



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

PEDRO FARIAS WANDERLEY

**USO DE BUSINESS INTELLIGENCE PARA AVALIAÇÃO DE
INDICADORES DE DESEMPENHO NA EDUCAÇÃO BÁSICA:
UM ESTUDO DE CASO NO ESTADO DO ACRE**

CAMPINA GRANDE - PB

2021

PEDRO FARIAS WANDERLEY

**USO DE BUSINESS INTELLIGENCE PARA AVALIAÇÃO DE
INDICADORES DE DESEMPENHO NA EDUCAÇÃO BÁSICA:
UM ESTUDO DE CASO NO ESTADO DO ACRE**

**Trabalho de Conclusão Curso
apresentado ao Curso Bacharelado em
Ciência da Computação do Centro de
Engenharia Elétrica e Informática da
Universidade Federal de Campina
Grande, como requisito parcial para
obtenção do título de Bacharel em
Ciência da Computação.**

Orientador: Professor Dr. Cláudio de Souza Baptista

CAMPINA GRANDE - PB

2021



W245u Wanderley, Pedro Farias.

Uso de Business Intelligence para avaliação de indicadores de desempenho na educação básica: um estudo de caso no Estado do Acre. / Pedro Farias Wanderley. - 2021.

13 f.

Orientador: Prof. Dr. Cláudio de Souza Baptista.

Trabalho de Conclusão de Curso - Artigo (Curso de Bacharelado em Ciência da Computação) - Universidade Federal de Campina Grande; Centro de Engenharia Elétrica e Informática.

1. Business Intelligence - técnica. 2. Indicadores de desempenho - educação básica. 3. Sistemas de apoio à decisão. 4. Desempenho escolar. 5. Painéis de desempenho. 6. Modelagem de banco de dados. 7. Indicadores educacionais - banco de dados. 8. Data Definition Language. 9. Sistema de Business Intelligence. I. Baptista, Cláudio de Souza. II. Título.

CDU:004.424.25(045)

Elaboração da Ficha Catalográfica:

Johnny Rodrigues Barbosa
Bibliotecário-Documentalista
CRB-15/626

PEDRO FARIAS WANDERLEY

**USO DE BUSINESS INTELLIGENCE PARA AVALIAÇÃO DE
INDICADORES DE DESEMPENHO NA EDUCAÇÃO BÁSICA:
UM ESTUDO DE CASO NO ESTADO DO ACRE**

**Trabalho de Conclusão Curso
apresentado ao Curso Bacharelado em
Ciência da Computação do Centro de
Engenharia Elétrica e Informática da
Universidade Federal de Campina
Grande, como requisito parcial para
obtenção do título de Bacharel em
Ciência da Computação.**

BANCA EXAMINADORA:

**Professor Dr. Cláudio de Souza Baptista
Orientador – UASC/CEEI/UFCG**

**Professor Dr. Adalberto Cajueiro de Farias
Examinador – UASC/CEEI/UFCG**

**Professor Tiago Lima Massoni
Professor da Disciplina TCC – UASC/CEEI/UFCG**

Trabalho aprovado em: 25 de Maio de 2021.

CAMPINA GRANDE - PB

ABSTRACT

In education, the massive amount of data produced annually is a critical asset to the area's progress. And progress in this area is essential to the human, economic, and social development of a country. Decision Support Systems, in particular the ones that apply Business Intelligence (BI) techniques, have a vital role in management. These systems can translate massive amounts of information into a more straightforward and friendly language, aiding managers' decision-making processes and information sharing within the community. With that in mind, we present a BI solution that uses interactive panels to display educational key performance indicators (KPIs). These panels contain information about primary and secondary school performance from cities in the Brazilian state of Acre. Our solution has the potential to help public managers in the evaluation of the efficiency and efficacy of public policies in education. It can also help in planning new guidelines and actions to improve performance in primary and secondary schools.

Uso de Business Intelligence para Avaliação de Indicadores de Desempenho na Educação Básica: Um Estudo de Caso no Estado do Acre

Pedro Farias Wanderley
pedro.wanderley@ccc.ufcg.edu.br
Universidade Federal de Campina Grande

Cláudio de Souza Baptista
baptista@computacao.ufcg.edu.br
Universidade Federal de Campina Grande

RESUMO

Na área de educação, a quantidade massiva de dados que são produzidos anualmente é um ponto chave para o progresso deste segmento, que é tão importante para o desenvolvimento humano, econômico e social de uma nação. Sistemas de Apoio à Decisão, principalmente os que utilizam técnicas de *Business Intelligence* (BI), têm um papel crucial na gestão, pois conseguem traduzir tais informações para uma linguagem mais amigável e direta, facilitando os processos decisórios de gestores e levando informação para a população em geral. Neste trabalho, apresentamos uma solução de BI voltada para a construção de painéis de indicadores de desempenho (KPI) que, de forma interativa, fornecem informações de desempenho escolar na educação básica em municípios do estado do Acre. Essa solução tem potencial para auxiliar os gestores na avaliação da eficiência e eficácia das políticas públicas educacionais implantadas, bem como para planejar novas diretrizes e ações para melhorar o desempenho na educação básica.

Palavras-chave

Business Intelligence, ETL, Indicadores de Desempenho, KPI, Educação

1 INTRODUÇÃO

Na gestão pública, uma das maiores dificuldades enfrentadas pelos gestores federais, estaduais e municipais, é o acompanhamento dos índices e indicadores de cada um dos setores que compõem uma determinada localidade. A imensa gama de setores que contemplam a governança pública, como educação, saúde, segurança, e mobilidade urbana, carecem de ferramentas que sintetizam informações e facilitam a gestão de políticas públicas e processos decisórios.

No setor da educação, as esferas administrativas necessitam de um meio que viabilize um entendimento concreto e ágil da real situação educacional de suas localidades. Na maioria dos casos, as informações vinculadas à educação estão dispostas através de dados brutos de difícil entendimento e leitura, em forma de microdados, resultantes de pesquisas educacionais realizadas por entidades públicas, ou em séries históricas, muitas vezes com lapsos temporais. Em seu formato original, esses dados não têm tanto poder informacional para gestores públicos e a sociedade em geral, pois o

formato utilizado torna seu entendimento e leitura complexos e a informação apresentada abstrata.

O entendimento de dados obtidos através de avaliações e pesquisas censitárias, como o censo escolar¹, é crucial para que se possa ter uma noção do que precisa ser melhorado, o que deve ser mantido, quais as perspectivas para o futuro e quais ações devem ser tomadas a curto, médio e longo prazo. Além desse entendimento, um estudo detalhado sobre quais indicadores educacionais são relevantes para um determinado contexto, é fundamental para que haja um *feedback* construtivo e focal, atendendo cada localidade de forma particular, com ações voltadas às suas necessidades.

Sistemas de *business intelligence* (BI) combinam arquiteturas, bancos de dados, ferramentas analíticas, aplicativos e metodologias visando dar suporte à tomada de decisão. O principal objetivo do BI é possibilitar o acesso interativo à informações e dados, para assim permitir que gerentes de negócios e analistas possam conduzir análises apropriadas. Através do processo de análise dessas informações, históricas ou atuais, os indivíduos responsáveis pela tomada de decisão obtêm *insights* valiosos que os permitem tomar decisões melhores e mais informadas [11]. Sistemas de BI apresentam-se como soluções ideais para tarefas analíticas, como a análise de índices educacionais [12, 13].

A integração de soluções de BI no âmbito da gestão educacional tem uma capacidade imensa de não só compilar informações de diversas fontes e fornecer ferramentas de fácil acesso e entendimento para gestores e cidadãos, mas permite, acima de tudo, ter uma visão sistêmica da educação [15]. Esta visão sistêmica habilita a aferição das métricas de eficiência, eficácia e efetividade nas políticas públicas voltadas para a educação.

Este artigo propõe uma arquitetura de BI para indicadores de desempenho em educação, com integração de dados de diversas fontes, possibilitando ao gestor de educação um avanço no processo de tomada de decisão. Os indicadores aqui propostos advêm do Projeto Integrar², projeto que é fruto de uma parceria do Tribunal de Contas da União³ (TCU) com a Organização para Cooperação e Desenvolvimento Econômico⁴ (OCDE) e vários Tribunais de Contas dos estados brasileiros. Foi implementado um estudo de caso para o estado do Acre, pela facilidade no acesso às informações desse estado. No entanto, o sistema proposto pode ser facilmente adaptado para ser utilizado em qualquer estado do Brasil, bastando que se tenham os dados necessários para representar os indicadores de desempenho propostos.

Os autores retêm os direitos, ao abrigo de uma licença Creative Commons Atribuição CC BY, sobre todo o conteúdo deste artigo (incluindo todos os elementos que possam conter, tais como figuras, desenhos, tabelas), bem como sobre todos os materiais produzidos pelos autores que estejam relacionados ao trabalho relatado e que estejam referenciados no artigo (tais como códigos fonte e bases de dados). Essa licença permite que outros distribuam, adaptem e evoluam seu trabalho, mesmo comercialmente, desde que os autores sejam creditados pela criação original.

¹<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar>

²<https://sites.tcu.gov.br/integrar>

³<https://portal.tcu.gov.br>

⁴<https://www.oecd.org>

O presente trabalho teve como principal objetivo a construção de painéis interativos (do inglês, *dashboards*), que trazem informações atuais e séries históricas, sobre indicadores educacionais dos municípios do estado do Acre. O intuito deste projeto é levar informação para cidadãos comuns sobre a situação dos índices educacionais das suas localidades. Além disso, os painéis interativos funcionam como uma ferramenta de suporte à tomada de decisão de gestores educacionais, sempre visando a melhoria dos índices vinculados à educação básica.

O restante deste artigo está organizado como segue. Na seção 2 são apresentados os trabalhos relacionados. Na seção 3 é apresentada a proposta de solução do problema. Na seção 4, os principais resultados obtidos são apresentados. Na seção 5 são expostos os resultados do processo de avaliação da solução e por fim, na seção 6, são abordadas as considerações finais e perspectivas para trabalhos futuros.

2 TRABALHOS RELACIONADOS

Sistemas de apoio à decisão, utilizando-se de *business intelligence* e indicadores de desempenho, vêm sendo amplamente explorados no domínio da gestão educacional em todo o mundo. Diepen e Bredeeweg usaram algoritmos genéticos para determinar indicadores de aprendizagem para alunos da educação secundária [12]. Weeda et al. investigaram indicadores que melhor caracterizam a qualidade de código em programas elaborados por alunos da educação secundária [14]. Rezgui et al. usaram indicadores de desempenho para aprimorar processos de apoio à tomada de decisão baseados em sistemas de gestão de qualidade aplicado ao ensino superior [9]. Petiot apresentou um experimento para o cálculo de indicadores educacionais usando técnicas de mineração de dados e sistemas de apoio à decisão [8]. Vital et al. utilizaram técnicas de aprendizagem não supervisionada, através de algoritmos de clustering (hierárquico e *k-means*) para analisar o desempenho de estudantes [13]. Panic e Lozanov-Crvenkovic usaram modelos financeiros e regressão linear para formular modelos de predição do comportamento de indicadores relacionados ao desempenho de alunos no ensino superior [7]. Moscoso-Zea et al. propuseram uma infraestrutura de *business intelligence & analytics* (BI&A) e gestão do conhecimento para um *datawarehouse* educacional com repositório para digitalização do conhecimento e análise de componentes organizacionais como pessoas, processos e tecnologias [6]. Salaki e Ratnam focaram na importância de adotar um processo ágil de *analytics* no desenvolvimento de *business intelligence* e *datawarehouse* em instituições de ensino superior [10]. Zhu apresentou um sistema de apoio à decisão dirigido a dados como uma plataforma de auxílio do processo de tomada de decisão educacional [15].

3 SOLUÇÃO PROPOSTA

Nesta seção é apresentada nossa solução de BI para acompanhamento e melhoria das políticas públicas educacionais no estado do Acre. Como a solução teve como base o projeto Integrar, o primeiro passo do trabalho foi um estudo detalhado sobre os indicadores apontados no documento do projeto [2]. O intuito nessa primeira etapa foi entender e avaliar quais desses indicadores seriam essenciais para compor os painéis, de acordo com o contexto do estado, e quais seriam dispensáveis, visto que uma das grandes diferenças em

relação ao que fora proposto no projeto Integrar é que os painéis seriam mais sucintos, de forma a evitar um excesso de informações que viessem a confundir os usuários da ferramenta de BI.

3.1 Arquitetura do BI de Educação

A arquitetura da solução de BI de educação, mostrada na figura 1, consiste basicamente em três camadas: o processo de *Extract, Transform and Load* (ETL) dos dados originais; a implementação de um banco de dados relacional, e a criação do *front end* de BI com os indicadores educacionais, organizados em painéis interativos.

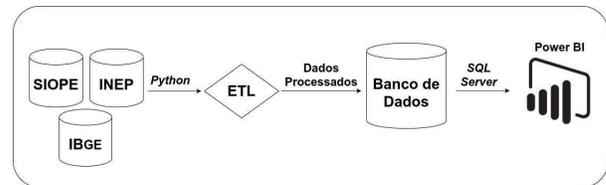


Figura 1: Arquitetura da solução

O processo de ETL foi realizado através de scripts escritos em Python, que acessaram os dados originais em planilhas ou bancos de dados relacionais, realizaram transformações nos dados e fizeram a carga dos dados no banco de dados.

O banco de dados foi implementado utilizando o sistema de gerenciamento de banco de dados (SGBD) Microsoft SQL Server 2019⁵. Esse SGBD possui um ferramenta facilitadora, o SQL Server Management Studio⁶, um ambiente integrado com interface gráfica para gerenciar infraestruturas sql, como realizações de consultas e criação de diagramas de bancos de dados.

Para o *front end*, foi utilizado o Power BI, que é um *framework* para desenvolvimento de painéis e relatórios voltados para a área de BI. Esse *framework* provê uma grande quantidade de funcionalidades para exibição dos indicadores. Além disso, o Power BI fornece suporte para *Data Analysis Expressions* (DAX), uma linguagem de expressão de fórmulas, que permite a criação de consultas dinâmicas, cálculos, dentre outras funcionalidades, integradas diretamente com os dashboards desenvolvidos na ferramenta.

3.2 Modelagem do Banco de Dados de Indicadores Educacionais

Após o estudo do material fornecido pelo projeto Integrar, que serviu de base para o desenvolvimento deste trabalho, foi iniciado o processo de modelagem do banco de dados. O banco de dados representa o ponto central no processo analítico posteriormente provido pelos painéis informativos. Nesta etapa, foram executados processos de *SQL Data Definition Language* (DDL), com criação de tabelas e colunas, com respectivos índices e *constraints*.

Depois de ter-se feito um estudo detalhado sobre as fontes de dados, e o que seria aproveitado das mesmas, foi desenhado o esquema que representaria a base de dados utilizada na construção dos painéis. Devido aos dados serem bem estruturados, e o volume de dados não ser *big data* (o tamanho do banco de dados é de aproximadamente 145 MB), optou-se por utilizar um banco de dados

⁵<https://www.microsoft.com/pt-br/sql-server/sql-server-downloads>

⁶<https://docs.microsoft.com/pt-br/sql/ssms/sql-server-management-studio-ssms>

Tabela (Indicador)	Descrição
ADEQUACAO_FORMACAO_DOCENTE	Porcentagem de docentes em cada um dos níveis de adequação de formação
ATENDIMENTO_EDUCACAO_INFANTIL	Quantidade de matrículas nas creches e pré-escolas
AVALIACAO_NAC_ALFABETIZACAO	Desempenho nas avaliações de escrita, leitura e matemática
DISTORCAO_IDADE_SERIE	Taxas de distorção idade-série nas diferentes etapas de ensino
ESFORCO_DOCENTE	Porcentagem de docentes em cada nível da variável categórica que indica o nível de esforço do docente
LOCALIDADE_ENSINO	Localidade do ensino (rural ou urbana)
MEDIA_ALUNOS	Média de alunos por turma em cada etapa de ensino
MUNICIPIO	Localidade, podendo ser município, microrregiões e mesorregiões
PROVA_BRASIL	Resultados nas provas de matemática e língua portuguesa da Prova Brasil
REDE_ENSINO	Rede de ensino, podendo ser: municipal, estadual, federal e privada
TIPO_ENSINO	Tipo de ensino escolar (parcial, integral)
TRANSICAO_EVASAO	Taxas de evasão nas diferentes etapas de ensino
TRANSICAO_REPETENCIA	Taxas de repetência nas diferentes etapas de ensino

Tabela 1: Tabelas do Banco de Dados

relacional, que é uma solução simples e eficiente para este tipo de dados. Foi escolhido o Microsoft SQL Server que é um SGBD relacional bastante utilizado no mercado. Esse sistema situa-se no primeiro quadrante do quadrante mágico do Gartner⁷ sendo portanto uma das ferramentas mais atrativas e utilizadas no mercado, tem uma interface amigável, bom desempenho, e provê facilidades de integração com o Power BI.

No primeiro momento, foram selecionados dez indicadores, e três entidades voltadas para os filtros da interface, os componentes do esquema estão dispostos na tabela 1.

O relacionamento entre as entidades é formado por três indicadores centrais, município, rede e localidade. O indicador município pode ser estendido para os sub indicadores de microrregião e mesorregião. Essas subcategorias de localidade vinculadas aos municípios em sistemas de BI podem ser exploradas por operações conhecidas como *drill down* e *drill up*, que aumentam ou diminuem a granularidade da informações dispostas em tela, fornecendo uma visão mais ou menos detalhada vinculada a um filtro. O indicador rede se refere a rede de ensino pública ou privada, e tem como subgrupo de filtros as sub redes: municipal, estadual e federal. Por fim, o indicador localidade diz respeito à localização do ambiente escolar, indicando se a escola encontra-se no meio rural ou urbano.

Os três componentes mencionados acima são essenciais no processo comparativo, possibilitando a introdução de relacionamentos cruzados entre os indicadores disponíveis, além de filtragens dinâmicas. A figura 2 apresenta o diagrama relacional do banco de dados modelado.

3.3 O Processo de ETL

Com as fontes definidas e a base de dados modelada, foi realizado um dos processos mais importante na construção de um sistema de BI, que é a etapa de ETL [4]. ETL é um processo de extração, transformação e carregamento de dados de diversas fontes para

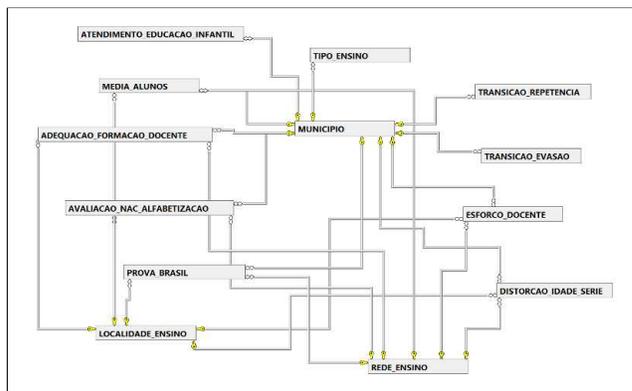


Figura 2: Diagrama Relacional do Banco de Dados de Indicadores da Educação Básica

uma base de dados. A etapa de extração de dados pode ser originada de diversas fontes de dados heterogêneas como planilhas, arquivos JSON, bancos de dados, páginas web, dentre outros. A etapa de transformação consiste na limpeza e tratamento de dados, que são advindos de diversas fontes, além de operações de modificação e cruzamento de informações. A etapa de carga consiste na inserção dos dados transformados ou não, no banco de dados, que abriga grandes quantidades de dados das diversas fontes utilizadas.

O primeiro passo do ETL foi a extração dos dados das fontes oficiais definidas na etapa de estudo de indicadores e fontes de dados. Um dos grandes facilitadores foram os formatos de arquivos que os dados estavam disponibilizados, geralmente arquivos *Excel* (XLSX) e *comma-separated values* (CSV). Com os dados extraídos do respectivo formato de arquivos, foi realizada uma etapa de pré-processamento, com o intuito de eliminar ruídos das informações, como, por exemplo, a ocorrência de valores nulos, vazios e inconsistentes, além de um processo de filtragem, pois para os painéis, o foco seria as informações do estado do Acre.

⁷<https://azure.microsoft.com/en-us/resources/gartner-2020-magic-quadrant-for-cloud-database-management-systems>

Utilizando a linguagem Python, e bibliotecas como *pandas*⁸ e *pyodbc*⁹, o pré-processamento dos dados, além das operações de *Data Manipulation Language* (DML) voltadas para o banco de dados foram implementadas, removendo o trabalho manual de processar informações advindas dos arquivos originários. Os scripts Python desenvolvidos nesta etapa foram essenciais para a otimização de tempo e robustez no processo de ETL, principalmente no processo de leitura de dados advindos dos arquivos carregados das diversas fontes, que por vezes, continham um grande volume de informações. A manipulação destas informações de forma manual dificultaria o tratamento coerente de todo conteúdo dos arquivos.

3.3.1 Fontes de Dados. Os dados primários educacionais estão disponíveis em diversas fontes. As principais fontes de dados utilizadas foram:

- INEP¹⁰: possui uma seção de dados abertos com séries históricas de diversos indicadores, formada principalmente por metadados, a exemplo dos dados do Censo Escolar.
- IBGE¹¹: utilizada para informações geográficas e de contexto. As informações foram buscadas diretamente do Instituto Brasileiro de Geografia e Estatística (IBGE), que fornece informações sobre os municípios e suas características.
- SIOPE¹²: contém dados sobre orçamentos e contas públicas vinculadas à educação, disponibilizados através do portal do Fundo Nacional de Desenvolvimento da Educação (FNDE), via o Sistema de Informações sobre Orçamentos Públicos em Educação (SIOPE).

3.3.2 Scripts Python. De posse dos arquivos das fontes primárias, foram desenvolvidos scripts em Python, com três funções primordiais, ler os dados dos arquivos originários, manipular e transformar alguns desses dados, caso seja necessário, e por fim, conectar com o banco de dados e fazer manipulações nas tabelas, principalmente scripts DML, mais voltados para inserção de dados.

O processo de leitura e transformação de dados foi feito com o auxílio da biblioteca *pandas* que fornece ferramentas de análise de dados e estrutura de dados de alta performance, onde é possível ler arquivos extensos e manipular informações neles contidas. Já para conectar com o banco de dados, foi utilizada a biblioteca *pyodbc* que facilita a conexão e manipulação direta com bancos de dados relacionais, simplificando o processo de acesso ao banco de dados. Os scripts desenvolvidos estão disponíveis no repositório¹³ do projeto.

3.4 Projeto dos Dashboards

Com os indicadores definidos e os dados carregados no banco de dados, foi necessário um trabalho de *User Experience* (UX) e *User Interface* (UI), para desenvolver o *front end* da aplicação. Por se tratar de um número grande e diversificado de indicadores, foi definido, nesta etapa, a estrutura da interface e organização dos seus componentes.

O agrupamento de indicadores seguiu o que foi proposto originalmente pelo projeto Integrar, que recomendou a separação dos indicadores em dois grupos: indicadores de acesso à educação e indicadores de proficiência. Os indicadores de acesso dizem respeito à viabilidade, facilidade e condições em geral que possibilitam o ingresso, a permanência e o progresso de um estudante nas inúmeras etapas de ensino que compõem a matriz curricular da educação básica. Por sua vez, os indicadores de proficiência estão atrelados aos resultados obtidos por esses estudantes em avaliações censitárias, como a Prova Brasil e a Avaliação Nacional da Alfabetização (descontinuada em 2016), que permitem a dedução de *insights* sobre os índices de aprendizagem do corpo discente. Com essa definição, os usuários finais podem seguir uma sequência lógica no momento de visualização dos painéis, facilitando o entendimento das informações dispostas.

Como o foco do trabalho são os indicadores educacionais, foram excluídos do planejamento dos painéis os indicadores de contexto e indicadores de financiamento, ambos propostos pelo projeto Integrar.

Definido o *layout* da interface, foi iniciado o desenvolvimento dos painéis. Para a homologação de cada versão finalizada, utilizou-se uma metodologia baseada em provas de conceito (POC), devido ao enorme número de possibilidades de funcionalidades fornecidas pelo Power BI. À medida em que as POCs eram desenvolvidas, as mesmas eram submetidas à análise de um usuário independente, que tinha a função de opinar sobre as visualizações escolhidas, ponto fundamental para garantir uma melhor experiência do usuário.

4 RESULTADOS

Os *dashboards* desenvolvidos contemplaram os dois grupos definidos no projeto Integrar: os indicadores de acesso e os indicadores de proficiência. Foram implementados os seguintes painéis:

- Painel Distorção Idade-Série: informações sobre as taxas de distorção idade-série na educação básica;
- Painel Esforço Docente: informações sobre as taxas de cada um dos níveis de esforço docente;
- Painel Adequação da Formação Docente: informações sobre as taxas de formação docente adequada para cada um dos grupos;
- Painel Prova Brasil: resultados e comparativos dos resultados nas avaliações de língua portuguesa e matemática da prova Brasil;
- Painel Avaliação Nacional da Alfabetização: resultados das avaliações aplicadas nas modalidades de leitura, escrita e matemática;
- Painel Evasão Escolar: taxas de evasão escolar nas diferentes etapas de ensino;
- Painel Repetência Escolar: taxas de repetência escolar nas diferentes etapas de ensino;
- Painel Atendimento à Educação Infantil: informações sobre a quantidade de alunos matriculados nas creches e pré-escolas do estado;
- Painel Média de Alunos por Turma: informações sobre a média de alunos por turma nas etapas de ensino da educação básica;

⁸<https://pandas.pydata.org/docs>

⁹<https://github.com/mkleehammer/pyodbc/wiki>

¹⁰<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos>

¹¹<https://www.ibge.gov.br/>

¹²https://www.fnnde.gov.br/fnde_sistemas/siope

¹³https://github.com/pedrofawanderley/scripts_tcc_indicadores_educacionais_Acre/tree/main/scripts

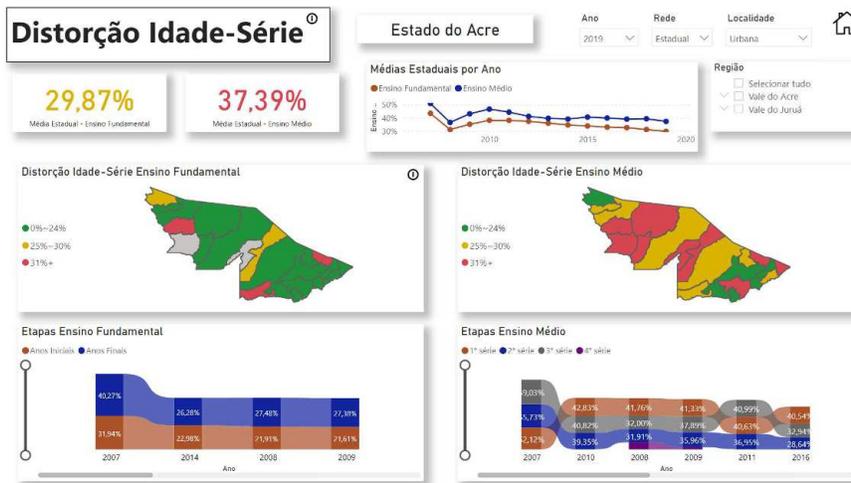


Figura 3: Painel do indicador de Distorção Idade-Série

- Painel Ensino Integral: informações sobre a oferta de ensino integral para as etapas de ensino da rede pública;

Devido à restrição de espaço deste documento são destacados a seguir apenas alguns dos painéis implementados. Todos os painéis desenvolvidos estão disponíveis no repositório¹⁴ do projeto.

4.1 Painel Distorção Idade-Série

O indicador de distorção idade-série acompanha a taxa de alunos que têm idade acima da esperada para a etapa de ensino em que estão matriculados. O painel desenvolvido para esse indicador é composto por um conjunto de filtros, município, rede de ensino, localidade de ensino e ano. Além disso, como existem dados de todo o estado do Acre, é possível realizar operações de *drill up* e visualizar dados das mesorregiões e microrregiões do estado, conforme apresentado na figura 3.

Na figura 3, os dois indicadores abaixo do título trazem informações sobre a média estadual da taxa de distorção idade-série, para o ensino fundamental (superior esquerdo) e ensino médio (superior direito), onde sua coloração muda de acordo com o valor da taxa. O gráfico de linhas acima dos mapas faz um comparativo entre as médias estaduais das taxas de distorção idade-série ao longo dos anos, para o ensino fundamental e médio, podendo assim ter uma visão geral sobre o comportamento das taxas para cada etapa de ensino em função do tempo. Os dois mapas trazem as taxas de distorção idade-série vinculadas ao ensino fundamental (mapa à esquerda) e ao ensino médio (mapa à direita), de acordo com a cor é possível se ter uma noção se o nível do indicador está crítico (vermelho), regular (amarelo) ou ideal (verde) para cada um dos municípios.

Por fim, os gráficos de faixas na parte inferior do painel mostram um comparativo entre as taxas de distorção idade-série nos anos iniciais e finais do ensino fundamental (inferior esquerdo) e um comparativo entre as quatro séries do ensino médio (inferior direito), mostrando uma série histórica com dados vinculados a um

município, caso seja ativado o filtro do mesmo. Os filtros disponíveis para esse painel são: rede de ensino, localidade de ensino, ano, município, mesorregião e microrregião.

4.2 Painel Prova Brasil (Língua Portuguesa)

A Prova Brasil é uma avaliação censitária aplicada bianualmente, conta com provas de língua portuguesa e matemática, e tem como principais objetivos medir a qualidade do ensino público, bem como observar o nível de aprendizagem dos alunos. A figura 4 exibe este painel com mapas e gráficos referentes ao desempenho na Prova Brasil, mais especificamente na avaliação de língua portuguesa.

Os mapas trazem os valores das médias municipais, o mapa superior é referente aos anos iniciais do ensino fundamental, já o mapa inferior é referente aos anos finais, a cor em cada município segue uma escala definida por uma equipe científica do movimento Todos pela Educação¹⁵. Mais à direita na figura 4, os indicadores trazem informações sobre as médias estaduais na prova, também diferindo pela etapa de ensino. A coloração do valor da média segue a mesma escala dos mapas temáticos.

Os gauges, no centro do painel, indicam se o município alcançou a média estadual para sua etapa de ensino, no caso, os anos iniciais e finais. Caso o valor da média municipal ultrapasse ou seja igual à média estadual, a cor do gauge será verde, caso o valor fique abaixo da média, o preenchimento fica vermelho. Na figura 4, para o município acreano de Sena Madureira, o resultado dos anos iniciais do ensino fundamental ficou abaixo da média estadual, já nos anos finais, o valor ficou maior que a média do estado.

O gráfico de faixas (inferior direito) é um comparativo das médias de resultados municipais entre os anos iniciais e finais, e tem um caráter temporal, sendo possível uma avaliação dos resultados de anos anteriores. Todos os componentes desse painel podem ser filtrados pelo ano, rede e localidade de ensino, município, mesorregião e microrregião.

¹⁴https://github.com/pedrofwanderley/scripts_tcc_indicadores_educacionais_Acre

¹⁵<https://todospelaeducacao.org.br/>

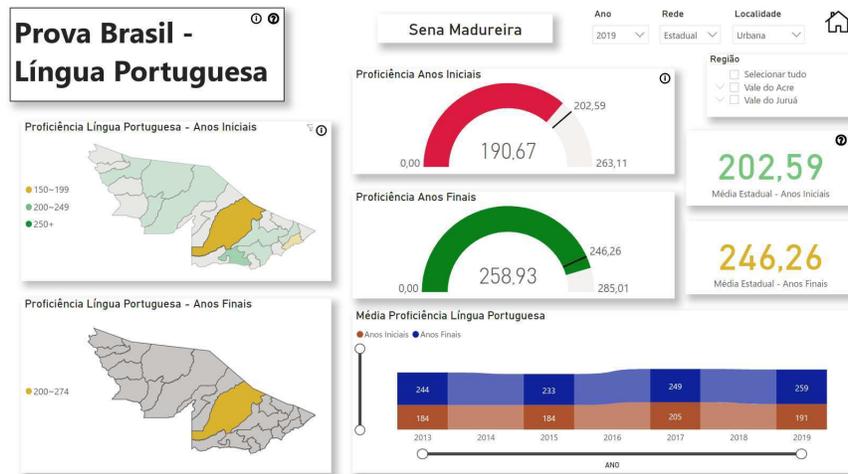


Figura 4: Painel do indicador de Prova Brasil, para a avaliação de Língua Portuguesa

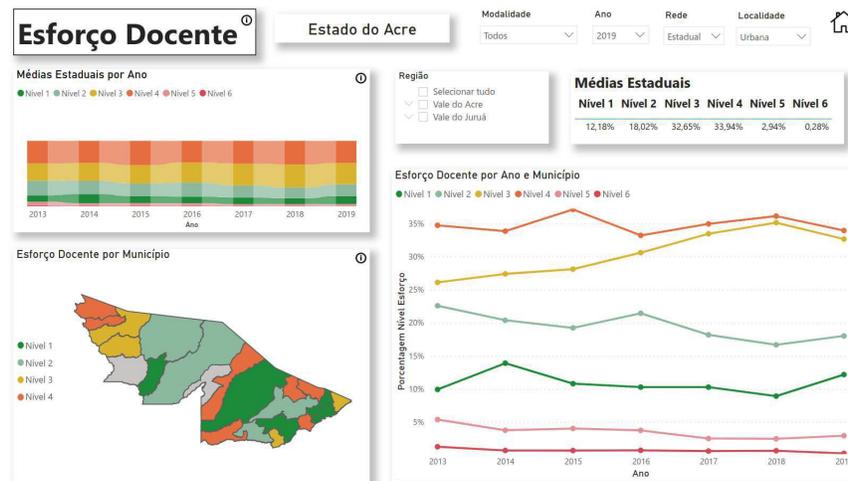


Figura 5: Painel do indicador de Esforço Docente

4.3 Painel Esforço Docente

Este indicador é calculado utilizando dados disponíveis no Censo Escolar da Educação Básica e leva em consideração quatro pontos, segundo nota técnica¹⁶ do INEP: número de escolas que o docente atua, número de turnos de trabalho, número de alunos atendidos e o número de etapas de ensino que o mesmo leciona.

Existem seis níveis de esforço docente, que, em ordem crescente, representam a carga de trabalho de um docente, sendo o nível 1, o de menos esforço e o nível 6, o nível mais crítico. A figura 5 mostra o painel deste indicador. O primeiro gráfico (superior esquerdo) traz os valores das médias estaduais ao longo dos anos, e representa a porcentagem de docentes em cada um dos níveis de esforço docente. O mapa mostra qual dos níveis teve uma maior porcentagem de docentes em cada um dos municípios.

No lado superior direito da figura 5 a tabela traz os valores das médias estaduais para cada um dos níveis de esforço docente, vinculadas ao ano selecionado no filtro da tela. O gráfico abaixo da tabela mostra uma série histórica das porcentagens de cada nível de esforço docente para um determinado município, respeitando os filtros de rede e localidade de ensino. Além da rede e localidade de ensino os componentes do painel esforço docente podem ser filtrados pela etapa de ensino, ano, município, mesorregião e microrregião.

4.4 Painel Evasão Escolar

A evasão escolar faz referência ao aluno que foi reprovado em um dado ano, ou abandonou a escola, e no ano seguinte não fez a matrícula novamente para dar continuidade aos estudos.

O painel para este indicador, mostrado na figura 6, traz dois mapas temáticos, o superior está vinculado às porcentagens de evasão no ensino fundamental e o inferior está vinculado às porcentagens

¹⁶ https://download.inep.gov.br/informacoes_estatisticas/indicadores_educacionais/2014/docente_esforco/nota_tecnica_indicador_docente_esforco.pdf

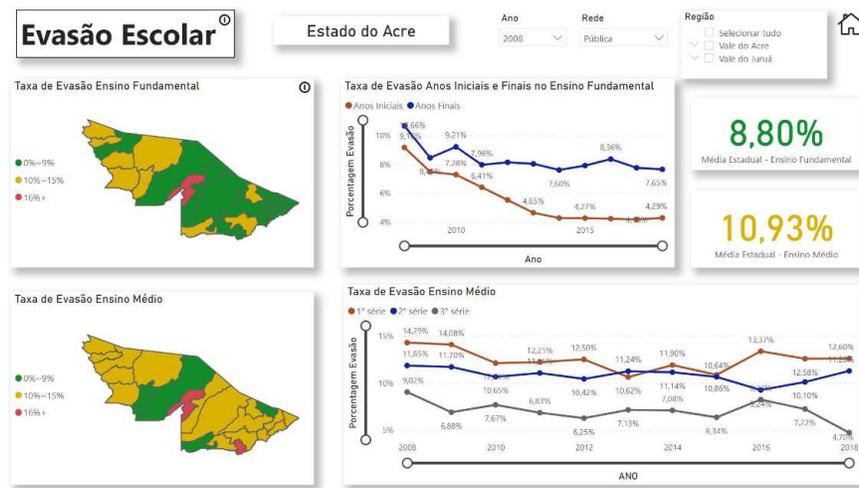


Figura 6: Painel do indicador de Evasão Escolar

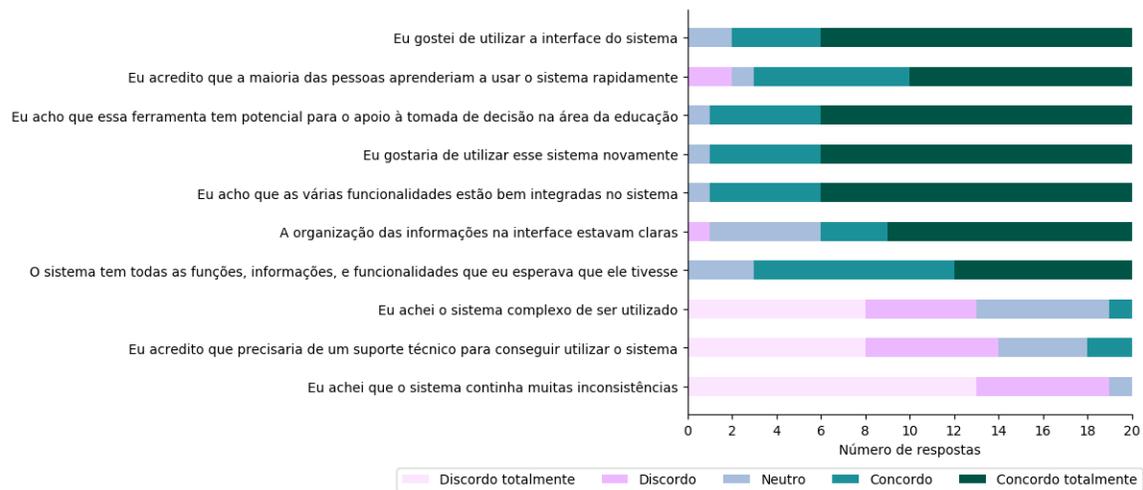


Figura 7: Resultados do questionário aplicado

de evasão no ensino médio, onde a cor que preenche o município muda de acordo com o valor daquela localidade específica, indo de desejável (verde) a crítico (vermelho). Os gráficos de linhas trazem séries históricas sobre os valores deste indicador, também voltados para o ensino fundamental e médio. No gráfico superior, é mostrada uma comparação entre os anos iniciais e finais do ensino fundamental, já no gráfico inferior um comparativo entre os valores de evasão escolar para a 1ª, 2ª e 3ª série do ensino médio.

Os dois indicadores mais à esquerda da figura 6, representam os valores das médias estaduais dos índices de evasão para o ensino fundamental e médio, mudando sua coloração de acordo com a escala de valores presente na legenda do mapa. Por fim, os componentes desse painel podem ser filtrados por município, microrregião, mesorregião, rede de ensino e ano.

5 AVALIAÇÃO DE USABILIDADE

O processo de avaliação de usabilidade da ferramenta de BI proposta consistiu na aplicação de um questionário que visou averiguar o nível de usabilidade do sistema. O questionário foi baseado no método *System Usability Scale* (SUS) [1, 5]. O SUS é um questionário que se propõe a avaliar a usabilidade de sistemas. O questionário é composto por 10 afirmativas, e 5 opções de respostas, que indicam o nível de concordância do usuário com o que está sendo afirmado.

Para a avaliação do painel foi utilizada uma escala Likert [3] que quantifica o grau de concordância dos usuários para com as afirmativas feitas no questionário. Para cada afirmativa o usuário poderia selecionar um número de 1 a 5, onde 1 significa discordância total com o que está sendo afirmado, e 5 concordância total, o valor 3, portanto, é um valor neutro.

O processo de avaliação contou com 20 indivíduos, que anonimamente responderam ao questionário. Dentre os quais, alguns indivíduos eram gestores educacionais. Para a avaliação, foi disponibilizada uma versão do sistema, acompanhada de um vídeo instrucional demonstrando como se utiliza a ferramenta, sobretudo as consultas dinâmicas que o Power BI possibilita realizar.

A figura 7, traz informações sobre as perguntas do formulário, bem como as respostas para cada um dos itens. Pode-se concluir que o resultado foi excelente, considerando o alto grau de concordância dos usuários com o que diziam as afirmações do questionário, com exceção dos últimos três itens mostrados na figura 7, que tratavam sobre a complexidade e inconsistências no sistema, nessas afirmativas o grau de discordância foi alto, o que representa um ótimo indicio.

Os resultados para duas afirmativas presentes no questionário representam bem o potencial da solução para atender os dois principais públicos para os quais a solução foi desenvolvida: usuários cidadãos e gestores. A primeira afirmativa diz respeito ao item 3 ("*Eu acho que essa ferramenta tem potencial para o apoio à tomada de decisão na área da educação*") que afirma que o sistema tem potencial para auxiliar processos de tomada de decisão no âmbito educacional, esse item alcançou uma média de respostas de 4,65 indicando a validade para utilização da ferramenta por gestores educacionais. A segunda afirmativa representada pelo item 8, faz referência a complexidade do sistema ("*Eu achei o sistema complexo de ser utilizado*") e obteve como média de resposta o valor 2, indicando uma alta discordância com o que está sendo dito, demonstrando por tanto ser um sistema simples de ser utilizado para ambos os públicos.

Diante das respostas dos usuários conclui-se que a ferramenta de BI educacional proposta atende aos dois públicos-alvo para os quais a ferramenta foi proposta: usuários cidadãos e gestores em educação. Salientamos que, para os gestores, pode haver um treinamento específico com uso da ferramenta de BI, permitindo aos mesmos atingir os objetivos de apoio à tomada de decisão em políticas públicas na área de educação.

6 CONCLUSÃO

Devido ao curto período para a realização deste trabalho, bem como dificuldades enfrentadas, como informações imprecisas e séries históricas curtas, o principal objetivo de desenvolver painéis com narrativas de fácil entendimento para cidadãos e gestores, assim como trazer informações cruciais para a melhor gestão da educação básica nos municípios do estado Acre, se tornou o foco deste projeto.

A avaliação da ferramenta por usuários trouxe uma *feedback* muito positivo, e que comprovou a eficácia e importância da solução. Os painéis conseguiram com clareza levar informações para os usuários finais, que foram aptos a compreender como anda a situação educacional no estado do Acre. Desse modo, o BI educacional mostra-se como um importante instrumento para auxílio de gestores educacionais do estado do Acre.

Considerando que uma das principais limitações deste trabalho foi a quantidade limitada de indicadores analisados, recomenda-se para trabalhos futuros um estudo que envolva outros indicadores vinculados ao âmbito educacional, além de indicadores de contexto e econômicos, de modo a construir uma ferramenta que englobe o

máximo de indicadores que de algum modo possam influenciar nos índices educacionais não só do estado Acre, mas de qualquer outra localidade do país.

Recomenda-se também, a incorporação de um modelo ou algoritmo para direcionamento de gestores, sobre quais ações serão mais eficazes para o alavanque da educação básica em cada localidade específica. Através da análise de indicadores, como os expostos neste documento, e cruzamento de informações, utilizando técnicas de *machine learning*, o sistema utilizando o algoritmo, seria capaz de recomendar o direcionamento de recursos para um determinado setor da educação, otimizando o processo de tomada de decisão dos gestores, principalmente em função da grande quantidade de variáveis que se deve levar em consideração para esses processos decisórios.

Por fim recomenda-se a criação de um modelo educacional nacional, utilizando como base municípios que obtiveram resultados satisfatórios, vinculados à escolarização e aprendizado. Para tal modelo, técnicas de *machine learning* seriam ideais para análise de dados educacionais, com objetivo de encontrar uma solução ideal para impulsionar índices educacionais em todo território nacional.

AGRADECIMENTOS

Agradeço a todo o corpo docente e funcionários do curso de Ciência da Computação, da Universidade Federal de Campina Grande (UFCG) por me instruírem e compartilharem seus conhecimentos durante todo o período da graduação, essenciais na minha formação profissional. Agradeço especialmente ao meu mestre e orientador Prof. Dr. Cláudio de Souza Baptista por todas as oportunidades e ensinamentos compartilhados ao longo dos últimos cinco anos, que me forneceram uma base primordial na minha formação acadêmica e acima de tudo na minha construção humana e profissional.

Agradeço também a todos os funcionários da UFCG, coordenação, serviços gerais, segurança, por tornarem o meu percurso na universidade mais prazeroso. Agradeço a minha namorada Renata, por toda a paciência e suporte, principalmente nesse período final do curso. Agradeço por fim a minha família, aos meus pais por todo o suporte e estímulo durante a graduação e a minhas irmãs, em especial minha irmã e companheira de curso Leticia, por toda a confiança depositada em mim, além de todo aprendizado e ensinamentos compartilhados.

REFERÊNCIAS

- [1] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction* 24, 6 (2008), 574–594. <https://doi.org/10.1080/10447310802205776> arXiv:<https://doi.org/10.1080/10447310802205776>
- [2] Tribunal de Contas da União (TCU). 2020. *Projeto Integrar: propostas para o fortalecimento do controle externo de políticas públicas descentralizadas / Tribunal de Contas da União, Organização para Cooperação e Desenvolvimento Econômico*. TCU, Secretaria-Geral de Controle Externo (Segecex), Secretaria de Controle Externo da Educação (SecexEduc), Brasília.
- [3] Ankur Joshi, Saket Kale, Satish Chandel, and Dinesh Pal. 2015. Likert Scale: Explored and Explained. *British Journal of Applied Science Technology* 7 (01 2015), 396–403. <https://doi.org/10.9734/BJAST/2015/14975>
- [4] Ralph Kimball and Joe Caserta. 2004. *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data*. John Wiley Sons, Inc., Hoboken, NJ, USA.
- [5] James R. Lewis and Jeff Sauro. 2009. The Factor Structure of the System Usability Scale. In *Human Centered Design*, Masaaki Kurosu (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 94–103.
- [6] Oswaldo Moscoso-Zea, Jorge Castro, Joel Paredes-Gualtor, and Sergio Luján-Mora. 2019. A Hybrid Infrastructure of Enterprise Architecture and Business

- Intelligence & Analytics for Knowledge Management in Education. *IEEE Access* 7 (2019), 38778–38788. <https://doi.org/10.1109/ACCESS.2019.2906343>
- [7] Andrea Andrejevic Panic and Zagorka Lozanov-Crvenkovic. 2018. Panel Regression Analysis of the Financial and Performance Indicators in Higher Education in Central East Europe. In *16th IEEE International Symposium on Intelligent Systems and Informatics, Sisy 2018, Subotica, Serbia, September 13-15, 2018*. IEEE, 317–320. <https://doi.org/10.1109/SISY.2018.8524766>
- [8] Guillaume Petiot. 2018. Merging Information Using Uncertain Gates: An Application to Educational Indicators. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations - 17th International Conference, IPMU 2018, Cádiz, Spain, June 11-15, 2018, Proceedings, Part I (Communications in Computer and Information Science)*, Jesús Medina, Manuel Ojeda-Aciego, José Luis Verdegay Galdeano, David A. Pelta, Inma P. Cabrera, Bernadette Bouchon-Meunier, and Ronald R. Yager (Eds.), Vol. 853. Springer, 183–194. https://doi.org/10.1007/978-3-319-91473-2_16
- [9] Abdelkerim Rezgui, Jorge Marx Gómez, and Raji Ben Maaouia. 2017. KPI-Based Decision Evaluation System to Enhance QMSs for Higher Educational Institutes. *Int. J. Decis. Support Syst. Technol.* 9, 2 (2017), 39–55. <https://doi.org/10.4018/IJDSST.2017040103>
- [10] Reynaldo Joshua Salaki and Kalai Anand Ratnam. 2018. Agile Analytics: Applying in the Development of Data Warehouse for Business Intelligence System in Higher Education. In *Trends and Advances in Information Systems and Technologies - Volume 1 [WorldCIST'18, Naples, Italy, March 27-29, 2018] (Advances in Intelligent Systems and Computing)*, Álvaro Rocha, Hojjat Adeli, Luis Paulo Reis, and Sandra Costanzo (Eds.), Vol. 745. Springer, 1038–1048. https://doi.org/10.1007/978-3-319-77703-0_101
- [11] R. Sharda, D. Delen, and E. Turban. 2018. *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*. Pearson.
- [12] Pepijn Van Diepen and Bert Bredeweg. 2016. Performance indicators for online secondary education: a case study. In *Benelux Conference on Artificial Intelligence*. Springer, 169–177.
- [13] T PanduRanga Vital, BG Lakshmi, H Swapna Rekha, and M DhanaLakshmi. 2019. Student Performance Analysis with Using Statistical and Cluster Studies. In *Soft Computing in Data Analytics*. Springer, 743–757.
- [14] Renske Weeda, Cruz Izu, Maria Kallia, and Erik Barendsen. 2020. Towards an Assessment Rubric for EiPE Tasks in Secondary Education: Identifying Quality Indicators and Descriptors. In *Koli Calling '20: 20th Koli Calling International Conference on Computing Education Research, Koli, Finland, November 19-22, 2020*, Nick Falkner and Otto Seppälä (Eds.). ACM, 30:1–30:10. <https://doi.org/10.1145/3428029.3428031>
- [15] Yaqiong Zhu. 2018. A Data Driven Educational Decision Support System. *International Journal of Emerging Technologies in Learning (iJET)* 13, 11 (2018), 4–16.