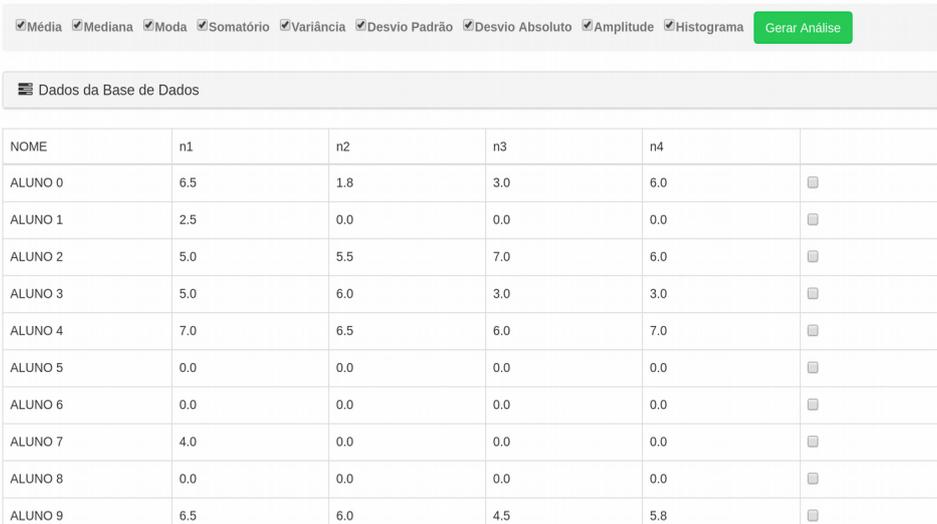
Painel principal Resumo das bases de dados cadastradas



Fonte: Elaborado pelo autor (2018).

Figura 3. Apresentação de uma base de dados cadastrada no sistema Web proposto.

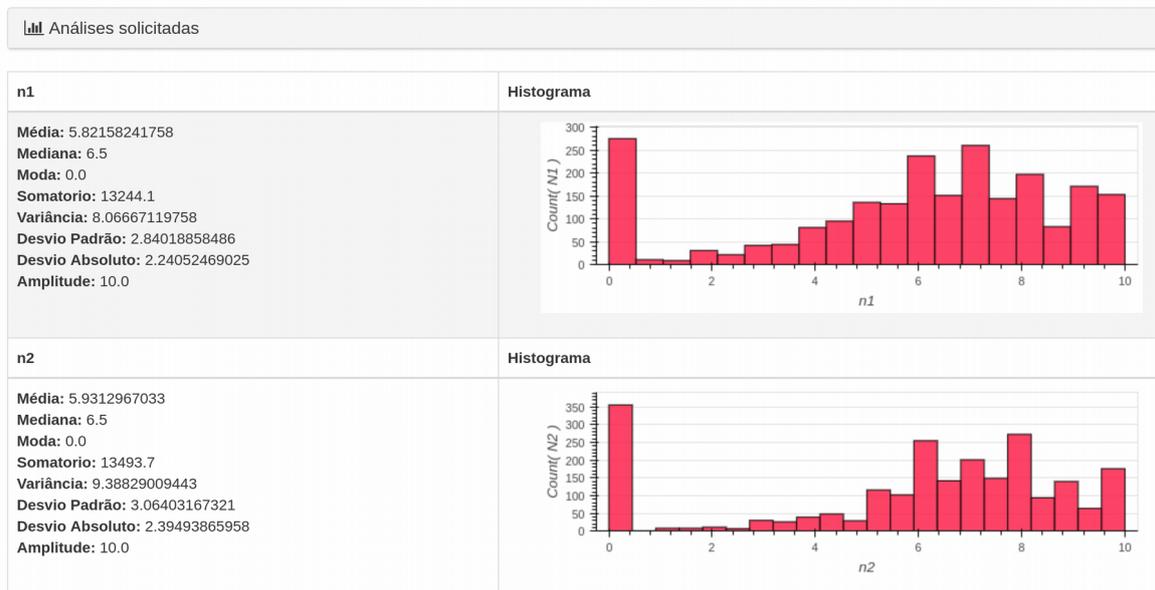
notas de todos os alunos Selecione a análise estatística desejada:



Fonte: Elaborado pelo autor (2018).

Figura 4. Exemplo de relatório de análise estatística.

Resultado de análise: notas de todos os alunos



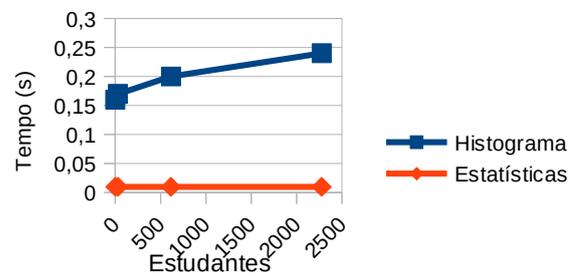
Fonte: Elaborado pelo autor (2018).

Para avaliar o sistema Web proposto, elaborou-se um estudo de caso, onde analisou-se as notas da disciplina de Português dos estudantes do ensino médio do IFCE campus Crato, nos 3 primeiros bimestres letivos, com o objetivo de oferecer suporte ao Setor Pedagógico, no acompanhamento do desempenho individual, por turmas e por nível de escolaridade, e com isso, quando necessário, buscar mecanismos para reduzir os índices de retenção.

No estudo, calculou-se as estatísticas descritivas média, desvio padrão, nota máxima, nota mínima, mediana, 25° e 75° percentis, bem como plotou-se o histograma das notas em quatro cenários: 1 estudante, 31 estudantes de uma turma do 1° ano, 615 estudantes de todas as turmas do 1° ano e 2275 estudantes de todas as turmas dos 1°, 2° e 3° anos.

A Figura 5 apresenta um gráfico de *Speedup* com os resultados de desempenho para o cálculo das estatísticas descritivas e geração de histogramas para cada um desses cenários. As curvas no gráfico na Figura 5 mostram que, tanto para o cálculo das estatísticas e quanto para geração dos histogramas, nos 4 cenários, os tempos ficaram abaixo de 0,25 segundos. Na figura, percebe-se ainda que os comportamentos das curvas tendem a ser lineares, mostrando assim um grande potencial de escalabilidade do tamanho das bases de dados que podem ser manipuladas pelo sistema de Web proposto.

Figura 5. Gráficos de *Speedup* para cálculos e para geração de gráficos de estatísticos.



Fonte: Elaborado pelo autor (2018).

5. CONCLUSÕES

Esse trabalho apresentou uma proposta de sistema para apoio às análises estatísticas na Web. Em seu projeto, buscou-se aplicar um conjunto de técnicas de projetos de interfaces visando suportar usabilidade, objetividade, amigabilidade, acessibilidade e responsividade, bem como de computação de alto desempenho para permitir análises estatísticas mais eficientes.

Para avaliá-lo, utilizou-se um estudo de caso de análise de notas de alunos do ensino médio do IFCE campus Crato e os resultados mostraram que o sistema proposto dispõe de uma interface simplificada para suportar as tarefas de cadastro de bases de dados, análise

estatística e geração de relatórios, bem como apresentou alto desempenho e escalabilidade.

A infraestrutura do sistema de análises proposto atualmente está em fase de testes. Em paralelo, o novo ciclo de desenvolvimento inclui adicionar o suporte de análises estatísticas e geração de gráficos mais complexos. Além disso, será necessária a inclusão de um módulo de agendamento de tarefas ao sistema proposto, uma vez que objetiva-se compartilhar seus serviços entre grupos e núcleos de pesquisa da região do Cariri.

Como trabalhos futuros, cogita-se a inclusão de Big Data e Inteligência Artificial, visando a recomendação de tipos de análises estatísticas de acordo com diferentes perfis (estrutura da base de dados cadastrada, tipo de usuário, linha de pesquisa cadastrada, histórico de uso do sistema, entre outros.). Outros esforços concentram-se em otimizações de desempenho, como a distribuição de cálculos estatísticos entre diferentes nós computacionais e gerenciamento mais eficiente de consumo de memória.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, R.; BERNARDINO, J.; FURTADO, P. Testing SQL and NoSQL approaches for big data warehouse systems. **International Journal of Business Process Integration and Management**, 7(4), 2015. pp.322-334.

AMÂNCIO-VIEIRA, S. F.; BORINELLI, B.; DE NEGREIROS, L. F.; DALMAS, J. C. A Relação entre custo direto e desempenho escolar: uma análise multivariada nas escolas de ensino fundamental de Londrina/Pr. **Educação em Revista**, 31(1), 2015. pp.169-194.

BORGES, L. E. **Python para Desenvolvedores**. Novatec, São Paulo 2014.

CAPOVILLA, A. G. S.; DIAS, N. M. Desenvolvimento de estratégias de leitura no ensino fundamental e correlação com nota escolar. *Psicol. rev.* (Belo Horizonte), Belo Horizonte, v. 13, n. 2, p. 363-382, dez. 2007. Disponível em: <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-11682007000200010&lng=pt&nrm=iso>. Acesso em: 21 maio 2018.

CASTRUCCI, P. L. **Modelos Computacionais para Gestão: Princípios e Aplicações**. Editora Manole, 2005.

CYBIS, W. A.; BETIOL, A. H.; FAUST, E. **Ergonomia e Usabilidade: Conhecimentos, Métodos e Aplicações**. Novatec, 2015.

DA SILVA, G. P. Análise de evasão no ensino superior: uma proposta de diagnóstico de seus determinantes. **Avaliação: Revista da Avaliação da Educação Superior**, 18(2), 2013.

ESMERALDO, G.; ESMERALDO, D.; ALVES, M.; FERRAZ, J. **Uma Solução de Datacenter para Suporte às Pesquisas na Região do Cariri**. In: Revista Conexões, 9(1), 2015.

FEITOSA, R. G. F.; ESMERALDO, G. A.; DE OLIVEIRA, C. C. F. ESTIMATIVA DE DESEMPENHO ACADÊMICO A PARTIR DE MÉTODOS ESTATÍSTICOS PREDITIVOS: UM ESTUDO DE CASO COM ALUNOS DO 1º ANO DO ENSINO MÉDIO DO INSTITUTO FEDERAL DO CEARÁ, CAMPUS CRATO. **Revista Conexões - Ciência e Tecnologia**, 7(1), 2013.

FLANAGAN, D. **JavaScript - O guia definitivo**, 6ª edição. Editora Bookman, 2013.

FREEMAN, E. **Use a cabeça!**: programação em HTML 5: Desenvolvendo aplicativos para web com JavaScript/Eric Freeman, Elisabeth Robson - Rio de Janeiro: Alta Books, 2014.

GAMMA, E.; HELM, R.; JOHNSON, R. E.; VLISSIDES, J. **Design Patterns: Elements of Reusable Object-oriented Software**. Pearson Education, 2015.

MONIRUZZAMAN, A. B. M.; HOSSAIN, S. A. Nosql database: New era of databases for big data analytics-classification, characteristics and comparison". **International Journal of Database Theory and Application**, Vol. 6, No. 4, 2013.

OZGUR, C.; KLECKNER, M.; LI, Y.; **Selection of Statistical Software for Solving Big Data Problems: A Guide for Businesses, Students, and Universities**. *SAGE Open*, 5(2), 2015.

PRADHANANGA, Y.; KARANDE, S; KARANDE, C. **High performance analytics of bigdata with dynamic and optimized hadoop cluster**. In: IEEE International Conference on Advanced Communication Control and Computing Technologies, 2016. pp. 715-720.

SILVESTRE, A. L. Análise de Dados e Estatística Descritiva. Editora Escolar, 2007.

LUTZ, M. **Learning Python: Powerful Object-Oriented Programming**. 5. ed. United States of America: O'Reilly Media, Inc., 2013. 1600 p.

R Core Team. **R: A Language and Environment for Statistical Computing.** Vienna, Austria, 2017.
Disponível em: <<https://www.R-project.org/>>.

SILVA, M. S. **CSS3: desenvolva aplicações web profissionais com uso dos poderosos recursos de estilização das CSS3.** Novatec Editora, São Paulo, 2012.

TAJRA, S. F. **Internet na educação: o professor na era digital.** 1. ed. São Paulo: Érica, 2002.