

UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO
EM INFORMÁTICA

Maria das Graças Cavalcante da Silva

**MODELOS DE PERSONALIZAÇÃO DE VISUALIZAÇÃO DE
DADOS PARA O PROFESSOR**

Maceió
2016

MARIA DAS GRAÇAS CAVALCANTE DA SILVA

**MODELOS DE PERSONALIZAÇÃO DE VISUALIZAÇÃO DE
DADOS PARA O PROFESSOR**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal de Alagoas, como requisito para parcial para obtenção do grau de Mestre em Informática.

Orientador: Prof. Dr. Alan Pedro da Silva
Coorientador: Prof. Dr. Ig Ibert Bittencourt
Santana Pinto

Maceió
2016

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central

Bibliotecária Responsável: Janaina Xisto de Barros Lima

S586m Silva, Maria das Graças Cavalcante da.
Modelos de personalização de visualização de dados para o professor /
Maria das Graças Cavalcante da Silva. – 2016.
91 f. : il.

Orientador: Alan Pedro da Silva.
Coorientador: Ig Ibert Bittencourt Santana Pinto.
Dissertação (Mestrado em Informática) – Universidade Federal de Alagoas.
Instituto de Computação. Programa de Pós-Graduação em Informática.
Maceió, 2016.

Bibliografia: f. 82-84.
Anexos: f. 85-91.

1. Modelos computacionais. 2. Ambiente virtual de aprendizagem.
3. Banco de dados – Visualização de dado. 4. Educação. I. Título.

CDU: 004.42:37



UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL
Programa de Pós-Graduação em Informática – PpgI
Instituto de Computação

Campus A. C. Simões BR 104-Norte Km 14 BL 12 Tabuleiro do Martins
Maceió/AL - Brasil CEP: 57.072-970 | Telefone: (082) 3214-1401



Membros da Comissão Julgadora da Dissertação de Mestrado de Maria das Graças Cavalcante da Silva, intitulada: “*Modelos de Personalização de Visualização de Dados para o Professor*”, apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal de Alagoas em 28 de junho de 2016, às 10h00min, no Miniauditório do Instituto de Computação da UFAL.

COMISSÃO JULGADORA

Prof. Dr. Alan Pedro da Silva
UFAL – Instituto de Computação
Orientador

Prof. Dr. Ig Ibert Bittencourt Santana Pinto
UFAL – Instituto de Computação
Orientador

Prof. Dr. Rafael de Amorim Silva
UFAL – Instituto de Computação
Examinador

Prof. Dr. Patrícia Leone Espinheira Ospina
UFAL – Universidade Federal de Pernambuco
Examinadora

*À minha família, em especial aos meus pais João e Benedita,
aos meus irmãos Girleide, Gilberto, Givaldo, Gilson, Girleno
e Giclênio e ao meu noivo Jonathas, por todo amor, apoio,
confiança e incentivo em todos os momentos.*

AGRADECIMENTOS

A Deus, por sua infinita bondade e misericórdia, que me fez caminhar mesmo quando não tinha mais disposição e força. Por ter colocado pessoas maravilhosas, que com suas experiências e atenção, ajudaram-me a prosseguir.

Aos meus orientadores, Alan Pedro e Ig Ibert, por todo o conhecimento que me passaram, por toda a evolução que me proporcionaram, pelas broncas e pelo caminho que sempre me mostravam. Ensinando-me que o mais importante de tudo é o “processo”. A vocês uma gratidão imensurável.

Aos meus pais, João e Benedita, uma mistura de severidade com uma doçura infinita, que apesar de todas as dificuldades que a vida lhes impôs, eles sempre buscaram e mediaram para que todos os seus filhos trilhassem um caminho honesto. Obrigada por chegarem ao fim desta etapa ao meu lado. Obrigada minha mãe, por não ter me deixado!

Aos meus irmãos, Girleide, Gilberto, Givaldo, Gilson, Girleno e Giclênio, os G's da minha vida, por terem sido o alicerce necessário ao lado de meus pais, enquanto não pude está ao lado deles. Obrigada por me ajudarem a cursar o mestrado. Obrigada a uns por terem “me criado” e a outros por me ensinarem a tentar sempre melhorar.

Ao meu noivo, Jonathas, que dentro do possível e impossível esteve ao meu lado, me apoiando, aguentando minhas reclamações, meus desesperos, meus choros e minhas ausências... Obrigada meu amor, essa conquista também é sua.

Aos meus pequenos, meus sobrinhos, Maria Rita, Ana Letícia, Lara Eduarda, Jhully Grabielly, João Pedro e João Lucas... Vocês são fundamentais em minha vida. Todas as vezes que nos encontrávamos, uns pediam colo, outros para que brincássemos, nisto vocês me fortificavam.

À minha FAMÍLIA (pais, irmãos, noivo, sobrinhos, e cunhados Clara, Joana, Laíse e Robson), vocês que são os pilares do meu caráter, que são a luz da minha vida, eu os amo incondicionalmente!

Aos meus amigos, Armando Barbosa, André Vinicius, Danila Feitosa, Ezequiel Batista, Josmário Albuquerque e Thiago Ávila, que nos meus momentos de angústia e medo, me apoiaram, me incentivaram.

Aos amigos Josilan Paulino, Berenice Pimentel, Marcus Braga e Olival Freitas, não tenho palavras para descrever a gratidão que sinto a vocês. Vocês foram pessoas maravilhosas, colocadas em minha vida, no momento mais preciso e oportuno. Que Deus os abençoe e que vocês continuem apoiando os sonhos de outras tantas “Marias”.

Aos membros do NEES, Meu Tutor, amigos e companheiros neste processo de aprendizado. Um reconhecimento especial a Ranilson Paiva, Patrícia Ospina, Jário José e

Daniel Borges, os que eu sempre estava perturbando, perguntando, chamando... E que ao mesmo tempo, tinham tanto a me ensinar.

Espero ter conseguido agradecer a todos(as) que foram fundamentais nesta caminhada, mas caso não tenha mencionado alguém, me perdoe. Contudo, tenha certeza de minha eterna gratidão!

A TODOS vocês, MUITO Obrigado!

“O que você deixa para trás não é o que é gravado em monumentos de pedra, mas o que é tecido nas vidas de outros”.

Péricles

RESUMO

A presente dissertação trata da criação de modelos de personalização de Visualizações de dados para os professores, no intuito de viabilizar auxílio na tomada de decisão pedagógica, em que, fundamentado neste auxílio, os professores poderão analisar corpus com base em visualizações apropriadas, a partir de características pessoais. Para tanto, realizamos pesquisas na área de visualização de dados, em que optamos por dividir as técnicas de visualização de dados em dois grandes grupos, técnicas tradicionais e técnicas não tradicionais. Durante a pesquisa, necessitamos adaptar um processo de recomendação pedagógica, de forma que o mesmo pudesse dá suporte exclusivamente ao professor por meio de visualização de dados, a partir disto, foi elaborado e aplicado um survey, com trezentas e trinta e três participantes, dos quais duzentas e trinta e cinco eram professores. Realizaram-se diversas filtragens de dados e análises estatísticas baseadas em modelos de regressão beta. Nestas análises foram identificadas variáveis pertinentes às características dos professores, tais que fizeram com eles escolhessem entre visualizações tradicionais e não tradicionais. Depois desta fase, realizou-se uma modelagem de classificação, do tipo árvore de decisão, a fim de descobrirmos quais os conjuntos de características podem estar relacionados à preferência do professor por uma determinada visualização. Com isso criamos quatro classes, a classe (i) preferência, (ii) a classe percepção, (iii) a classe tempo e (iv) a classe ajuda. As quais modelam as características relacionadas ao professor, quando ele busca analisar um dado em uma visualização de acordo com suas preferências, uma visualização que dentro de suas características o faça perceber melhor os dados, uma visualização que o ajude a ver as informações e uma visualização que reduza o seu tempo de decisão. Com os resultados, fomos capazes de identificar quais características pertinentes aos professores tem influência na preferência de uma visualização e quais conjuntos de características levam a possíveis escolhas de visualizações possibilitando-nos, assim, criarmos modelos de personalização de visualizações de dados, como forma de melhorar a tomada de decisão pedagógica.

Palavras-chaves: Visualização de dados. Características. Modelos. Personalização.

ABSTRACT

This thesis studies the creation of personalization models of data visualization for instructors in order to assist pedagogical decision making. Thus, instructors will be able to analyze data based on appropriate data visualizations from personal characteristics. For this purpose, we researched on the literature for data visualization and decided to divide its techniques into two major groups, traditional techniques, and non-traditional techniques. During the research, we adapted a pedagogical recommendation process to provide exclusive support to instructors through the use of data visualization. We also designed and placed a survey involving 333 participants, of which 235 are instructors. After that, we used beta regression models to perform statistical analysis and data filtering. The study identified variables related to instructors characteristics that led them to choose between traditional visualizations and non-traditional ones. After that phase, we performed a decision tree classification modeling to identify a set of characteristics related to instructors preferences in certain visualizations. Therefore, we created four classes: (i) preference, (ii) perception, (iii) time, and (iv) help. These classes mold specific characteristics related to the instructor when he is analyzing data according to his preferences, a visualization that makes him better understand the data, other to help him perceive the data, and a visualization to reduce his decision-making time. The results allowed identifying which characteristics relevant to the instructors have some influence on a visualization preference. Further, which characteristic set led us to identify certain visualization alternatives that enable us to create personalization models of data visualization to improve pedagogical decision making.

Key-words: Data Visualization. Characteristics. Models. Personalization.

LISTA DE ILUSTRAÇÕES

Figura 1 – Atividades do professor.	17
Figura 2 – Situação do professor.	17
Figura 3 – Situação dos alunos ao final do ano letivo.	18
Figura 4 – Mapa figurativo onde são expostas as sucessivas perdas de homens do Exército francês na campanha da Rússia 1812-1813.	23
Figura 5 – Gráfico de barras ou gráfico de colunas	25
Figura 6 – Gráfico de linha	25
Figura 7 – Hierarquia de bolhas	26
Figura 8 – Square pie ou Gráfico de unidade	27
Figura 9 – (A) Gráfico de Barras e (B) Gráfico de Caixas Coloridas	29
Figura 10 – Criação do sistema, sob a concepção do usuário.	31
Figura 11 – Diagrama de caso de uso do sistema.	31
Figura 12 – Processo de Recomendação Pedagógica	35
Figura 13 – Processo de Recomendação de Visualização de Dados	37
Figura 14 – Estágios da criação do modelo	39
Figura 15 – Diagrama de Atividades de distribuição de blocos de acordo com o processo	44
Figura 16 – Diagramas de dispersão da variável resposta para cada tipo de visualização.	59
Figura 17 – Gráficos de resíduos. Modelo 2 estimado.	62
Figura 18 – Gráficos de resíduos. Modelo 3 estimado.	64
Figura 19 – Gráficos de resíduos. Modelo 4 estimado.	65
Figura 20 – Gráficos de resíduos.	68
Figura 21 – Arvore de Decisão - Classe Preferência	75
Figura 22 – Arvore de Decisão - Classe Percepção	76
Figura 23 – Arvore de Decisão - Classe Ajuda	77
Figura 24 – Arvore de Decisão - Classe Tempo	78
Figura 25 – Características dos professores que escolheram os três diferentes tipos de visualização de dados..	80

LISTA DE TABELAS

Tabela 1	– Comparando trabalhos relacionados	32
Tabela 2	– Definição das hipóteses	46
Tabela 3	– Nível de Fatores	46
Tabela 4	– Instanciando Variáveis - Características Pessoais	50
Tabela 5	– Instanciando Variáveis - Questões e Respostas	51
Tabela 6	– Modelos de regressão beta. Modelo 1: comparação entre os efeitos dos três tipos de visualizações. Base de comparação: Visualização tipo II. Modelo 2: Avaliação do efeito isolado da visualização do tipo I.	61
Tabela 7	– Modelos de regressão beta. Modelo 3: Avaliação do efeito isolado da visualização do tipo II. Modelo 4: comparação entre os efeitos dos três tipos de visualizações. Base de comparação: Visualização do tipo I.	63
Tabela 8	– Modelo de regressão beta. Modelos 5, 6 e 7: Avaliação das qualificações atribuídas aos três tipos de visualizações de acordo com o perfil do entrevistado.	66
Tabela 9	– Parâmetros para a criação da classificação	70

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Motivação e Contextualização do Trabalho	14
1.2	Problemática	15
1.3	Cenário Ilustrativo	16
1.4	Objetivos	19
1.5	Escopo do Trabalho	20
1.6	Organização da Dissertação	20
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	Visualização	21
2.2	Visualização de Dados	22
2.2.1	Técnicas de Visualização de Dados	24
2.2.1.1	Gráfico de Barras ou Gráfico de Colunas (tradicional)	24
2.2.1.2	Gráfico de linha (tradicional)	25
2.2.1.3	Hierarquia de bolhas (não tradicional)	26
2.2.1.4	Square pie ou Gráfico de unidade (não tradicional)	26
2.3	Personalização	27
3	TRABALHOS RELACIONADOS	28
3.1	Exploring the Role of Individual Differences in Information Visualization	28
3.2	Individual User Characteristics and Information Visualization: Connecting the Dots through Eye Tracking	29
3.3	Recommender Narrative Visualization	30
3.4	Comparativo dos trabalhos relacionados	31
4	PROPOSTA	34
4.1	Formalização da proposta	35
4.1.1	Adaptação do Processo	37
4.1.2	Passos para a criação do modelo	38
4.1.2.0.1	Estágio I	39
4.1.2.0.2	Estágio II	40
4.1.2.0.3	Estágio III	40
4.1.2.0.4	Estágio IV	40
4.1.2.0.5	Estágio V	40
4.1.2.0.6	Estágio VI	41
4.1.2.0.7	Estágio VII	41

4.1.2.0.8	<i>Estágio VIII</i>	41
4.2	Modelo de aplicação do processo Recomendação de Visualização de Dados baseado em survey	41
4.2.1	<i>Distribuição dos blocos no processo de Recomendação de Visualização de Dados</i>	42
5	MÉTODO DE PESQUISA	45
5.1	Planejamento do Survey	45
5.1.1	<i>Questões de Pesquisa e Hipóteses</i>	45
5.1.2	<i>Seleção de Variáveis</i>	46
5.1.3	<i>Escolha das Visualizações de Dados</i>	47
5.1.4	<i>Preparação dos dados para a criação do Survey</i>	47
5.1.5	<i>Elaboração do Survey</i>	48
5.1.6	<i>Aplicação do Survey</i>	49
5.1.7	<i>Preparação para a análise dos dados</i>	49
5.1.8	<i>Instrumentação para as análises dos dados</i>	52
5.1.9	<i>Ameaças a Validade</i>	52
6	ANÁLISES E RESULTADOS DO SURVEY	53
6.1	Modelos de Regressão	53
6.2	Modelo de regressão beta	55
6.3	Análise de Diagnóstico	56
6.4	Análise estatística descritiva	57
6.4.1	<i>Descrição das variáveis - Resposta Global</i>	57
6.4.2	<i>Descrição das variáveis - Respostas Parciais</i>	58
6.4.3	<i>Análise descritiva</i>	58
6.5	Modelos: Scores parciais de pontuação considerando separadamente cada tipo de Visualização	59
6.5.1	<i>Perfil do avaliador e escores de pontuação dos três tipos de visualizações</i>	65
6.6	Modelo de Classificação	69
6.6.1	<i>Classe Preferência</i>	71
6.6.2	<i>Classe Percepção</i>	72
6.6.3	<i>Classe Ajuda</i>	73
6.6.4	<i>Classe Tempo</i>	74
7	DISCUSSÕES	79
7.1	Respondendo as Hipóteses	80
8	CONCLUSÕES	82
8.1	Contribuições do Trabalho	82
8.2	Perspectivas Futuras	82

Referências **84**

ANEXOS **87**

1 INTRODUÇÃO

O presente trabalho está situado na área da Informática na Educação, dando sua contribuição a partir do conceito de Visualização de Dados. Nesse sentido, os esforços da comunidade de Visualização de Dados não têm sido direcionados, de forma bem específica, para a tomada de decisão pedagógica, apesar de todas as potencialidades que o mesmo oferece para tomada de decisão de maneira geral. Assim, este trabalho propõe Modelos de Personalização de Visualização de Dados para o Professor, como forma de aproximar a Visualização de Dados aos problemas pedagógicos enfrentados por estes.

Além disso, o trabalho visa justificar e descrever a proposição e o planejamento de Modelos de Personalização de Visualização de Dados para Professores. Esta modelagem se destina a prover auxílios aos professores na análise de visualizações de dados de situações pedagógicas, possibilitando um direcionamento de visualização a eles, mediante as suas características pessoais.

Nas próximas seções, serão apresentados os fatos que motivaram esta pesquisa, e o contexto no qual a mesma está inserida. Seguidamente serão descritos os problemas identificados por pesquisadores da área, o problema geral e o problema específico da pesquisa. Ainda neste capítulo, o escopo do trabalho será definido, enumerados os objetivos que se deseja alcançar, por fim, será descrito como o restante da proposta está organizado.

1.1 MOTIVAÇÃO E CONTEXTUALIZAÇÃO DO TRABALHO

A Declaração Universal dos Direitos Humanos ¹ assegura que todo indivíduo tem direito a educação (BRASIL, 1948). Isto é um direito universal, onde todos, sem exceção, devem desfrutar disto. (RIZZI; GONZALEZ; XIMENES, 2009) complementam ainda que tratar a educação como um direito humano, significa que não deve depender de condições econômicas ou sociais dos estudantes, nem mesmo está submetida às regras de mercado. Em outros termos, a educação não pode estar limitada à condição social, nacional, cultural, de gênero ou étnico-racial da pessoa.

De acordo com o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP ², no Brasil, a questão do acesso à escola não é mais um problema, dado que quase a totalidade das crianças está inserida no sistema educacional. Entretanto existem problemas educacionais críticos ocorrendo tais como: (i) Altas taxas de repetências. (ii) Grande quantidade de alunos que abandonam a escola antes da conclusão da educação básica. (iii) Baixa proficiência em testes padronizados (INEP, 2011a).

¹ Disponível em: <http://unesdoc.unesco.org/images/0013/001394/139423por.pdf>

² Endereço de acesso: www.inep.gov.br/

No intuito de analisar a situação das escolas brasileiras e os pontos críticos citados anteriormente, foi desenvolvido pelo INEP o Índice de Desenvolvimento da Educação Básica - IDEB, o qual sintetiza informações de desempenho em exames padronizados com informações sobre rendimento escolar (taxa média de aprovação dos estudantes na etapa de ensino)(INEP, 2011b).

A lei das diretrizes e bases da educação nacional - LDB, estabelecida na década de 90, em seu artigo 13 ³, descreve o papel do professor dentro das escolas brasileiras com os seguintes: (i) Participação na elaboração da proposta pedagógica em sua instituição. (ii) Elaboração e cumprimento do plano de trabalho, segundo a proposta pedagógica da sua instituição. (iii) **Zelar pela aprendizagem dos alunos.** (iv) **Estabelecer estratégias de recuperação para os alunos de menor rendimento**⁴. (v) Ministrando os dias letivos e horas-aula estabelecidas. (vi) Participação integral nos períodos dedicados ao planejamento, à avaliação e ao desenvolvimento profissional. (vii) Colaborar com as atividades de articulação da escola com as famílias e a comunidade (BRASIL, 1996).

Refletindo então sobre o que a LDB descreve como papel do professor, pode-se compreender que o seu papel, de forma resumida está associado a um gestor do conhecimento e aprendizagem dos estudantes. Portanto, o mesmo pode ainda ser visto como um dos principais responsáveis pela superação de possíveis problemas educacionais.

1.2 PROBLEMÁTICA

De acordo com (CURY, 2003), os professores, apesar das suas dificuldades, são insubstituíveis porque a gentileza, a solidariedade, a tolerância, a inclusão, os sentimentos altruístas, enfim, todas as áreas da sensibilidade não podem ser ensinadas por máquinas e sim por seres humanos. Assim, é possível perceber que o professor, muitas vezes precisa responder a exigências que estão além de sua formação, desempenhando funções de agente público, assistente social, enfermeiro, psicólogo, entre outras (OLIVEIRA, 2004).

De acordo com (MADDEN, 2014) cabe aos professores não apenas resgatar o conhecimento que obtiveram durante sua preparação acadêmica e profissional para transmitir a seus alunos, assim como, é preciso também saber escolher qual estratégia pedagógica é a mais apropriada para acolher as necessidades particulares de seus alunos.

Neste contexto, é inegável que os professores são tomadores de decisões, e com tal, eles precisam diariamente refletir sobre ações que viabilizem a aprendizagem de seus alunos, que os façam aprenderem, e que este "aprender" seja uma experiência impactante (SCHILDKAMP; LAI; EARL, 2012).

(PAIVA; BITTENCOURT; SILVA, 2013) entendem que a tomada de decisões pedagó-

³ Disponível em: <http://www2.senado.leg.br/bdsf/bitstream/handle/id/70320/65.pdf>

⁴ Grifo colocado pela autora.

gicas eficazes, exigem um bom suporte computacional, o qual por sua vez irá promover um melhor suporte informacional. Este suporte informacional depende diretamente do levantamento das informações relevantes, as quais só podem ser detectadas e obtidas a partir de dados educacionais (GARCÍA et al., 2011). Assim, um grande volume de dados deverá ser analisado, para se chegar a informações relevantes, para que a partir destas informações sejam estudadas estratégias que atuem diretamente sobre problemas preocupantes na educação.

Consoante (ANNIE, 2014) a tomada de decisão é um processo bastante complexo, uma vez que exige do professor a compreensão das necessidades a serem analisadas. Nesse sentido, a visualização de dados tem se mostrado como um dos instrumentos adequados para proporcionar o suporte em tomadas de decisão complexas a partir de grandes volumes de dados, com bem enfatiza (BIENKOWSKI; FENG; MEANS, 2012).

O que a literatura vem indicando é o uso de mineração de dados para subsidiar conhecimento, de tal forma que o professor possa ser auxiliado em tomadas de decisões como estas (PAIVA et al., 2012). No entanto, isto não é adequado para o uso em sala de aula, porque exige, dos professores, um conhecimento específico da área de Inteligência Artificial. Assim, para aproximar a mineração de dados e o seu conhecimento gerado em tomadas de decisão pedagógica, foi desenvolvido um processo que facilita a sua utilização, proporcionando melhorias na geração de informação personalizada nesse contexto, conforme (PAIVA; BITTENCOURT; SILVA, 2013) (PAIVA et al., 2014). Portanto, este trabalho propõe a utilização do respectivo processo para tentar viabilizar o uso de visualização de dados na tomada de decisão pedagógica do professor.

1.3 CENÁRIO ILUSTRATIVO

Buscando uma melhor compreensão da proposta e uma melhor assimilação dos desafios enfrentados pelos professores no processo de tomadas de decisões pedagógicas, a pesquisa apresentou a necessidade de criar esta seção, com um cenário ilustrativo abaixo, que ilustra os desafios os quais muitos professores atualmente estão enfrentando nas escolas brasileiras.

Um dado professor, em suas atividades cotidianamente, precisa elaborar suas aulas, organizar e corrigir provas, preencher diários escolares e por fim, detectar alunos que estão em risco de reprovação e evasão. Nesse sentido, para que exista a detecção destes alunos, o professor precisa analisar uma grande quantidade de dados para que decisões sejam tomadas de forma assertiva e ágil, no intuito de que seus alunos cheguem a concluir o ano letivo, o curso ou a disciplina (conforme a figura 1).

Portanto, consideremos o seguinte cenário:

- Um professor leciona matemática;

Figura 1 – Atividades do professor.

- ✓ O Professor precisa elaborar suas aulas...
- ✓ Elaborar e corrigir provas...
- ✓ Preencher cadernetas escolares...
- ✓ Detectar alunos que estão em risco de reprovação e evasão.

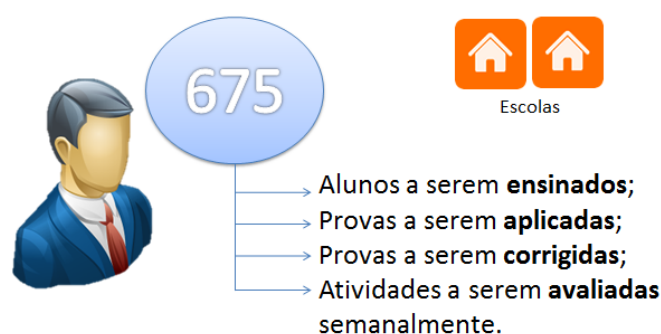


Fonte: Elaborado pela autora.

- Trabalha em duas escolas (em cidades diferentes);
- Tem quinze turmas (com séries diferentes);
- E cada uma de suas turmas tem em média quarenta e cinco alunos.

Além disso, este professor precisa se deslocar diariamente da cidade “A” para a cidade “B”, percorrendo em média cerca de quarenta quilômetros. Ele tem seiscentos e setenta e cinco alunos, este número representa a quantidade de alunos a serem ensinados e a quantidade de provas a serem aplicadas e avaliadas, em diferentes níveis de aprendizado (figura 2).

Figura 2 – Situação do professor.



Fonte: Elaborado pela autora.

Portanto, dado que este professor precisar analisar o desempenho de cada turma das escolas que ele leciona. Será que ele consegue em uma simples análise distinguir rapidamente, qual turma tem melhor desempenho em cada um dos conteúdos específicos? Ou então, identificar quantos alunos desta mesma turma aprenderam cada um dos conteúdos específicos? Ou de forma ainda mais aprofundada, será que ele consegue detectar quais alunos irão conseguir a aprovação na disciplina ao final do ano letivo?

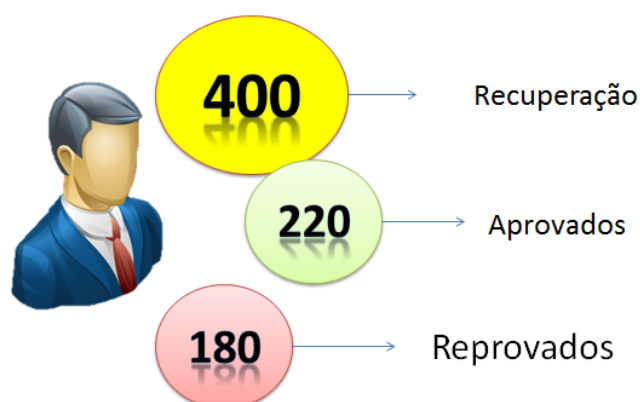
No entanto, para responder as questões acima, faz-se necessário que o professor consiga recuperar informações apropriadas, para depois agrupá-las e analisá-las de tal forma que seja obtido respostas relevantes. Embora, essa tarefa pode requerer do professor um tempo que ela não tenha, e além disso, pode exigir uma habilidade computacional e estatística que ele também não tenha.

Desta maneira, o professor conta com três grandes desafios: (i) O tempo e (ii) a forma como as informações necessárias sobre cada turma serão obtidas. E o (iii) método como estas informações serão analisadas e visualizadas. Assim, além de todo seu planejamento diário, é preciso organizar um tempo para fazer a análise dos resultados, de tal forma que se possa ser suficientemente, é preciso na identificação de quais de suas ações em sala de aula estão dando certo em cada turma e quais devem e podem ser melhoradas.

No que se refere à tomada de decisão dos professores, (ANNIE, 2014) acredita-se que alguns critérios estão envolvidos, em que estes influenciam na aprendizagem do aluno, tais como (i) propósito de dar consideração a uma questão, (ii) identificação do resultado final desejado, (iii) determinação das opções para chegar ao resultado final, e (iv) seleção da opção mais adequada para atingir o objetivo desejado.

Partindo-se do pressuposto que tal desafio dificilmente será superado pelo suposto professor. Ou seja, supondo que não foi viável recuperar a informação adequada e também fazer o devido tratamento a ela. Vários alunos não seriam devidamente acompanhados e acolhidos no momento correto, podendo ter consequências, tais como: (i) Ser reprovado. (ii) Desistir de estudar (figura 3). (iii) E até mesmo seguir outros caminhos mais destrutivos. Acredita-se que este caso pode ser bem representativo no dia a dia das escolas dado o estudo "Educação Brasileira: Indicadores e Desafios" (BRASIL, 2013).

Figura 3 – Situação dos alunos ao final do ano letivo.



Fonte: Elaborado pela autora.

Após análise e reflexão do cenário descrito acima, é evidente a necessidade de proporcionar suporte tecnológico a esses professores de forma que se possa auxiliá-los

na tomados de decisão pedagógica. Propiciando que decisões possam ser tomadas de modo assertivo⁵ e ágil⁶, onde o professor possa visualizar diferentes alternativas para um dado problema.

Considerando o que foi mencionado, identificamos o nosso **problema geral**: Como podemos auxiliar os professores na tomada de decisão pedagógica para ajudar suas turmas? Para tal questionamento, precisam-se minimizar os pontos críticos citados anteriormente, a área de visualização de dados é vista, no escopo desta pesquisa, como uma alternativa que pode oferecer o suporte necessário ao processo de tomada de decisão dos professores.

Entretanto, será que uma visualização de dados, é interpretada da mesma forma por todos os professores? Será que todos os professores preferem ver as mesmas visualizações de dados? Com base em nosso problema geral, e nos questionamentos citados, chegamos ao nosso **problema específico**: Como podemos proporcionar visualização de dados educacionais de forma personalizada para auxiliar professores nas tomadas de decisões pedagógicas?

Para guiar nossa pesquisa, dentro dos problemas definidos acima, sentimos a necessidade de criar as seguintes questões:

- Será que o perfil do professor influencia na escolha de uma visualização específica a ser vista?
- Que características dos professores podem estar presente na preferência de uma determinada visualização?
- Será possível fazer uma previsão computacional de preferência de visualização de dados a partir de características?

1.4 OBJETIVOS

Portanto, o objetivo geral desta dissertação é estudar como viabilizar Visualização de Dados de forma adequada aos professores, considerando suas particularidades, afim de que o estímulo, em termos de informação, produzido por uma visualização específica possa ser disponibilizada de forma otimizada para um professor com um perfil específico. Assim, para atender a tal propósito, foram estabelecidos os objetivos específicos abaixo, afim de que se possam produzir os primeiros resultados sobre estas questões:

- Estudar técnicas de visualização de dados;
- Verificar a viabilidade de se utilizar visualização de dados para o professor;

⁵ Entendemos "modo assertivo" como ação coerente com o problema evidenciado.

⁶ Por "ágil" nos referimos ao período no qual a decisão precisa ser tomada para que resultados de impacto positivo possam ser aplicados a uma dada situação.

- Verificar as preferências do professor para diferentes técnicas de visualização de dados;
- Extrair características que, eventualmente, influenciam na percepção do professor em diferentes métodos de visualização;
- Tentar propor modelos que possam orientar a associação de métodos específicos de visualização a perfis específicos de professor.

1.5 ESCOPO DO TRABALHO

O escopo desta pesquisa visa a alcançar os objetivos definidos na seção anterior (Seção 1.3). Entretanto, afirmamos está fora do escopo desta observação abordar as seguintes questões:

- Analisar a eficiência de técnicas de visualização de dados;
- Analisar outro público além do público professor;
- Permitir que professores processem ou analisem, livremente, os dados educacionais.

1.6 ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação está organizada da seguinte forma: no **Capítulo 2** apresentaremos a fundamentação teórica, que aborda o conhecimento utilizado nessa dissertação. **O Capítulo 3**, os trabalhos pertinentes ao tema. **O Capítulo 4** abordará e descreverá a proposta, que tratará da criação de Modelos de Personalização de Visualização de dados para o Professor. No **Capítulo 5**, exibiremos o método de pesquisa qualitativa utilizado (*Survey*).

No **Capítulo 6** abordaremos as análises e resultados do *Survey*. O **Capítulo 7** apresentaremos uma breve discussão sobre os resultados obtidos. Finalmente, o **Capítulo 8** apresentaremos as conclusões, seguido das referências e anexo deste trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Nesse capítulo abordaremos os principais tópicos acerca dos fundamentos teóricos utilizados nessa dissertação e necessários para sua apropriação e compreensão. Nesse sentido, distribuímos as seções da forma como segue: na **seção 2.1**, será apresentada uma introdução à visualização, seguida de visualização de dados, introduzida na **seção 2.2**. A **seção 2.2.1** expõe as técnicas de visualização de dados estudados e utilizados nas pesquisas realizadas para a escrita da dissertação. Logo após, abordaremos os conceitos acerca da personalização, na **seção 2.3**.

2.1 VISUALIZAÇÃO

Ao relatar que as imagens têm sido usadas como um mecanismo de comunicação bem antes da formalização da linguagem escrita, (WARD; GRINSTEIN; KEIM, 2010) definem a visualização como a comunicação de informações, utilizando representações gráficas. Eles apontam que em uma única imagem pode conter uma grande quantidade de informação, e estas informações podem ser processadas muito mais rapidamente se comparadas à leitura de uma página contendo unicamente palavras.

(DEFANTI; BROWN, 1991) definem o campo da visualização como um método de computação que oferece enxergar o que anteriormente era considerado incompreensível por meio apenas de dados brutos. Este método permite que seja enriquecido o processo de descoberta científica, promovendo e possibilitando reflexões profundas e inesperadas.

Entende-se que o objetivo da visualização é alavancar métodos científicos existentes, fornecendo uma nova visão científica através de métodos visuais. As visualizações são partes cada vez mais importantes dos sistemas cognitivos. Estima-se que 50% dos neurónios do cérebro estão associados à visão, assim, visualizações nos possibilitam adquirir mais informações através da visão do que por meio de todos os outros sentidos combinados (DEFANTI; BROWN, 1991), (WARE, 2012).

De acordo com (KIRK, 2012), a visualização deve ser tratada como uma ferramenta de descobertas. Uma vez que é por meio dela, que se busca retratar dados de forma que seja possível vê-los sob uma nova luz (uma nova perspectiva), sendo possível observar visualmente padrões, exceções e histórias que estão por trás de seu estado bruto.

2.2 VISUALIZAÇÃO DE DADOS

A visualização de dados é definida como uma representação visual de um conjunto de dados numéricos diferentes, que permite a tomada de decisão, a comparação entre dados e a identificação de possíveis tendências. É por meio da visualização de dados, que é possível a identificação, melhoria e reconhecimento de padrões, o que proporciona rapidez de forma perceptível na compreensão dos dados analisados com o mínimo de esforço (KRUM, 2013).

Segundo (ALEXANDER; YUK; DIAMOND, 2014) a visualização de dados pode ser definida como o estudo da forma de representar os dados através da utilização de uma abordagem artística e interativa, em detrimento às abordagens mais tradicionais, transformando dados em informação significativa e útil. (KRUM, 2013) descreve ainda que em 2001, um dos pioneiros de visualização de dados moderna, professor emérito de ciência política, estatística e ciência da computação pela Universidade de Yale, Dr. Edward R. Tufte explicou claramente o fenômeno de compreensão dos dados, quando declarou que a visualização de dados é:

"De todos os métodos de análise e de comunicação de informação estatística, bem concebidos gráficos de dados são usualmente os mais simples e ao mesmo tempo os mais poderosos".

As visualizações de dados beneficiam tanto seus usuários quanto seus projetores. (SIMON, 2013) cita alguns desses benefícios, dos quais destacamos dois deles. O primeiro é “eficiência de grandes visualizações”, elas propiciam que as pessoas olhem para grandes quantidades de dados de uma forma rápida. E o segundo, “são os efeitos visuais que elas oferecem”, possibilitando ajudar especialistas em visualizações ou grupos (sendo estes grupos de especialistas ou não) a alcançarem mais *insights*⁷ sobre a natureza de um determinado problema e descobrirem um novo.

Ainda sobre a eficiência das visualizações, é relevante destacar que uma boa visualização pode criar uma visão compartilhada de uma determinada situação, e ajuda a alinhar pessoas sobre ações necessárias a serem tomadas (SIMON, 2013). Elas viabilizam que pessoas de diferentes áreas enxerguem os mesmos dados sobre a mesma perspectiva. Assim, o objetivo da visualização de dados é de proporcionar a comunicação da informação, de modo claro e eficiente ao ponto de vista dos usuários, ajudando-os a analisar e a raciocinar sobre os dados e as evidências representadas por eles (FRIENDLY; DENIS, 2001).

Pode-se considerar que uma visualização adequada é uma espécie de narrativa, a qual proporciona uma resposta clara a uma pergunta sem detalhes irrelevantes (FRY, 2007). Assim, busca-se expor o essencial e indispensável a ser visto. Como exemplo, temos a figura 4, , tida como uma das principais e mais conhecidas visualizações de dados já criadas, ela foi

⁷ Pode ser entendido como a capacidade de compreensão/conhecimento a respeito de algo.

- Filtrar: reconhecido pela filtragem de todos os dados considerados interessantes.
- Minerar: determinado pela aplicação de métodos de estatística ou mineração de dados como uma maneira de discernir padrões ou colocar os dados no contexto matemático.
- Representar: determinado pela escolha uma representação visual pela qual os dados serão apresentados.
- Refinar: especificado pela melhoria na representação escolhida para torná-lo mais claro e visualmente atraente.
- Interagir: designado pela adição de métodos para manipular os dados ou controlar quais recursos serão visíveis.

2.2.1 Técnicas de Visualização de Dados

O dicionário *Oxford*⁹, define o termo técnica como "um método particular de fazer algo, especialmente quando se tem que aprender habilidades especiais". Ligando este termo a área de Visualização de dados é compreensível que as técnicas são usadas, com o objetivo de melhor expor os dados e absorção o conhecimento.

Neste trabalho, foram pesquisadas e classificadas as técnicas, em dois grandes grupos: As técnicas tradicionais e as técnicas não tradicionais. As técnicas tradicionais (tem como sinônimo o termo convencional), segundo *Oxford* referem-se a algo criado seguindo métodos e ideias mais antigas, em vez de os métodos mais modernos ou diferentes. Já as técnicas não tradicionais, de acordo com a definição acima, admitimos como técnicas mais modernas e diferentes, não tão usuais.

Com base nos estudos de (KHAN; KHAN, 2011) as técnicas convencionais¹⁰ de visualização de dados são as que em sua essência possuem características similares e possibilitam um entendimento comum.

A seguir, serão apresentadas as técnicas de visualização de dados utilizadas durante essa pesquisa. Ressaltamos que todas as técnicas foram descritas de acordo com os conceitos apresentados por (KIRK, 2012).

2.2.1.1 Gráfico de Barras ou Gráfico de Colunas (tradicional)

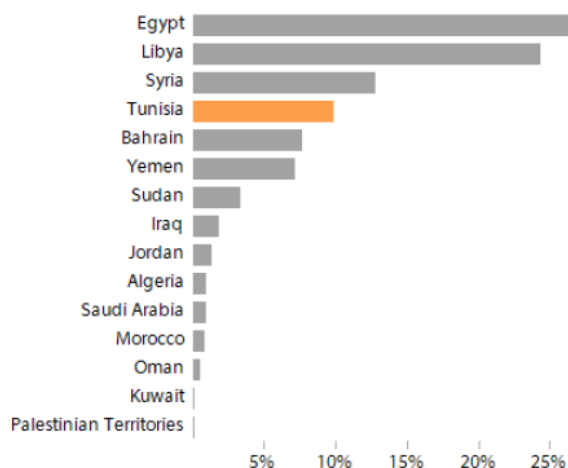
Os gráficos de barras, muitas vezes conhecidos também com gráficos de colunas, têm como principais características visuais, o comprimento (altura) e cor de matriz. São esses elementos que transmitem os dados, permitindo ao leitor realizar comparações precisas

⁹ Disponível em: <http://www.oxfordlearnersdictionaries.com/us/>

¹⁰ Termo definido pelo dicionário *Oxford* como "segundo o que é tradicional ou a maneira como algo foi feito por um longo tempo", o que pode ser entendido como algo comumente usado

entre as categorias presentes em sua estrutura gráfica. Estas categorias mostram os valores relativos e os absolutos que estão sendo representados. Ao usar o comprimento como a variável, estamos representando um valor quantitativo, sendo importante mostrar toda a sua extensão. Ao fazermos uso da cor, estamos buscando chamar a atenção para valores de categorias específicas mediante a narrativa de um determinado dado.

Figura 5 – Gráfico de barras ou gráfico de colunas

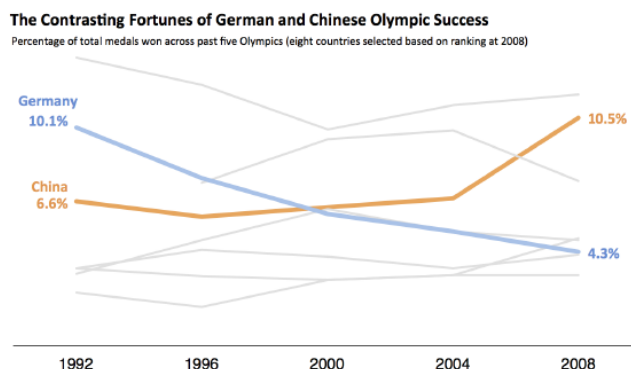


Fonte: (Kirk, 2012).

2.2.1.2 Gráfico de linha (tradicional)

Esta técnica é usada para comparar uma variável quantitativa contínua e o tamanho dos valores. As características visuais que fazem parte desta técnica englobam a posição, a inclinação e a cor de matriz. Os pontos verticais presentes nestes gráficos, são unidos usando linhas, que mostram a trajetória, alternando através das "pistas" resultantes. Este tipo de gráfico ajuda a contar histórias, as quais dão ênfase no padrão relativo da jornada dos dados.

Figura 6 – Gráfico de linha

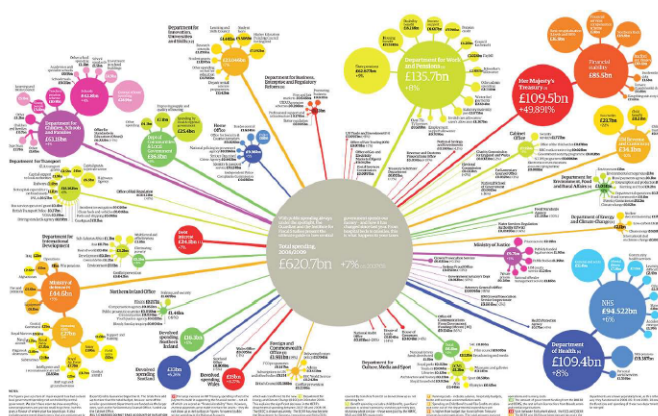


Fonte: (Kirk, 2012).

2.2.1.3 Hierarquia de bolhas (não tradicional)

A técnica de hierarquia de bolhas é usada muitas vezes para retratar uma organização e/ou uma estrutura através de uma exibição hierárquica. Destacam-se como principais variáveis na área visual, a posição das bolhas e cor de matriz. O uso de bolhas, também chamados de círculos, algumas vezes representam departamentos constituintes, dimensionado de acordo com o seu valor quantitativo, os quais são coloridos para distinguir visualmente os diferentes departamentos de dados.

Figura 7 – Hierarquia de bolhas

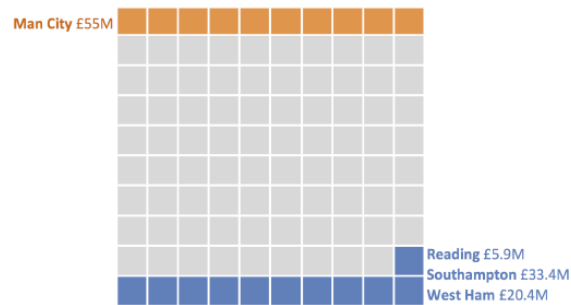


Fonte: (Kirk, 2012).

2.2.1.4 Square pie ou Gráfico de unidade (não tradicional)

O Square pie, possui diversas nomenclaturas, entretanto, essa técnica tem como características visuais, a posição, cor de matriz e símbolos. O desenvolvimento de um gráfico por meio dessa técnica, envolve uma grade de unidade, podendo ser quadrados ou símbolos, os quais buscam representar partes de um todo. O uso das cores e de símbolos enfatiza o estabelecimento uma composição visual de valores quantitativos e categóricos, do qual se pode ter ideia de percentual ou quantidade absoluta de algo.

Figura 8 – Square pie ou Gráfico de unidade



Fonte: (Kirk, 2012).

2.3 PERSONALIZAÇÃO

(MONTGOMERY; SMITH, 2009) no artigo "Prospects for Personalization on the Internet" apresentam algumas definições do termo personalização, das quais destacamos a definição dada por (ROBERTS; ZAHAY, 2012), que propõe a seguinte definição: A personalização é o processo de preparação de uma comunicação individualizada para uma pessoa específica com base em preferências explícitas ou implícitas.

De uma forma geral, a personalização muitas vezes é definida como a particularização de serviços de acordo com as características ou preferências de cada usuário (JOUNG; ZARKI; JAIN, 2009). Assim a personalização pode, por meio das definições anteriores, ser interpretada como uma forma de extrair as principais características de um usuário no intuito de fornecer a ele, algum serviço, sistema, ou até mesmo conteúdo, de forma que, o que está sendo oferecido seja pertinente e baseado em seu perfil.

Nesse texto, estar-se tratando como personalização, quando se busca levantar as características pessoais dos professores, para com base nelas dar-se andamento as pesquisas.

3 TRABALHOS RELACIONADOS

As seções apresentadas neste capítulo, trazem trabalhos relacionados à pesquisa desenvolvida nessa dissertação, e uma breve discussão das diferenças entre eles e o foco da presente pesquisa.

3.1 EXPLORING THE ROLE OF INDIVIDUAL DIFFERENCES IN INFORMATION VISUALIZATION

Estudo realizado por pesquisadores da *University of British Columbia*¹¹, descreve um estudo sobre usuários, tendo com objetivo a avaliação de duas técnicas diferentes de visualização de dados, desenvolvidas para descrever as mudanças ambientais complexas em um sistema interativo projetado para fomentar a sensibilização para o desenvolvimento sustentável descreve um estudo sobre usuários, tendo como objetivo a avaliação de duas técnicas diferentes de visualização de dados, desenvolvidas para descrever as mudanças ambientais complexas em um sistema interativo projetado para fomentar a sensibilização para o desenvolvimento sustentável (CONATI; MACLAREN, 2008).

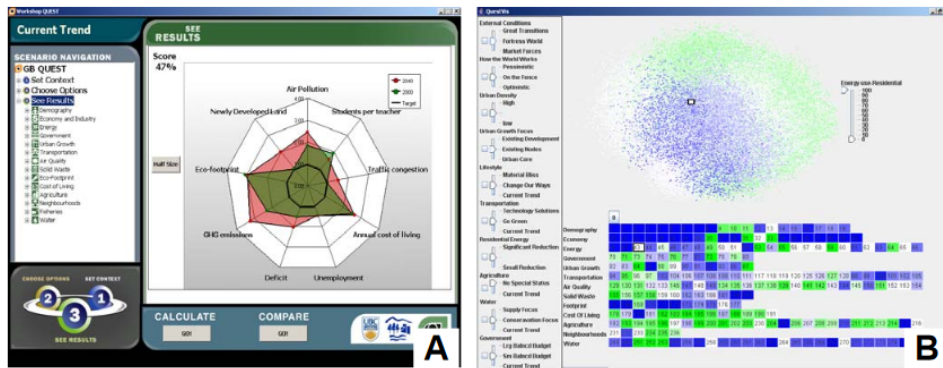
O presente estudo investiga se as diferenças individuais de usuários podem ser utilizadas como indicadores de eficácia de uma visualização, em escolher entre diferentes visualizações. As diferenças individuais frisadas no artigo referem-se às habilidades cognitivas, das quais são enumeradas dez habilidades, sendo elas: Memória visual, visualização espacial, velocidade perceptual, desincorporação, necessidade de cognição, estilo de aprendizagem.

As técnicas utilizadas foram as de gráfico de radar e a de caixas coloridas (Figura 9). O desempenho dos usuários foi comparado de acordo com cada visualização de forma isolada, a partir do sistema com o qual os utentes estavam inseridos. Como critério para desempenho do usuário, os autores optaram por avaliar habilidades cognitivas. Realizando modelos lineares gerais como forma de análise, em que se destacou que, dentre as habilidades utilizadas, aquela cuja significação mostrou-se relevante foi a rapidez de percepção.

O presente estudo traz considerações bastante relevantes para a área de visualização de dados, entretanto o seu foco não é educação, o artigo busca medir a eficácia de visualizações, não sendo este o foco da presente pesquisa. Toda a observação está sendo voltada para habilidades cognitivas, cujas características podem contribuir para um melhor desempenho do utilizador de visualização de dados.

¹¹ Site da Universidade: <https://www.ubc.ca/>

Figura 9 – (A) Gráfico de Barras e (B) Gráfico de Caixas Coloridas



Fonte: (CONATI; MACLAREN, 2008)

3.2 INDIVIDUAL USER CHARACTERISTICS AND INFORMATION VISUALIZATION: CONNECTING THE DOTS THROUGH EYE TRACKING

O estudo investiga a relação entre um conjunto de habilidades (medidas) cognitivas do usuário e experiência com tipo de visualização, dificuldade da tarefa, e os padrões de atenção do usuário ao utilizar duas diferentes técnicas de visualização (gráfico de barras e gráfico de radar). De acordo com (TOKER et al., 2013), o objetivo deste estudo foi o de investigar se características dos usuários podem impactar nos padrões de movimentos do olho durante o processamento de uma visualização, e identificar quais habilidades são influenciadas por quais características dos usuários e de que forma a influência é medida pela dificuldade da tarefa e o tipo de visualização.

Os autores analisaram o rastreamento do olhar para descobrir quais características podem ser influenciadas por atributos individuais do usuário. As características dos usuários investigadas para tanto, foram a experiência prévia com cada tipo de visualização e habilidades cognitivas, como a velocidade perceptual (uma medida da velocidade quando executasse tarefas de percepção), a memória de trabalho verbal (uma medida de armazenamento e manipulação capacidade de informação verbal) e memória de trabalho visual (uma medida de armazenamento e manipulação capacidade de informação visual e espacial).

Para a realização do estudo, foram recrutados estudantes da universidade, sendo dezoito do sexo feminino e dezessete do sexo masculino, com idades entre dezenove e trinta e cinco anos. Três testes diferentes foram realizados no intuito de medir a memória verbal, memória visual e a velocidade de percepção, sendo que dois foram por meio eletrônicos (computadores) e um por meio tradicional (papel). Para o rastreamento do movimento do olhar, foi utilizado o Tobii T120, o qual é incorporado a uma tela, proporcionando o rastreamento discreto. Como forma de interação dos participantes com as visualizações,

foram aplicadas perguntas textuais a eles, os quais buscavam responder mediante aos dados apresentados nas visualizações.

De acordo com a análise apresentadas no artigo, algumas das características cognitivas do usuário têm de fato uma influência significativa sobre o comportamento do olhar do usuário, das quais os autores destacam a velocidade perceptual. Os autores apontam ainda, que as capacidades cognitivas podem ser confiavelmente detectadas por técnicas de rastreamento do olhar, usando informações do olhar em tempo real, para a criação de visualizações adaptadas aos usuários.

3.3 RECOMMENDER NARRATIVE VISUALIZATION

O estudo foi desenvolvido pela *Carleton University*¹² em conjunto com a *IBM*^{13,14}, o qual propõe uma fundamentação de um framework para a criação de um sistema de recomendação de visualização narrativas personalizadas. Este sistema irá recomendar aos usuários uma melhor visualização com base nos dados extraídos de perfis de redes sociais.

Para a realização deste estudo, os pesquisadores se fundamentam no conceito de que o passo lógico na pesquisa de visualização é a apresentação (para esses teóricos diz respeito a forma como se conta a história de algo) ou pelo menos deveria ter um foco de igual importância com a exploração e análise de dados. Para os autores (KHATAEI; LAU, 2013), por meio de um sistema de recomendação de visualização será possível extrair informações, filtrar de acordo com o perfil, e transformar em visualizações narrativas personalizadas para o usuário, conforme a figura 10.

Para tanto, os autores descrevem que o sistema de recomendação de visualizações narrativas, precisará utilizar APIs¹⁵ para a extração das informações que estão contidas em redes sociais. Por fim, a recomendação da visualização é feita, possibilitando também ao visualizador que escolha do nível de detalhes de cada narrativa (figura 11).

Como resultado da análise do estudo, os autores constaram que um sistema de recomendação para contar história, pode influenciar na nova era da educação e na visualização, em que a visualização de dados pode ser compreendida como um método de superação de problemas na exploração de dados. Tais articuladores acreditam ainda, que introduzir a personalização pode gerar grandes oportunidades de criar visualizações individuais, não mais olhando para o todo, e sim, compreendendo o background, o nível de escolaridade e outras características de cada indivíduo. Entretanto, o estudo apresenta como limitação o fato de não ter um mecanismo próprio que possibilite a personalização, em que segundo os

¹² Site da Universidade: <http://carleton.ca/>

¹³ A sigla significa International Business Machines, onde corresponde uma empresa dos Estados Unidos voltada para a área de informática.

¹⁴ Site da empresa: <http://www.ibm.com>

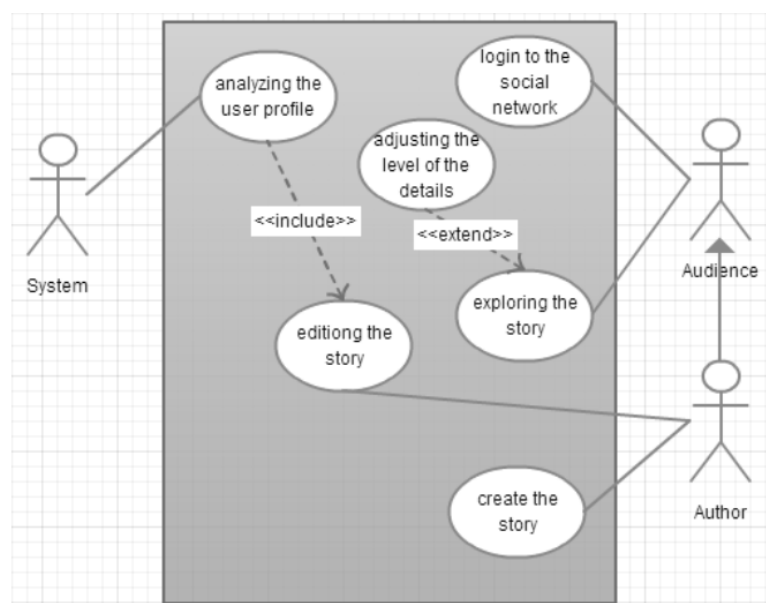
¹⁵ API é uma sigla em inglês que significa Application Programming Interface (Interface de Programação de Aplicativos) sendo instruções e padrões de programação para acesso a um aplicativo ou software.

Figura 10 – Criação do sistema, sob a concepção do usuário.



Fonte: (KHATAEI; LAU, 2013)

Figura 11 – Diagrama de caso de uso do sistema.



Fonte: (KHATAEI; LAU, 2013)

autores as características dos usuários serão coletadas por meio de site de relacionamentos, a exemplo das redes sociais.

3.4 COMPARATIVO DOS TRABALHOS RELACIONADOS

Os trabalhos descritos acima expõem três visões bastante próximas e ao mesmo tempo diferentes umas das outras. A presente pesquisa tem como foco o auxílio aos professores, os quais os leve a melhores tomadas de decisões, de forma que as características pessoais dos

docentes possam ser analisadas, viabilizando uma melhor estruturação para a proposta.

Como forma de sumarizar os principais pontos de observação dos trabalhos já apresentados, foi criada uma tabela, em que se aponta os pontos fortes e fracos de cada trabalho, a fim de que assim, seja possível uma melhor compreensão dos mesmos, como mostrado na tabela abaixo (tabela 1):

Tabela 1 – Comparando trabalhos relacionados

Título	Ponto forte	Ponto fraco
Exploring the role of Individual Differences in Information Visualization	Analisa diferenças de usuários ao interagirem com visualizações.	Mede a eficácia do tipo de visualização, tem foco em apontar que as diferenças podem estar ligadas a fatores cognitivos, e seu público não é o professor.
Individual User Characteristics and Information Visualization: Connecting the dots Through Eye Tracking	Investigam medidas cognitivas dos usuários e experiência do usuário com determinada visualização.	Consideram apenas fatores cognitivos, focam na medição da eficácia da visualização e seu público não é o professor.
Recommender Narrative Visualization	Propõe um sistema de personalização de recomendação de visualização com base em informações dos usuário, cogita a possibilidade de existirem diferentes características dos usuários e a necessidade de atender a todos.	Não tem mecanismo para a coleta de dados do usuário, dependendo de redes sociais para tanto, seu foco é passar algo ao professor, mas não levanta a necessidade de auxílio na tomada de decisão

Fonte: Elaborado pela autora.

Vimos que, dos três trabalhos relacionados, os dois primeiros apontam características individuais, de uma forma mais focada para a área da cognição, onde trazem pesquisas com base em fatores cognitivos. Entretanto tais trabalhos buscam medir a eficiência de determinadas visualizações. Nossa pesquisa difere destas pesquisas por levantar indícios que complementem estes trabalhos, buscando encontrar outras características, diferentes das estudados, que possam está associadas à características cognitivas individuais já levantadas. Entretanto tais fatores cognitivos não são o nosso alvo. Em nossa pesquisa, buscamos explorar características pessoais pertinentes aos professores, que quando agrupadas possam ser consideradas como um perfil de um determinado professor.

O último trabalho relacionado, por sua vez, apesar de propor um sistema de recomendação de visualização narrativa, e de questionar a criação de visualização de acordo com

características pessoais de professores, não foca na necessidade de exploração dos dados mas sim, de se passar uma história ao usuário. O presente trabalho teve como preocupação fazer com que a exploração dos dados faça com que os usuários reflitam sobre os corpora e como estão visualmente representados, para que assim, consigam não apenas entender os dados, mas também possam descobrir o que pode estar relacionado a eles.

A preocupação, não é apenas criar um sistema que recomende visualização, ou identificar qual visualização é mais eficiente. Dentro dessa pesquisa, tem-se como foco identificar quais características pessoais possivelmente levam os professores a preferirem determinadas visualizações a outras, após esta identificação, propôs-se uma classificação das características, no intuito de detectar qual visualização melhor se encaixa para um determinado conjunto de perfis pertinentes a diferentes professores.

4 PROPOSTA

De acordo com os trabalhos relacionados citados no capítulo anterior é possível compreender que atualmente a visualização de dados vem sendo usada não mais no intuito de apenas mostrar graficamente os dados, tornando-os mais interessantes e atraentes. A visualização de dados passou a ser vista como uma forma de interiorizar uma informação, a fim de fazer com que o visualizador ou utilizador, possa de fato compreender o que está sendo mostrado (LIU; STASKO, 2010).

Os autores (ZIEMKIEWICZ et al., 2012) relatam que, as visualizações estão sendo utilizados como auxiliares cognitivos no processo de resolução de problemas, os usuários passaram a contar com visualizações para ajudá-los a resolver os problemas cada vez mais difíceis. Entretanto, a maior parte das pesquisas sobre a área de visualização e seus usuários estão voltados em qualificar a eficácia de uma determinada visualização. Como exemplo disto, (CONATI; MACLAREN, 2008) ressaltam que a maior parte das pesquisas sobre visualizações se posiciona sobre os fatores que definem eficácia visualização estende-se sobre as propriedades dos dados a serem visualizados ou as tarefas a serem realizadas, em que alguns estudos começaram a considerar as diferenças individuais do usuário como um caminho promissor de investigação.

Existe na literatura a evidência anedótica que diferentes usuários podem ter diferentes preferências. Uma vez que, dada uma tarefa e dados correspondentes, visualizações diferentes podem funcionar melhor para usuários diferentes, dadas as características do usuário, como habilidades cognitivas, conhecimentos e preferências pessoais (CONATI; MACLAREN, 2008).

Desta forma é compreensível que as características pessoais e preferências podem divergir de um usuário para outro, uma vez que nem toda visualização é a mais apropriada a todo visualizador. Seguindo esta premissa, a proposta desta dissertação é a criação de Modelos de Personalização de Visualização de Dados para o professor, que disponibilize ao docente visualizações de dados personalizadas, obtidas de acordo com as características individuais e pessoais pertinentes ao seu perfil.

A proposta aqui apresentada levanta que outras características além das investigadas atualmente podem possibilitar ao usuário uma melhor compreensão dos dados e das informações a serem mostradas, não sendo necessário projetar novas visualizações, mas sim, direcionar as que melhor se adaptam as características presentes nos perfis de cada usuário. O foco, portanto, é proporcionar aos professores a possibilidade de que eles visualizem as informações de suas turmas, com base em suas próprias características.

Nas seções deste capítulo, serão explicadas com mais detalhes o que se propõe, bem como as etapas necessárias para se alcançar o objetivo perseguido.

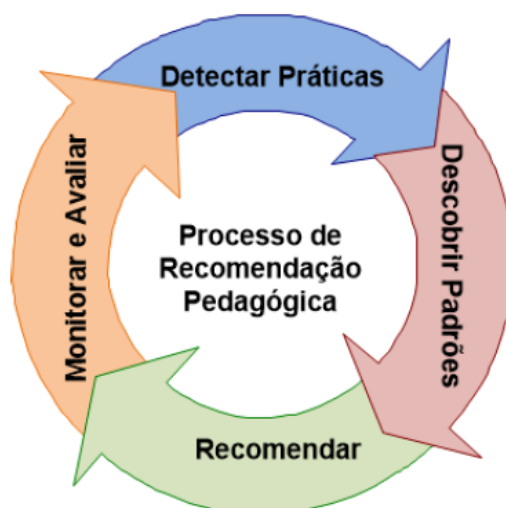
4.1 FORMALIZAÇÃO DA PROPOSTA

Uma das principais metas é identificar as características predominantes do perfil de cada professor para que assim, seja possível a criação do modelo de classificação de preferências. Mediante a isto, seguimos como método uma adaptação do Processo de Recomendação Pedagógica, criado por (PAIVA; BITTENCOURT; SILVA, 2013) (PAIVA et al., 2014).

O Processo de Recomendação, mostrado na figura 12, é de natureza interativa e incremental, o qual é composto por quatro etapas, como descritas a baixo:

- Primeira etapa: detectar práticas no processo de aprendizagem;
- Segunda etapa: descobrir padrões responsáveis pelas práticas detectadas;
- Terceira etapa: recomendar conteúdo pedagógico com base nos padrões descobertos;
- Quarta etapa: monitorar e avaliar a efetividade/eficiência das recomendações.

Figura 12 – Processo de Recomendação Pedagógica



Fonte: (PAIVA; BITTENCOURT; SILVA, 2013).

A etapa de **Detecção de Práticas** no processo de aprendizagem corresponde à identificação de interações dos alunos, dentro do ambiente de aprendizagem, onde é investigado se os resultados mostram progresso para a experiência de aprendizagem, individual ou em grupo, ou se os resultados observados indicam prejuízo ao processo de aprendizado dos alunos e turmas. As práticas detectadas podem ser tanto positivas quanto negativas. As

práticas são positivas quando promovem, ou facilitam, o aprendizado e o desenvolvimento acadêmico. Por sua vez, as práticas negativas são as que se comportam de forma contrária prejudicando o aprendizado. Quando se constata a ocorrência de uma prática, hipóteses devem ser formuladas para que sejam identificados ou explicados os motivos para a prática está acontecendo. Estas hipóteses serão confirmadas ou refutadas nas próximas etapas do processo.

A etapa de **Descobrimto de Padrões** está diretamente relacionada à etapa anterior, em que se busca descobrir o motivo pelo qual cada prática está ocorrendo. Nesta etapa é necessária à utilização de mineração e análise estatística dos dados, a fim de confirmar ou rejeitar as hipóteses criadas. Para tal, é necessário conhecer o ambiente de aprendizagem, investigando se existe a capacidade de coletar dados sobre o desempenho dos alunos e sobre a forma como eles se comportam. Dados que precisam ser corretos, úteis e em quantidade apropriada. Após a coleta dos dados, é crucial definir se eles precisarão de algum tipo de tratamento antes de serem utilizados no processo de mineração. É necessário utilizar técnicas da estatística descritiva, com o objetivo de conhecer os dados escolhidos. Com os dados escolhidos e tratados, o próximo passo é definir a tarefa de mineração apropriada para a análise das hipóteses. Feito isso e conhecendo os algoritmos de mineração de dados, saber-se-á qual deles fornece os resultados desejados.

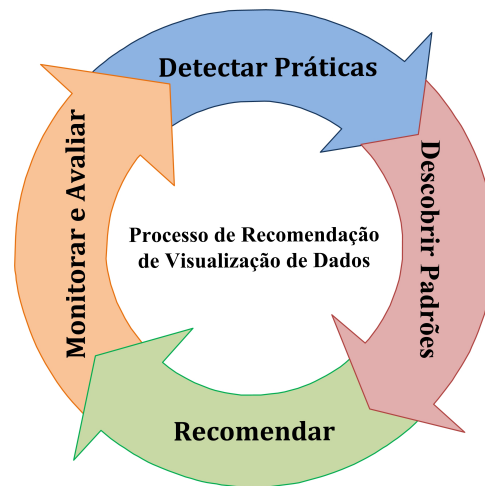
A etapa de **Recomendação** é a etapa responsável por disponibilizar aos professores opções de recomendações pedagógicas para que estas sejam encaminhadas aos usuários finais. Essas recomendações são criadas com bases na análise dos padrões descobertos na etapa anterior. Recomendações Pedagógicas consistem de entidades externas ou próprias do ambiente de aprendizado, cujo objetivo é oferecer orientações para que se possa superar dificuldades pedagógicas já detectadas ou previstas. É importante frisar que uma recomendação pedagógica para ser considerada útil, deve ser capaz de aprimorar a experiência de aprendizagem dos alunos aos quais ele foi direcionado. Estes alunos podem ter necessidades distintas e, certamente características e capacidades diferentes. Percebe-se, então, que a personalização dessas recomendações é um atributo capaz de maximizar suas chances de sucesso.

A etapa de **Monitoramento e Avaliação** é responsável por avaliar se os alunos estão progredindo positivamente em decorrência das recomendações recebidas, e também, avaliar o quão relevante estão sendo essas recomendações para as situações pedagógicas para as quais elas foram propostas. Nesta etapa, duas tarefas são realizadas: (i) O monitoramento do progresso dos alunos e avaliação da relevância das recomendações. Este monitoramento busca medir o quão bem os alunos estão progredindo após a recomendação. (ii) E a avaliação permite classificar uma recomendação quanto a sua relevância.

4.1.1 Adaptação do Processo

O processo foi adaptado para que a essa proposição pudesse ser alcançada, visto que optou-se por criar um processo corretado de Recomendação de Visualização de Dados (figura 13), de forma a torná-lo voltado para auxiliar o professor a visualizar melhor os dados expondo a eles uma visualização que esteja alinhada de acordo com suas características.

Figura 13 – Processo de Recomendação de Visualização de Dados



Fonte: Elaborado pela autora, adaptado de (PAIVA; BITTENCOURT; SILVA, 2013).

Para tanto, foram realizadas as seguintes etapas:

- **Detectar Práticas:** buscou-se identificar se as visualizações de dados estão proporcionando aos professores um resultado satisfatório na identificação de um problema. Se elas proporcionam aos docentes um resultado positivo em análises realizadas, ou se essa decorrência é um resultado contrário, em que se obtêm um resultado negativo em análises realizadas. O resultado positivo é pontuado mediante à quantidade de acerto para uma questão vinculada a uma visualização em específico, ou para uma decisão a ser tomada em um sistema, decisão que condicionará a situação de suas turmas, proporcionando melhorias ou não de desempenho. Já o resultado negativo é pontuado mediante à quantidade de erros para uma questão também vinculada a uma visualização em específico.
- **Descobrir Padrões:** analisou-se os dados dos professores baseado na interação com as visualizações, a fim de estabelecer padrões característicos dos perfis dos professores. Para tanto, foram utilizadas a mineração e análise estatística de dados. A prioridade é a identificação de quais características estão presentes nos perfis dos professores que obtiveram um resultado positivo nos acertos com base nas visualizações a eles expostas.

- **Recomendação:** serão sumarizados os padrões as práticas observadas para proporcionar ao professor uma recomendação de visualização pertinente ao seu perfil. Essa recomendação é resultado de um possível mapeamento entre as características colhidas com o número de acertos.
- **Monitorar e Avaliar:** monitoramento do desempenho dos professores com base nas visualizações recomendadas a cada um deles de acordo com o seu perfil, avaliando se de fato se as características foram mapeadas de forma consistentes.

4.1.2 Passos para a criação do modelo

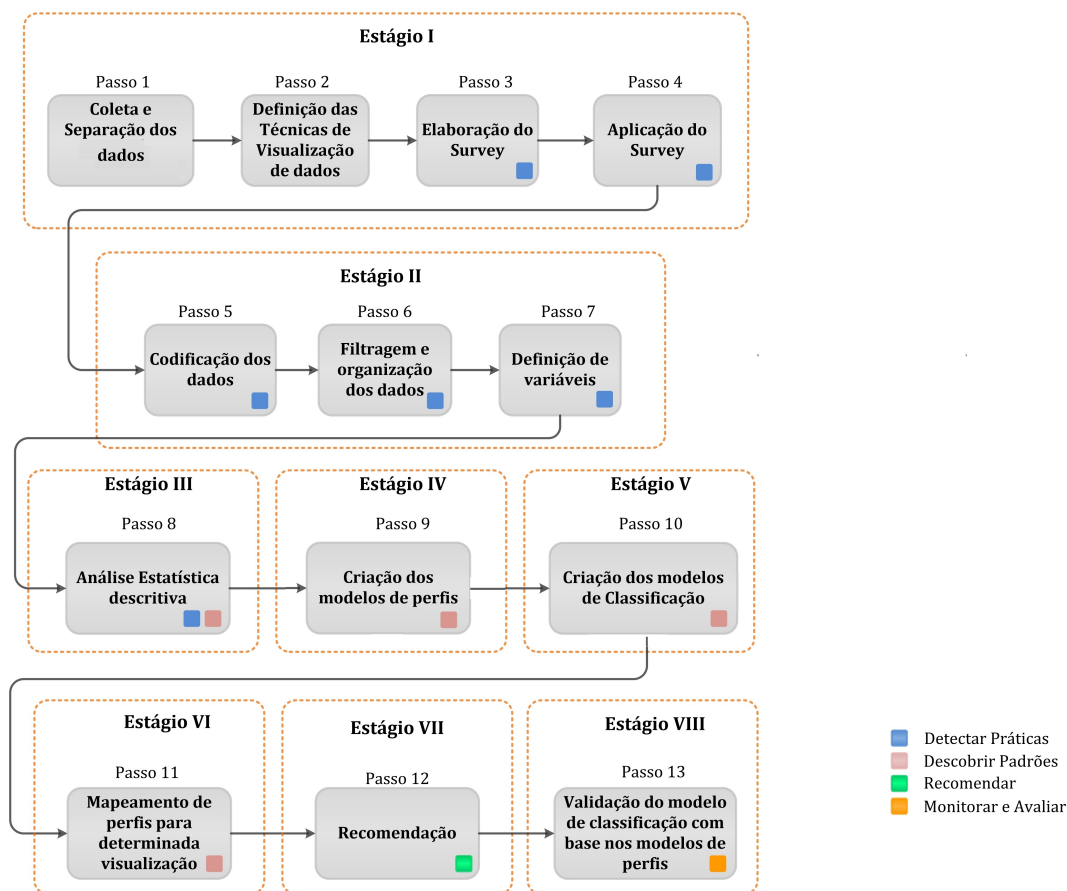
Para a criação modelo de classificação, estabeleceu-se oito estágios, conforme mostrado na figura 14. Estes estágios estão incorporados no processo de recomendação de visualização de dados, sendo eles explicados da seguinte forma:

- Estágio I: Correspondente aos passos de Coleta e Separação de dados (1). Definição das Técnicas de visualização de dados (2) e Elaboração do *Survey* (3). Aplicação do *Survey*(4).
- Estágio II: Corresponde aos passos de Codificação dos dados (5). Filtragem e organização dos dados (6) e Definição de variáveis (7).
- Estágio III: : Corresponde ao passo de Análise estatística descritiva.
- Estágio IV: Corresponde ao passo de Criação dos modelos de Classificação.
- Estágio V: Corresponde ao passo de Criação dos modelos de perfis.
- Estágio VI: Corresponde ao passo do mapeamento dos perfis criados com as visualizações.
- Estágio VII: Corresponde ao passo de recomendação da visualização.
- Estágio VIII: Corresponde aos passos de Validação dos modelos de classificação com base nos modelos de perfis.

O processo de recomendação de visualização de dados tem início (conforme é mostrado na figura 14) no estágio I com a etapa de Detectar práticas, a partir do passo 3, indo até o estágio III - passo 8. A etapa de Descobrir padrões inicia no estágio III - passo 8 e é concluída no estágio VI - passo 11. Após a finalização da última etapa, iniciamos a etapa de Recomendar, a respectiva etapa tem seu início e fim no estágio VII - passo 12. A última etapa a ser realizada é a etapa de Monitorar e Avaliar, sendo iniciada e finalizada no estágio VIII - passo 13.

Nas seções a seguir, serão especificados com mais detalhes cada estágio.

Figura 14 – Estágios da criação do modelo



Fonte: Elaborado pela autora.

4.1.2.0.1 *Estágio I*

No estágio I são realizadas as primeiras atividades referentes à pesquisa, é neste estágio que se define quais dados que serão usados, coletando e separando os dados de acordo com a necessidade do trabalho. Estas atividades fazem parte do passo 1.

O passo 2 é dado pelo definir quais as técnicas de visualização de dados serão usadas para mostrar graficamente situações pertinentes aos dados, mostrando, por exemplo, o número de pessoas que obtiveram uma nota inferior em um teste. Estas técnicas devem ser pensadas, levando em consideração o que se quer mostrar e o impacto disso aos visualizadores.

O passo 3 é o momento de elaboração do *Survey*. Este passo compreende as atividades de definição do público que o *survey* será aplicado, definição de problemas a serem atacados, criação de visualizações que mostrem estes problemas, levantamento dos questionamentos sobre problemas, com foco na interpretação das visualizações criadas.

As atividades de aplicação *survey*, estão compreendidas no passo 4. Este é o momento que se define em que local o *survey* será aplicado e se viabiliza que de fato esta aplicação

possa acontecer.

4.1.2.0.2 *Estágio II*

Com o passo de aplicação de *survey* finalizado, é necessário que os dados sejam codificados. O passo 5, referente à codificação, acontece como um pré-requisito para as investigações pertinentes aos dados, estabelecendo a criação de planilhas e tabelas quando necessário. Quando o *survey* é aplicado de forma impressa, este passo é caracterizado pela passagem das respostas dos *survey* para o computador.

No passo 6, as atividades de filtragem e organização dos dados são realizadas. O intuito é retirar os *surveys* que não tenham sido completamente preenchidos, com as informações referentes ao perfil e as respostas para cada questão. Após isso são selecionadas apenas as respostas do público investigado.

O passo 7 é estabelecido com definição de variáveis. Em que, mediante as questões do *surveys*, serão estabelecidas variáveis para viabilizar a análise (Estágio III). Após todas as variáveis estabelecidas, os dados são convertidos para formatos numéricos, de forma a estarem de acordo com o formato aceito pelo software estabelecido para a análise.

4.1.2.0.3 *Estágio III*

Neste estágio, caracterizado pelo passo 8, são realizadas as análises descritivas de todas as variáveis estabelecidas e pontuadas no *survey*. É buscado investigar todas as informações relevantes obtidas para cada questionamento feito. Investigando os valores respondidos para cada variável, bem como buscando, se preciso, agrupar variáveis para a criação de outras que possibilitem uma melhor e mais profunda análise dos dados.

4.1.2.0.4 *Estágio IV*

Este estágio compreende o passo 9, em que se busca criar os modelos de classificação de preferência. Tais modelos serão projetados única e exclusivamente com base nos dados codificados, nas variáveis estabelecidas e nas análises realizadas. Neste estágio, são exigidas ações de mineração de dados, bem como a definição do *software* para ser usado.

4.1.2.0.5 *Estágio V*

No estágio V foi criado os modelos de perfis, nos quais se utiliza para esta criação uma modelagem estatística dos dados, buscando um modelo válido baseado em cada uma das visualizações criadas.

Este estágio é representado pelo passo 10, o qual busca criar modelos estatísticos que consigam levantar as principais características de cada participante do *survey*. As quais tais

características serão pertinentes à análise das respostas obtidas em cada questão do *survey*. Este estágio é responsável por uma análise estatística mais profunda, e o estabelecimento do software para isso usado.

4.1.2.0.6 *Estágio VI*

No estágio VI, caracterizado pelo passo 11, busca-se fazer todo o mapeamento das informações levantadas em ambos modelos criados. Neste mapeamento serão agrupadas todas as características pessoais detectadas dos professores, de acordo com os modelos criados, com as visualizações mais propícia, das visualizações investigadas. Criando assim, um caminho entre as características pessoais dos professores e os tipos de visualizações.

4.1.2.0.7 *Estágio VII*

Este estágio caracterizado pelo passo 12, busca realizar a recomendação do tipo de visualização mais apropriada aos conjuntos de características pessoais dos professores. A respectiva recomendação ocorrer com base no mapeamento realizado no estágio anterior.

4.1.2.0.8 *Estágio VIII*

No passo 13, busca-se analisar a veracidade do modelo de classificação criado de acordo com os perfis estatisticamente estabelecido. Para tal foi projetado percorrer toda a classificação baseado nas características extraídas dos modelos de perfis, obtendo assim a validação do modelo de preferências.

4.2 MODELO DE APLICAÇÃO DO PROCESSO RECOMENDAÇÃO DE VISUALIZAÇÃO DE DADOS BASEADO EM SURVEY

Na utilização do processo de Recomendação de Visualização de Dados em um método de pesquisa *survey* é importante que todos os questionamentos levantados no *survey* estejam alinhados. Alinhamento chamado de bloco. Cada bloco deve conter em si, todas as informações que possam estar diretamente envolvidas no processo, tais informações podem ser pessoais, escolaridade, formação e etc.

O alinhamento em bloco deve acontecer a fim de após a aplicação do *survey*, o aplicador possa conseguir com mais facilidade, identificar que informações pertencem a quais etapas. Entretanto a organização das informações em cada bloco deverá depender exclusivamente o propósito o qual o processo está aplicado, dependendo assim dos critérios de organização do aplicador.

O aplicador deve separar também por blocos as questões relativas às visualizações utilizadas, visto que em sua análise, isso fará com que sejam levantadas mais rapidamente

as respostas sobre cada tipo de visualização utilizada, por exemplo questões relativas as visualizações A (questão 1, questão 2... questão n.) formarão um bloco, questões relativas as visualizações B (questão 1, questão 2... questão n.) formarão outro bloco e assim por diante. Até que todas as informações a serem coletadas estejam compreendidas em blocos, os quais melhor auxiliarão nas análises das respostas durante as etapas do processo de recomendação de visualização de dados.

4.2.1 Distribuição dos blocos no processo de Recomendação de Visualização de Dados

Os blocos deveram ser distribuídos nas etapas do processo Recomendação de Visualização de Dados de acordo com os critérios utilizados na formação de cada um dos blocos.

Na etapa de Detectar Padrões, conforme o diagrama de atividades (figura 15) devem ter blocos que investigam as visualizações. Estes contêm exclusivamente os questionamentos criados para cada visualização, no qual o aplicador do *survey* poderá identificar o que está acontecendo, de práticas positivos ou negativos com um professor na investigação do problema visualizado.

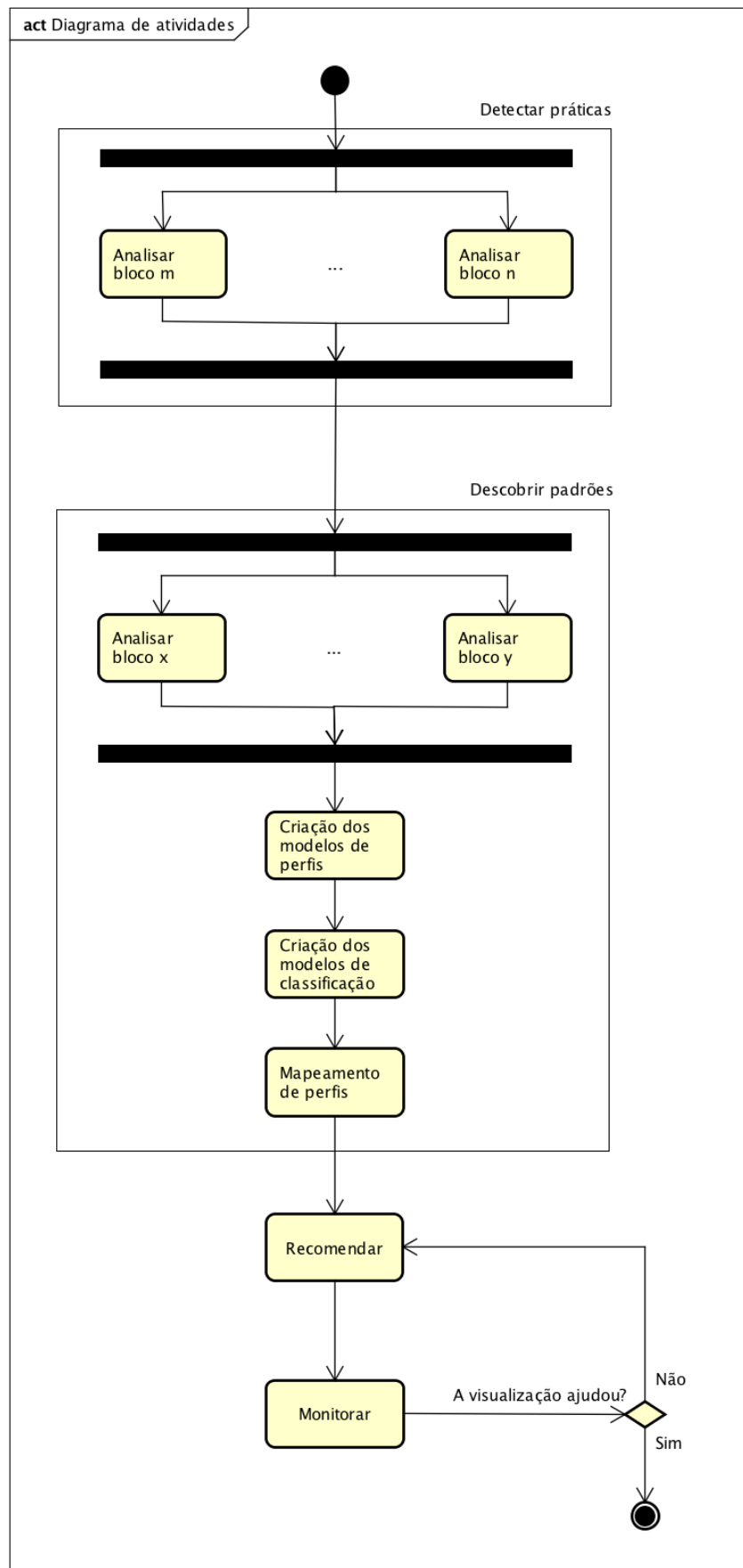
Na etapa de Descobrir Padrões, conforme o diagrama de atividades (figura 15), devem ter blocos que investigam o professor. Estes blocos contêm exclusivamente os questionamentos criados sobre as características de cada professor, onde o aplicador do *survey* poderá identificar por que o professor a tem práticas positivas ou negativas na investigação de um problema visualizado.

As etapas de Recomendar e Monitoramento não possuem blocos. Estas etapas são etapas dependentes e sequenciais das etapas de Detectar Práticas e Descobrir padrões, obrigatoriamente, só podem ocorrer após a criação dos modelos de classificação, criação dos modelos de perfis, mapeamento de perfis para determinada visualização. Sendo que a etapa recomendar irá exibir ao professor a visualização que melhor se enquadre em suas características e a etapa Monitorar e Analisar ocorre para qualificar a etapa Recomendar, visto que investiga se a visualização exibida para o professor contribuiu para em melhor entendimento.

O diagrama de atividades presente na figura 15, mostra de forma mais clara a distribuição dos blocos nas duas etapas. É possível compreender que para sair da etapa de Detectar Práticas, primeiro é necessário analisar todos os blocos para ela estabelecidos. Isto também acontece na etapa de Descobrir Padrões. É importante ressaltar que ambas buscam objetivos diferentes. Já as etapas de Recomendar e Monitorar e Avaliar, só irão acontecer depois de finalizadas todas as outras atividades. Elas podem acontecer várias vezes, caso a visualização não esteja correspondendo ao esperado, de acordo com a opinião dos

professores.

Figura 15 – Diagrama de Atividades de distribuição de blocos de acordo com o processo



Fonte: Elaborado pela autora.

5 MÉTODO DE PESQUISA

Visando realizar o processo descrito no capítulo 4, bem como viabilizar meios para a criação dos modelos de personalização de visualização de dados para professores, escolhemos como método de pesquisa o *Survey*.

O *Survey* método quantitativo de pesquisa, é descrito por (FREITAS et al., 2000) como a obtenção de dados ou informações sobre características, ações ou opiniões de um determinado grupo de pessoas, indicando assim como representante de uma população alvo.

Para (RIBEIRO et al., 2012) o *Survey* é método voltado a obter opiniões ou medir atitudes de pessoas ou de grupos, visando à identificação da verdade pela quantidade por meio de aplicação de questionário.

5.1 PLANEJAMENTO DO SURVEY

Nesta seção, será detalhado o planejamento do *Survey* que foi projetado para este trabalho. Dentro do planejamento, encontra-se a definição da questão de pesquisa e hipóteses, seleção das variáveis, escolha das visualizações de dados, preparação dos dados para a criação do *Survey*, elaboração do *Survey*, aplicação do *Survey*, preparação para a análise dos dados, instrumentação para as análises dos dados e ameaças a validade.

5.1.1 Questões de Pesquisa e Hipóteses

Este *Survey* foi elaborado com a finalidade que encontrar meios que melhor possibilitem aos professores uma visualização que esteja ligada ao seu perfil. Tendo esse princípio como objetivo, nós buscamos responder as seguintes questões:

- Questão de pesquisa 1 - Será que o perfil do professor influencia na escolha de uma visualização específica a ser vista?

A questão de pesquisa acima implica nas seguintes hipóteses:

- **H1 – 0** : A preferência do professor é igual para visualizações tradicionais e não tradicionais;
- **H1 – 1**: A preferência do professor é diferente para visualizações tradicionais e não tradicionais.

- Questão de pesquisa 2 - Que características dos professores podem estar presente na preferência de uma determinada visualização?

A questão de pesquisa acima implica nas seguintes hipóteses:

- **H2 – 0:** As características dos professores são iguais na preferência entre visualizações de dados tradicionais e não tradicionais;
- **H2 – 1:** As características dos professores são diferentes na preferência entre visualizações de dados tradicionais e não tradicionais.
- Questão de pesquisa 3 - Será possível fazer uma previsão computacional de preferência de visualização de dados a partir de características?

A Tabela 2 define formalmente as hipóteses de pesquisa supracitadas. É importante destacar que pV corresponde a preferência de visualização, cV corresponde a características dos professores, T corresponde a visualização tradicionais, nT corresponde a visualizações não tradicionais.

Tabela 2 – Definição das hipóteses

Questões de Pesquisa	Hipótese Nula	Hipótese Alternativa
QP1	$pV(T) = pV(nT)$	$pV(T) \neq pV(nT)$
QP2	$cV(T) = cV(nT)$	$cV(T) \neq cV(nT)$

Fonte: Elaborado pela autora.

5.1.2 Seleção de Variáveis

Em nosso *Survey*, identificamos como variável independente, também chamada de fator, o nível de percepção. Esta variável identifica o nível de percepção dos participantes de acordo com a classificação das visualizações tradicionais e não tradicionais. Assim abaixo, temos a tabela 3, de nível de fatores, que melhor expõem estas informações:

Tabela 3 – Nível de Fatores

Fator	Nível	Descrição
Preferência	T	Visualizações tradicionais
	Tn	Visualização não tradicionais

Fonte: Elaborado pela autora.

Conforme o estabelecimento da variável dependente, identificamos com variáveis independentes as questões corretas, o quesito ajuda e quesito tempo. Sendo que:

- O quesito questões corretas: Essa variável corresponde a soma das questões corretas, respondidas pelos participantes para cada bloco de questões.
- O quesito ajuda: Essa variável representa o quanto o gráfico ajudou ao participante a responder as questões, sendo medida em uma escala de zero (0) a cem (100).
- O quesito tempo: Essa variável representa o tempo que o participante respondeu, sendo medida em uma escala de zero (0) a cem (100).

5.1.3 Escolha das Visualizações de Dados

Para a elaboração deste *Survey*, foram escolhidos três (3) diferentes representações visuais. Uma delas considerada tradicional e duas delas não tradicionais. A primeira delas foi criada com base na técnica de gráfico de barras também conhecido como gráfico de colunas. A segunda, foi desenhada baseada na técnica de hierarquia de bolhas. E por fim, na terceira delas, optou-se por agrupar características de duas técnicas, a técnica de gráfico de unidade e a técnica de gráfico de linha, para a criação de uma única visualização.

É importante ressaltar que, neste trabalho, não se busca qualificar a efetividade de nenhuma das técnicas utilizadas, mas entender como os professores percebem possíveis problemas dentro de uma turma, e se eles conseguem, por meio da visualização tomar decisões acertadas.

5.1.4 Preparação dos dados para a criação do *Survey*

Para a realização da elaboração do *Survey*, foram utilizados dados reais de duzentos e trinta e um (231) estudantes, os quais foram cedidos pelo ambiente MeuTutor^{16,17}. correspondiam ao uso do ambiente web em escolas públicas do interior de alagoas. Buscando uma melhor organização dos dados, optou-se por separa-los, classificando a localização das escolas em três grupos (Grupo I, Grupo II e Grupo III), sendo estes:

- Grupo I - São Miguel: composto por cento e quarenta e sete (147) estudantes;
- Grupo II - São Sebastião: composto por cinquenta e cinco (55) estudantes;
- Grupo III – Escolas pertencentes a outros municípios: composto por vinte e nove (29) estudantes.

¹⁶ MeuTutor é um ambiente web educacional gamificado que incide sobre a aprendizagem personalizada. O qual realiza o monitoramento das atividades de seus alunos em uma forma personalizada (NETO, 2014).

¹⁷ Disponível em: <http://meututor.com.br/>

Em cada um destes grupos, foi avaliado o cenário de desempenho dos alunos, em que se comparou o desempenho dos alunos com as interações deles dentro do sistema. Para o cálculo do desempenho foi considerada a razão entre o número de questões respondidas corretamente pelo montante total de questões respondidas através do ambiente online de aprendizagem.

Com o resultado obtido, categorizou-se os estudantes, de cada grupo, em três (3) subgrupos e atribui-se uma cor a cada subgrupo. Os estudantes com desempenho superior receberam a cor verde, os estudantes com desempenho médio receberam a cor amarela, e os estudantes com desempenho inferior receberam a cor vermelha. Após a categorização, comparou-se o desempenho dos alunos com suas interações dentro do sistema, com os seguintes recursos de aprendizagem:

- (i) o número percentual de vídeos assistidos;
- (ii) o número percentual de questões respondidas corretamente;
- (iii) o número percentual de questões respondidas incorretamente;
- (iv) o número percentual de acessos no ambiente.

Após a organização dos dados e categorização dos grupos criamos os três diferentes gráficos relacionavam o desempenho dos alunos com a interação com os recursos acima citados. Em que se buscou elaborar questões que fossem condizentes com as situações expostas nos gráficos, a fim de obter e coletar a sustentação necessária para as questões de pesquisas e hipóteses levantadas.

5.1.5 Elaboração do *Survey*

O *Survey* era composto por trinta e duas questões, em que estavam divididas em sete blocos. O primeiro bloco é composto pelas informações pessoais (nome, sexo, idade, tempo de trabalho como professor). O segundo bloco corresponde à escolaridade (em que tipo de instituição – pública ou privada, foram concluídos os ensinos fundamental e médio).

O terceiro bloco é referente à formação (se o participante cursou ou está cursando algum curso técnico, em que tipo de instituição este curso é ofertado – pública ou privada. Se o participante está cursando ou concluiu o ensino superior em uma instituição este curso é ofertado – pública ou privada, e o nome do(s) curso(s). O quarto bloco, buscava identificar se o participante era daltônico (uma opção a ser marcada e um teste com uma imagem de identificação de daltonismo).

O quinto, sexto e sétimo bloco traziam em si, diferentes gráficos, os quais com base neles, eram esperados que os participantes respondessem a sete (7) questões, cinco (5)

questões de acordo com as informações presentes nos gráficos e duas (2) para avalia-los. As cinco questões (5) eram de múltiplas alternativas e as duas questões (2) de avaliação dos gráficos eram uma escala, o participante tinha como alternativas as opções: muito pouco, pouco, indiferente, muito e muitíssimo.

5.1.6 Aplicação do *Survey*

O *Survey*¹⁸, foi aplicado no IV Congresso Brasileiro de Informática na Educação e X conferência Latino-America de Objetos e Tecnologias de Aprendizagem, ocorrido no Brasil, no período de 26 a 30 de outubro de 2015. Sendo distribuídos, mil (1000) exemplares, sob a autorização da equipe organizadora, dentro de pastas entregues por ela, no ato do credenciamento.

Durante os cinco dias de realização do evento, foram solicitados em meio as palestras, aos participantes que ainda não tivessem preenchido o *Survey*, que o fizessem, uma vez que estariam contribuindo diretamente para o enriquecimento da pesquisa e concorrendo a um sorteio de quinze (15) livros e pequeno brindes.

5.1.7 Preparação para a análise dos dados

Após o encerramento do evento, foram contabilizados o número de trezentos e trinta e três (333) *Surveys* respondidos, dos quais, o número de participantes professores que haviam respondido era equivalente a duzentos e trinta e cinco (235). Para que pudéssemos então dá andamento na análise estatística optamos por utilizar apenas os *Surveys*, que foram respondidos completamente, dos quais obtivemos o número de equivalente a cento e oitenta e cinco (185).

Durante esta etapa, nós instanciamos como variáveis todas as informações coletadas dos sete (7) blocos em uma planilha, para que se pudesse dá início à análise. Todas as variáveis correspondentes às características pessoais dos professores instanciadas (conforme a tabela 4), foram codificadas com valor correspondentes a zero (0) e um (1) exceto a idade. Resolvemos deixar a informação passada pelo professor, conforme a tabela 3. Um dado importante a ser informado é que para professores com mais de dez (10) anos de experiência não atribuiu-se variável a essa informação, uma vez que isto era subentendido caso professor tivesse valor zero (não) nas variáveis Exp1 (menos de 1 ano de experiência), Exp2 (de 1 a 5 anos de experiência) e Exp3 (de 6 a 10 anos de experiência).

Já as variáveis relacionadas às questões e suas respectivas respostas, dadas pelos professores, foram instanciadas com base na questão estabelecida, o nome atribuído da variável, e o seu valor. Este último correspondente a pontuação referente aos acertos obtido, (conforme a tabela 5).

¹⁸ O *Survey* aplicado encontra-se nos anexos desta dissertação

Tabela 4 – Instanciando Variáveis - Características Pessoais

Informação	Variável	Valor
Sexo	Sexo	0 = Masculino ou 1 = feminino
Idade	Idade	—
Tempo de experiência como professor	Menos de 1 ano - Exp 1	1 = sim ou 0 = não
Tempo de experiência como professor	De 1 a 5 anos - Exp2	1 = sim ou 0 = não
Tempo de experiência como professor	De 6 a 10 anos - Exp 3	1 = sim ou 0 = não
Ensino Fundamental	Esc 1	1 = Instituição privada ou 0 = Instituição pública
Ensino Médio	Exc2	1 = Instituição privada ou 0 = Instituição pública
Cursando ou concluiu curso técnico	Fom1	1 = sim ou 0 = não
Curso técnico em instituição pública	Form2	1 = sim ou 0 = não
Curso técnico em instituição privada	Form3	1 = sim ou 0 = não
Ensino superior (concluído ou em andamento)	Form4	1 = sim ou 0 = não
Ensino superior em instituição pública	Form5	1 = sim ou 0 = não
Ensino superior em instituição privada	Form6	1 = sim ou 0 = não
Ensino superior na área de Ciências exatas	CurSup1	1 = sim ou 0 = não
Ensino superior na área de Ciências Humanas	CurSup2	1 = sim ou 0 = não
Ensino superior na área de Ciências Biológicas	CurSup3	1 = sim ou 0 = não
Daltônico	Dalt	1 = não ou 0 = sim
Teste de daltonismo	Dalt1	1 = aprovado ou 0 = reprovado

Fonte: Elaborado pela autora.

Tabela 5 – Instanciando Variáveis - Questões e Respostas

Informação	Variável	Valor
Visualização 1 - Questão 1	QA1	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 1 - Questão 2	QA2	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 1 - Questão 3	QA3	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 1 - Questão 4	QA4	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 1 - Questão 5	QA5	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 1 - Questão 6	QA6	Entre 0, 25, 50, 75, 100.
Visualização 1 - Questão 7	QA7	Entre 0, 25, 50, 75, 100.
Visualização 2 - Questão 1	QB1	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 2 - Questão 2	QB2	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 2 - Questão 3	QB3	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 2 - Questão 4	QB4	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 2 - Questão 5	QB5	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 2 - Questão 6	QB6	Entre 0, 25, 50, 75, 100.
Visualização 2 - Questão 7	QB7	Entre 0, 25, 50, 75, 100.
Visualização 3 - Questão 1	QC1	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 3 - Questão 2	QC2	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 3 - Questão 3	QC3	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 3 - Questão 4	QC4	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 3 - Questão 5	QC5	Entre 0, 0.25, 0.50, 0.75, 1, sendo 0 para erros e 1 para acerto total.
Visualização 3 - Questão 6	QC6	Entre 0, 25, 50, 75, 100.
Visualização 3 - Questão 7	QC7	Entre 0, 25, 50, 75, 100.

Fonte: Elaborado pela autora.

5.1.8 Instrumentação para as análises dos dados

- Para a análise estática e criação dos modelos de regressão e regressão beta, utilizamos os *softwares* OxEdit^{19 20} e o R^{21 22}.
- Para os modelos de Classificação o utilizamos o *software* Weka^{23 24}

5.1.9 Ameaças a Validade

Embora tenhamos tentado amenizar as possíveis ameaças que por ventura pudessem ocorrer, é preciso citar as que foram identificadas em nosso experimento. A primeira delas foi organização do Survey, uma vez que todas as questões, foram organizadas seguindo a mesma ordem de apresentação de visualização e questões a serem respondidas. Desta forma era possível que segundo bloco de questões se tornasse mais cansativo que o primeiro, e o terceiro bloco a ser mais cansativo que todos os outros. E a segunda ameaça identificada foi a possibilidade de uma questão ter mais de uma resposta, o que fez com que alguns professores não conseguissem alcançar todos os pontos corresponde a determinadas questões.

¹⁹ Editor de texto, no qual foram organizados os dados e criados os cálculos matemáticos para a análise estatística

²⁰ Disponível em: <http://www.doornik.com/>

²¹ R é um ambiente de software livre para computação estatística e gráficos.

²² Disponível para download em: <https://www.r-project.org/>

²³ Weka é um conjunto de algoritmos de aprendizado de máquina para tarefas de mineração de dados

²⁴ Disponível para download em: <http://www.cs.waikato.ac.nz/ml/weka/>

6 ANÁLISES E RESULTADOS DO SURVEY

As análises e os resultados fazem parte da etapa de Detectar prática e de Descobrir Padrões do processo de Recomendação de Visualização de Dados.

No presente trabalho não foi possível realizar as etapas de Recomendar e Monitorar atividades, do processo de Recomendação de Visualização de dados devido ao tempo, estas que seriam atuantes após a criação dos modelos estatísticos de perfis e dos modelos de classificação.

6.1 MODELOS DE REGRESSÃO

Os valores observados em problemas reais podem ser vistos com resultados de um experimento. As variáveis relacionadas com este experimento podem ser relacionadas com base em um modelo matemático. Esse modelo matemático tenta reproduzir o verdadeiro processo gerador dos dados. No entanto, é necessário considerar além da expressão matemática a natureza aleatória dos dados. Essa a principal característica dos modelos de regressão.

Considere o modelo linear, dado por:

$$y = X\beta + \epsilon =$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \underbrace{\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}}_{\begin{pmatrix} x_1 & x_2 & \dots & x_k \end{pmatrix}} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Tem-se que y é um vetor de n observações da variável aleatória resposta ou de interesse, X é uma matriz $n \times k$ formada pelas covariadas, em que cada coluna é um conjunto de n observações da covariada x_t , $t = 1, \dots, k$, totalizando k covariadas. As covariadas não são variáveis aleatórias. Ainda temos β que é um vetor de k parâmetros também fixos e desconhecidos (não são variáveis aleatórias) e ϵ um vetor de n erros aleatórios.

O objetivo de um modelo de regressão é explicar o máximo possível o comportamento aleatório da resposta, de tal forma que o não explicado deve estar contido no erro aleatório ϵ .

Obviamente, o erro deve ser bem pequeno, tanto que, espera-se que ele seja zero. Isto é, em um modelo de regressão supõe-se que:

$$E(\epsilon) = \mu_\epsilon = 0.$$

De onde segue que,

$$E(y) = E(X\beta) + E(\epsilon) \Leftrightarrow E(y) = X\beta \Leftrightarrow \mu = X\beta.$$

Consequentemente, o modelo final é

$$\mu = X\beta.$$

Tem-se que $E(X\beta) = X\beta$ e

$$\text{var}(y) = \underbrace{\text{var}(X\beta)}_0 + \underbrace{\text{var}(\epsilon)}_{\sigma^2} \Leftrightarrow \text{var}(y) = \sigma^2.$$

Pois, $\text{var}(\epsilon) = \sigma^2$ e $\text{var}(X\beta) = 0$, variância de uma constante é zero.

A representação do modelo considerando a i -ésima observação é dada por:

$$\mu_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik}, \quad i = 1, \dots, n.$$

Para que o modelo acima seja conhecido é preciso estimar $\beta_1, \beta_2, \dots, \beta_k$. Tipicamente isto é realizado usando o método de máxima verossimilhança. Assim, $\hat{\mu}_i$ é obtido quando são obtidos: $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$, tal que:

$$\hat{\mu}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3} + \dots + \hat{\beta}_k x_{ik}, \quad i = 1, \dots, n.$$

Para usar o método de máxima verossimilhança é preciso definir o tipo de distribuição de probabilidades que a variável aleatória y segue. Apesar de a distribuição normal ser a mais conhecida, nem sempre essa distribuição é adequada para todas as variáveis aleatórias contínuas. Se $y \in (0, 1)$ podemos pensar na distribuição beta ou na distribuição simplex. Com o objetivo de viabilizar o uso de outras distribuições além da normal, apresentamos uma expressão mais geral para modelos lineares:

$$g(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik}.$$

Observemos que $g(\mu_i)$ é uma função de ligação, que conecta a média da variável resposta e o modelo envolvendo as covariadas e os β 's. Quando $y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, temos que $g(\mu_i) = \mu_i$, ou seja g é o que chamamos de função identidade. De fato, não houve necessidade de transformar a média.

No modelo normal assim como a resposta que pertence a todos os reais, $y \in (-\infty, +\infty)$, o mesmo ocorre com sua média $\mu \in (-\infty, +\infty) = \mathbb{R}$. E o mesmo deve ocorrer com $\widehat{\mu}_i, \widehat{\mu}_i \in (-\infty, +\infty) = \mathbb{R}$. Neste caso, da distribuição normal, $\widehat{\mu}_i$ pode assumir qualquer valor real e conseqüentemente, $\widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_k$ estão livres para também assumirem quaisquer valores reais.

Isto não acontece por exemplo se a variável resposta $y \in (0, 1)$, $\mu \in (0, 1)$ e temos que garantir que $\widehat{\mu} \in (0, 1)$. Com esta restrição os $\widehat{\beta}$'s não estão livres, pois deve ser garantido que $X\widehat{\beta}$ só assuma valores no $(0, 1)$. A alternativa é aplicar uma função g em μ_i de forma que $g(\mu_i) \in (-\infty, +\infty) = \mathbb{R}$. Então, os $\widehat{\beta}$'s estão liberados, como veremos abaixo.

6.2 MODELO DE REGRESSÃO BETA

A distribuição beta é tipicamente usada para modelar variáveis aleatórias que se distribuem de forma contínua no $(0, 1)$, tais como taxas, porcentagens e escores. A densidade beta é flexível podendo assumir diversas formas dependendo da combinação de valores de seus parâmetros. Seja y_1, \dots, y_n uma amostra de v.a's independentes tal que cada $y_i, i = 1, \dots, n$, segue a distribuição beta com densidade:

$$f(y; \mu_i, \phi_i) = \frac{\Gamma(\phi_i)}{\Gamma(\mu_i \phi_i) \Gamma((1 - \mu_i) \phi_i)} y^{\mu_i \phi_i - 1} (1 - y)^{(1 - \mu_i) \phi_i - 1}, \quad 0 < y < 1, \quad (6.1)$$

onde $0 < \mu_i < 1$ e $\phi_i > 0$. Aqui, $E(y_i) = \mu_i$ e $\text{var}(y_i) = (\mu_i(1 - \mu_i))/(1 + \phi_i)$.

Tem-se que ϕ representa um parâmetro de precisão, já que quanto maior ϕ menor a variância de y_i , conseqüentemente, ϕ^{-1} é um parâmetro de dispersão, (ALMEIDA; SOUZA, 2015) propõem que a média da variável resposta y_i , i.e., μ_i possa ser escrita como:

$$g(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik}, \quad i = 1, \dots, n.$$

Como o objetivo da função de ligação g é permitir que os $\widehat{\beta}_i$'s possam assumir qualquer valor real e dado que $\mu_i \in (0, 1)$, uma função de ligação que conduz essa média a todos os reais é a função de ligação logito, dada por:

$$\mu_i \in (0, 1) \leftrightarrow \log \left\{ \frac{\mu_i}{(1 - \mu_i)} \right\} \in (-\infty, +\infty) = \mathbb{R}.$$

Assim como utilizamos um modelo matemático para tentar explicar a média da variável resposta μ_i , que é um parâmetro desconhecido, também é possível e as vezes

necessário fazer o mesmo com a variância da resposta, que neste caso implica em modelar o parâmetro ϕ . De fato, é sugerido um modelo para ϕ quando suspeita-se que a dispersão não é constante para os dados, ou que é possível haver grupos com dispersões diferentes. Assim, (SMITHSON; VERKUILEN, 2005) propõem um modelo de regressão beta em que

$$h(\phi_i) = \gamma_1 + \gamma_2 z_{i2} + \gamma_3 z_{i3} + \dots + \gamma_q z_{iq}, \quad i = 1, \dots, n.$$

Neste caso, como $\phi > 0$ uma função de ligação adequada é $\log(\phi) \in IR$. Note que é necessário estimar os β_t 's ($t = 1, \dots, k$) e os γ_j 's ($j = 1, \dots, q$) para que ϕ_i e μ_i sejam estimados. Isto é feito utilizando o método de máxima verossimilhança^{25,26}.

Note que $\beta = (\beta_1, \dots, \beta_k)^\top$ e $\gamma = (\gamma_1, \dots, \gamma_q)^\top$ são vetores e $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_k)^\top$ e $\hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_q)^\top$ são os seus respectivos estimadores de máxima verossimilhança.

6.3 ANÁLISE DE DIAGNÓSTICO

Verificar se um determinado modelo é uma representação adequada dos dados, é um passo importante da análise estatística. A construção de um modelo de regressão envolve a definição da distribuição a variável de resposta, a escolha da função de ligação, a escolha das covariáveis. Vários fatores podem levar um modelo ajustado pobre. Por exemplo, a função de ligação inadequado, omissão de covariáveis importantes, a escolha errada da distribuição da variável resposta, pontos influentes, especificação incorreta da variância entre outros fatores.

Ou seja, modelos estatísticos estão baseados em certas suposições. A fim de ter confiança na análise é dever verificar se os pressupostos associados são válidos. Isso pode ser alcançado por meio de análise de diagnóstico. Esses diagnósticos podem ser realizados com base na análise de resíduos.

A maior parte dos resíduos são baseados nas diferenças entre o respostas observadas (y) e a média estimada ($\hat{\mu}$). Por exemplo $r_i = y_i - \hat{\mu}_i$, ou seja, o resíduo é uma medida de discrepância entre os dados reais e o modelo ajustado. Aqui vamos utilizar o resíduo proposto por (ESPINHEIRA et al., 2015), $r_{p,i}^{\beta\gamma}$ denominado resíduo combinado e baseado na diferença

$$(y_i^* - \hat{\mu}_i^*), \quad \text{em que } y_i^* = \log \left\{ \frac{y_i}{(1 - y_i)} \right\} \quad \text{e } \mu_i^* = E(y_i^*).$$

Os gráficos de resíduos versus índices das observações ou versus valores preditos ($\hat{\mu}_i$) são os mais básicos. Se um modelo está especificado corretamente, então estes gráficos não deve apresentar nenhuma tendência, os resíduos devem estar aleatoriamente distribuídos em torno do zero. A presença de quaisquer características sistemáticas tipicamente implica um falha de um ou mais pressupostos do modelo.

²⁵ Disponível em: <https://www.ime.usp.br/giapaula/cursospos.htm>

²⁶ Pesquisadores podem usar o pacote *betareg* o qual está disponível no software estatístico *R*

Outro gráfico de resíduos importante é o gráfico de probabilidade normal com envelope simulado, que pode ser usada mesmo quando as distribuições empíricas dos resíduos não são normais. Se o modelo está adequado aos dados, esperamos que a maioria dos resíduos estejam aleatoriamente distribuídos dentro das bandas do envelope.

6.4 ANÁLISE ESTATÍSTICA DESCRITIVA

6.4.1 Descrição das variáveis - Resposta Global

As questões específicas relacionadas a cada tipo de visualização irão ser utilizadas tanto para construir a *resposta* quanto como covariadas dos modelos.

Visualização I:

QA₁, QA₂, QA₃, QA₄ e QA₅ - são variáveis referentes a perguntas de múltiplas escolhas respondidas com base na interpretação da visualização I. Cada uma destas perguntas podem ter mais de uma resposta correta. Assim, as variáveis de QA₁ a QA₅ assumem valores do tipo 0, 0.25, 0.50, 0.75 e 1.0. No final, é construída a primeira variável resposta, Resp1a, que é a média entre as variáveis QA₁ a QA₅.

QA₆ e QA₇ - São escores que qualificam a visualização. QA₆ se a visualização foi útil para interpretação e QA₇ se o tempo fez diferença. Essas variáveis assumem valores de 0 a 100. No final, é construída a segunda variável resposta, Resp2a, que é a média entre as variáveis QA₆ a QA₇.

Visualização II:

QB₁, QB₂, QB₃, QB₄ e QB₅ - são variáveis referentes a perguntas de múltiplas escolhas respondidas com base na interpretação da visualização II. Cada uma destas perguntas podem ter mais de uma resposta correta. Assim, as variáveis de QB₁ a QB₅ assumem valores do tipo 0, 0.25, 0.50, 0.75 e 1.0. No final, é construída a primeira variável resposta, Resp1b, que é a média entre as variáveis QB₁ a QB₅.

QB₆ e QB₇ - São escores que qualificam a visualização. QA₆ se a visualização foi útil para interpretação e QA₇ se o tempo fez diferença. Essas variáveis assumem valores de 0 a 100. No final, é construída a segunda variável resposta, Resp2b, que é a média entre as variáveis QA₆ a QA₇.

Visualização III:

QC₁, QC₂, QC₃, QC₄ e QC₅ - são variáveis referentes a perguntas de múltiplas escolhas respondidas com base na interpretação da visualização III. Cada uma destas perguntas podem ter mais de uma resposta correta. Assim, as variáveis de QC₁ a QC₅ assumem valores do tipo

0, 0.25, 0.50, 0.75 e 1.0. No final, é construída a primeira variável resposta, Resp1c, que é a média entre as variáveis QC₁ a QC₅.

QC₆ e QC₇ - São escores que qualificam a visualização. Q QC₆ se a visualização foi útil para interpretação e QC₇ se o tempo fez diferença. Essas variáveis assumem valores de 0 a 100. No final, é construída a segunda variável resposta, Resp2c, que é a média entre as variáveis QC₆ a QC₇.

Neste ponto surge a construção da primeira variável resposta: RespGeral, que é a média entre Resp1a, Resp1b, Resp1c, Resp2a, Resp2b, Resp2c. Que representa o escore de pontuação geral atribuído ao indivíduo sem separação por tipo de visualização. Ou seja, é uma resposta que considera a média de pontuações das três visualização sem fazer distinções entre as visualização. Essa variável resposta será utilizada para identificar como o efeito do perfil do avaliador no escore médio de pontuação global.

6.4.2 Descrição das variáveis - Respostas Parciais

Agora o objetivo de verificar o efeito de cada tipo de visualização no escore de pontuação. Para isso foram criadas duas variáveis respostas, Resp1 e Resp2.

Resp1 é a media das questões Q1 a Q5 de cada figura e Resp2 é a média entre Q6 a Q7 de cada figura. Assim, criou-se um banco de dados em que são replicadas duas vezes o conjunto original de dados e incluído dummies para separar ou indicar as Resp1 e Resp2 para cada uma das três figuras.

As primeiras 185 valores de Resp1 e Resp2 são referentes a visualização I, de 186 a 370 são 185 valores de Resp1 e Resp2 referentes a visualização II e de 371 a 555 são 185 valores de Resp1 e Resp2 referentes a visualização III. Note que os valores de todas as outras variáveis são replicadas à cada 185 observações. Para separar ou identificar cada visualização são criadas três Dummies: IQ1: 1, se visualização I e 0, caso contrário. IQ2 1, se visualização II e 0, caso contrário. IQ3 1, se visualização III e 0, caso contrário. Só duas destas variáveis podem entrar no modelo a cada vez.

A variável resposta de interesse final é a média entre Resp1 e Resp2, para cada um dos três tipos de visualizações. Vamos chamar essa resposta de parcial. A Resposta Parcial foi nomeada como escore de pontuação e pode ser vista como uma medida do desempenho do tipo de visualização, permitido que as respostas das questões Q1 a Q5 fossem mais próximas das esperadas, com facilidade de interpretação Q6 e com uma relação com o tempo que não prejudicasse a interpretação da visualização associada.

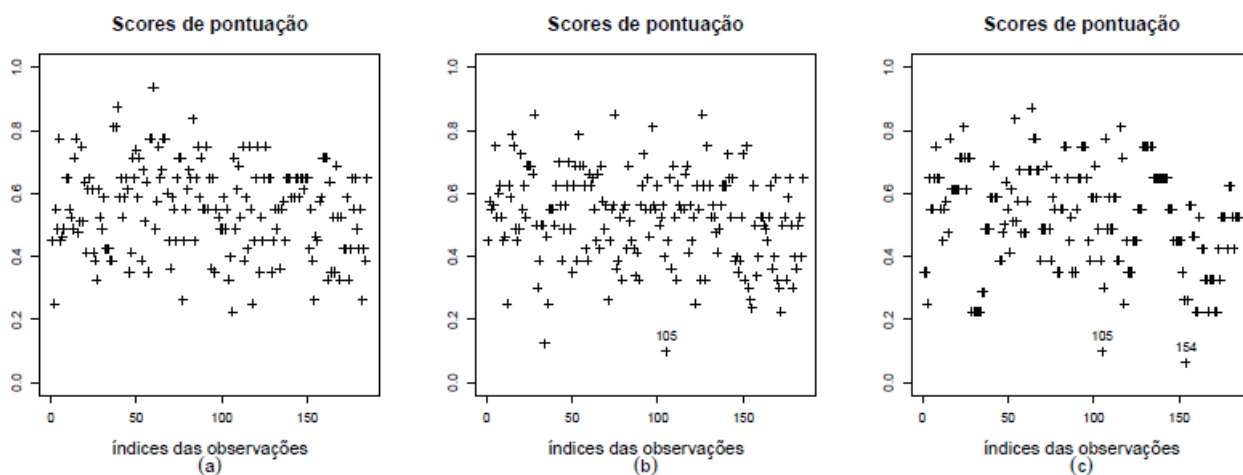
6.4.3 Análise descritiva

Inicialmente realizamos uma análise descritiva da variável resposta parcial (pois considera separadamente as pontuações de cada uma das três visualizações) considerando cada

um dos três tipos de visualizações, representadas respectivamente através das visualizações I, II e III.

Com base na análise descritiva foi identificado que os escores de pontuação são muito similares para os três tipos de visualizações. No entanto, visualização do tipo I possui valores dos escores de pontuação superiores aos escores relativos aos tipos de visualizações II e III. Onde, nos dois últimos tipos de visualizações, nota-se a ocorrência de outliers.

Figura 16 – Diagramas de dispersão da variável resposta para cada tipo de visualização.



Fonte: Elaborado pela autora.

Na Figura 16 encontram-se os diagramas de dispersão dos escores. A partir destes gráficos é possível identificar o caso 105 com o outlier da distribuição dos escores relativos a visualização do tipo II e quanto a visualização do tipo III, o outlier é o caso 156.

6.5 MODELOS: SCORES PARCIAIS DE PONTUAÇÃO CONSIDERANDO SEPARADAMENTE CADA TIPO DE VISUALIZAÇÃO

Inicialmente ajustou-se modelos de regressão beta considerando a variável resposta parcial, os escores de pontuação separados por cada tipo de visualização. Utilizamos como covariadas as questões relacionadas com as visualizações, ou seja as questões QA1, QA2, QA3,... QA7. O objetivo desta primeira modelagem é identificar se existe diferença significativa entre as visualizações do tipo I, II e III. Para tanto, investigamos quatro modelos de regressão beta apresentados nas expressões (6.2), (6.3) (6.4) e (6.4).

No Modelo 1 as covariadas IQ1 e IQ3 estão relacionadas com os tipos de visualizações I e III, respectivamente. Essas visualizações estão sendo compradas com a visualização do

tipo II. Na tabela 6, apresentamos o Modelo 1 estimado e os respectivos pValores.

Modelo 1 :

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 IQ1_i + \beta_3 IQ3_i + \beta_4 QA2_i + \beta_5 QB1_i + \beta_6 QB2_i + \beta_7 QB3_i + \beta_8 QB4_i + \beta_9 QB5_i + \beta_{10} QC1_i + \beta_{11} QC2_i + \beta_{12} QC2_i + \beta_{13} QC4_i + \beta_{14} QC5_i, \quad (6.2)$$

e

$$\log(\phi_i) = \gamma_1 + \gamma_2 IQ1_i, \quad i = 1, \dots, n. \quad i = 1, \dots, n.$$

Como com base na tabela 6, percebe-se que a covariada IQ3 não é significativa considerando qualquer nível de significância plausível, enquanto que a covariada IQ1 mostrou-se significativa ao nível de 10%. Ressaltando que a covariada IQ1 permite, para este modelo, a comparação entre o desempenho das visualizações dos tipos I e II, os resultados implicam que há diferença entre a visualização do tipo I e do tipo II, com superioridade para a visualização do tipo I, dado que a estimativa do coeficiente é positiva.

Por outro lado, não há diferença entre as visualizações dos tipos 2 e 3, já que, IQ3 permite para este modelo a comparação entre o desempenho das visualizações dos tipos III e II. Sendo assim, consideramos um modelo para avaliar isoladamente o efeito da visualização do tipo I, a saber: Modelo 2 (6.3). Os resultados do processo de estimação e inferência (p-Valores) do Modelo 2 também estão apresentados na tabela 6.

Modelo 2 :

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_1 + \beta_2 IQ1_i + \beta_4 QA2_i + \beta_5 QB1_i + \beta_6 QB2_i + \beta_7 QB3_i + \beta_8 QB4_i + \beta_9 QB5_i + \beta_{10} QC1_i + \beta_{11} QC2_i + \beta_{12} QC2_i + \beta_{13} QC4_i + \beta_{14} QC5_i, \quad (6.3)$$

e

$$\log(\phi_i) = \gamma_1 + \gamma_2 IQ1_i, \quad i = 1, \dots, n. \quad i = 1, \dots, n.$$

Com base nesta tabela, segue que a estimativa do coeficiente da covariada IQ1 é igual a $\hat{\beta}_2 = 0.1616$, com $p_{Valor} = 0.0367 < 0.05$. Como a covariada IQ1 é significativa ao nível de 5%, temos que o tipo de visualização I tem um desempenho diferenciado das visualizações dos tipos II e III e superior, dado a estimativa positiva desta covariada.

De fato, podemos dizer que da visualização do tipo I apresenta desempenho superior de $[\exp(0.1616) - 1.0] \times 100 = 18\%$ em relação aos demais tipos de visualizações, no sentido que este tipo de visualização conduziu à respostas de questões interpretativas da respectiva figura mais próximas das respostas esperadas, ou corretas (Q1, Q2, Q3, Q4 e Q5) e com maior facilidade (Q6) e menor influência do tempo (Q7) de compreensão da respectiva figura.

Tabela 6 – Modelos de regressão beta. Modelo 1: comparação entre os efeitos dos três tipos de visualizações. Base de comparação: Visualização tipo II. Modelo 2: Avaliação do efeito isolado da visualização do tipo I.

Modelo 1.				Modelo 2	
Comparando visualizações I, II e III. Base de comparação: Visualização II.				Avaliando o efeito isolado da visualização I.	
Parâmetros	Descrição	Estimativas	p_Valor	Estimativas	p_Valor
β_1	Constante	-1.8916	0.0000***	-1.8970	0.0000 ***
β_2	IQ1	0.1555	0.0653	0.1616	0.0367 *
β_3	IQ3	-0.0116	0.8636	-	-
β_4	QA2	0.2131	0.0007***	0.2133	0.0007 ***
β_5	QB1	0.3567	0.0000***	0.3569	0.0000 ***
β_6	QB2	0.4334	0.0000***	0.4341	0.0000 ***
β_7	QB3	0.3414	0.0002***	0.3417	0.0002 ***
β_8	QB4	0.2692	0.0001***	0.2692	0.0001 ***
β_9	QB5	0.4360	0.0000***	0.4361	0.0000 ***
β_{10}	QC1	0.3810	0.0000***	0.3803	0.0000 ***
β_{11}	QC2	0.3937	0.0000***	0.3928	0.0000 ***
β_{12}	QC3	0.3225	0.0000***	0.3221	0.0000 ***
β_{13}	QC4	0.3649	0.0000***	0.3648	0.0000 ***
β_{14}	QC5	0.4027	0.0000***	0.4025	0.0000 ***
γ_1	Constante	2.2542	0.0000***	2.2537	0.0000 ***
γ_2	IQ1	-0.9151	0.0000***	-0.9143	0.0000 ***

· significante a 10%. * significante a 5%. ** significante a 1%.
*** significante a mais de 1%

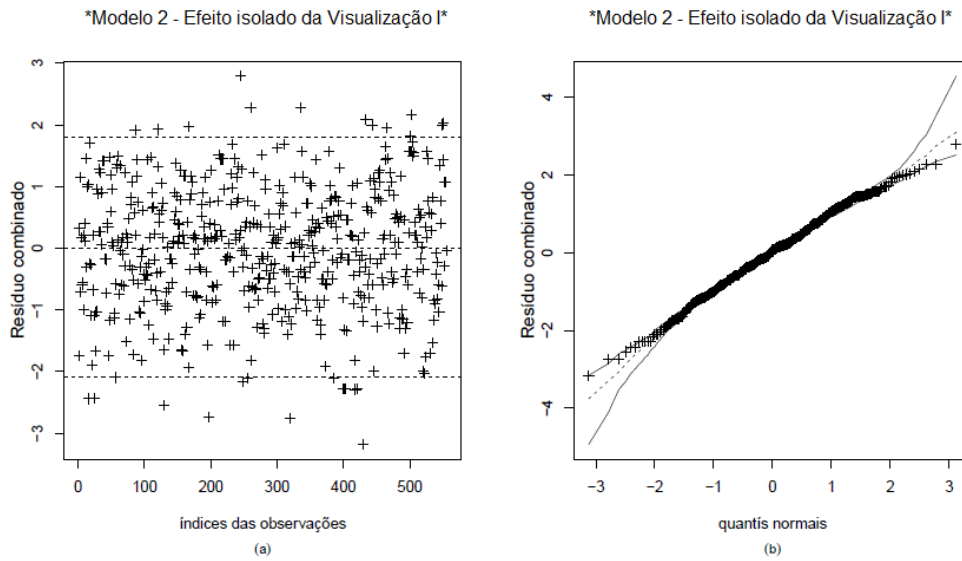
Fonte: Elaborado pela autora.

Para avaliar a qualidade da regressão beta referente ao Modelo 2 **estimado** construímos o gráfico de resíduos contra os índices das observações Figura 18(a) e o gráfico normal de probabilidades com envelopes simulados Figura 18(b). Analisando estes gráficos podemos concluir que Modelo 2 está adequado aos dados. Esta conclusão decorre do fato que os resíduos estão aleatoriamente distribuídos em torno de zero Figura 18(a), ou seja o resíduo está tendendo a zero, implicando que o modelo consegue explicar bem os dados e o que não é explicado, o resto, os resíduos, são próximos da nulidade.

Adicionalmente, os resíduos também estão aleatoriamente distribuídos dentro do envelope. O envelope é uma simulação do que seria a distribuição real dos dados, quando os resíduos do modelo que estamos supondo, os resíduos calculados com base nas estimativas apresentadas nas tabelas, caem dentro do envelope é um sinal de concordância entre o modelo que propomos e os dados reais.

O próximo passo foi avaliar o desempenho isolado da visualização do tipo II, o que foi feito considerando o Modelo 3. As estimativas do Modelo 3 (6.4) e os respectivos pValores

Figura 17 – Gráficos de resíduos. Modelo 2 estimado.



Fonte: Elaborado pela autora.

estão apresentados na tabela 7.

Modelo 3 :

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) =$$

$$\beta_1 + \beta_2 IQ2_i + \beta_3 QA2_i + \beta_4 QB1_i + \beta_5 QB2_i + \beta_6 QB3_i + \beta_7 QB4_i + \beta_8 QB5_i + \beta_9 QC1_i + \beta_{10} QC2_i + \beta_{11} QC3_i + \beta_{12} QC4_i + \beta_{13} QC5_i,$$

e

$$\log(\phi_i) = \gamma_1 + \gamma_2 IQC2_i + \gamma_3 QC3_i + \gamma_4 QC4_i, \quad i = 1, \dots, n.$$

Com base na tabela 7, notamos que a covariada IQ2, a qual está relacionada com o desempenho da visualização do tipo II, apresenta estimativa do coeficiente negativo, isto implica dizer que em comparação com as visualizações do tipo I e III, o desempenho deste tipo de visualização é inferior. Mas, já comparamos as visualizações dos tipos II e III, que não se mostraram significativamente diferentes. Então, este desempenho inferior da visualização do tipo II é em relação ao tipo I. Avaliamos com base nos gráficos de resíduos (figura 18) a qualidade do Modelo 3 estimado apresentado na tabela 7. A análise da Figura 17(a) revela muitas observações com resíduo alto, por exemplo -4. Isto é um indício que o Modelo 3 poderia ser melhorado. Neste sentido, consideramos o Modelo 4.

Assim, finalmente consideramos um modelo em que a visualização do tipo I é a base de comparação com os demais tipos. Para isto construímos o Modelo 4 (6.4), o qual considera as covariadas IQ2 e IQ3. Na tabela 7, apresentamos o Modelo 4 estimado e os respectivos p_Valores.

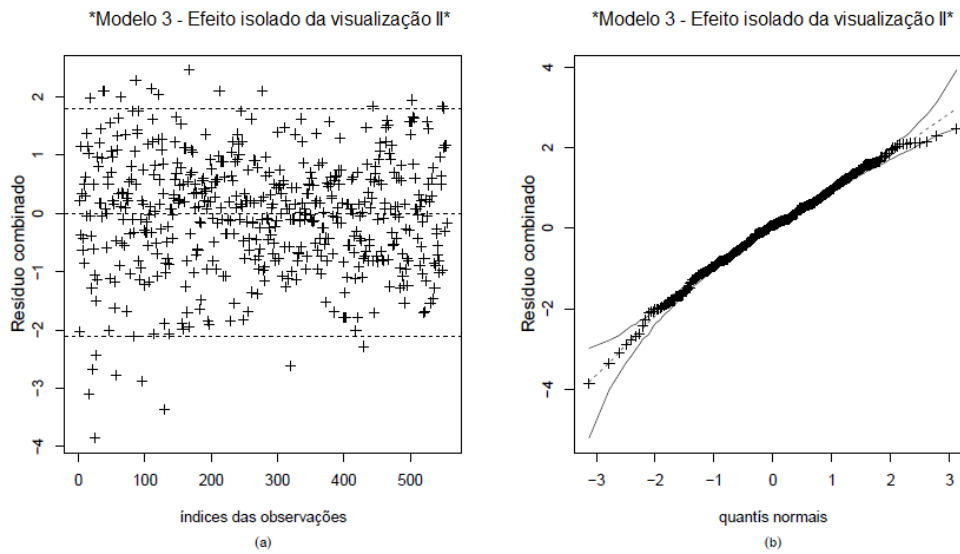
Tabela 7 – Modelos de regressão beta. Modelo 3: Avaliação do efeito isolado da visualização do tipo II. Modelo 4: comparação entre os efeitos dos três tipos de visualizações. Base de comparação: Visualização do tipo I.

Modelo 3.				Modelo 4			
Avaliando o efeito isolado da visualização II.				Comparando visualizações I, III e III. Base de comparação: visualização I.			
Parâmetros	Descrição	Estimativas	p_Valor	Parâmetros	Descrição	Estimativas	p_Valor
β_1	Intercepto	-1.5264	0.0000***	β_1	Intercepto	-1.1419	0.0000 ***
β_2	IQ2	-0.1290	0.0524 ·	β_2	IQ2	-0.6123	0.0000 ***
β_3	QA2	0.4318	0.0000 ***	β_3	IQ3	-0.5296	0.0000 ***
β_4	QB1	0.2779	0.0002 ***	β_4	QB1	0.4320	0.0000 ***
β_5	QB2	0.3627	0.0000 ***	β_5	QB2	0.4986	0.0000 ***
β_6	QB3	0.2592	0.0078 ***	β_6	QB3	0.5172	0.0000 ***
β_7	QB4	0.1690	0.0237 *	β_7	QB4	0.5666	0.0000 ***
β_8	QB5	0.4478	0.0000 ***	β_8	QC1	0.6578	0.0000 ***
β_9	QC1	0.2709	0.0034 **	β_9	QC2	0.4640	0.0000 ***
β_{10}	QC2	0.2702	0.0009***	β_{10}	QC3	0.6840	0.0000 ***
β_{11}	QC3	0.2355	0.0021**	γ_1	Intercepto	2.0368	0.0000 ***
β_{12}	QC4	0.3321	0.0000***	γ_2	QB2	-0.3028	0.0830 ·
β_{13}	QC5	0.3805	0.0000***	γ_3	QC2	-0.4605	0.0035 *
γ_1	Constante	1.5022	0.0000***	γ_4	QC3	0.5575	0.0001 **
γ_2	QC2	-0.3972	0.0019 *				
γ_3	QC3	0.2573	0.0451 *				
γ_4	QC4	0.3057	0.0202 *				
γ_5	QC5	0.2478	0.0576 ·				

· significante a 10%. * significante a 5%. ** significante a 1%.
*** significante a mais de 1%

Fonte: Elaborado pela autora.

Figura 18 – Gráficos de resíduos. Modelo 3 estimado.



Fonte: Elaborado pela autora.

Modelo 4 :

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) =$$

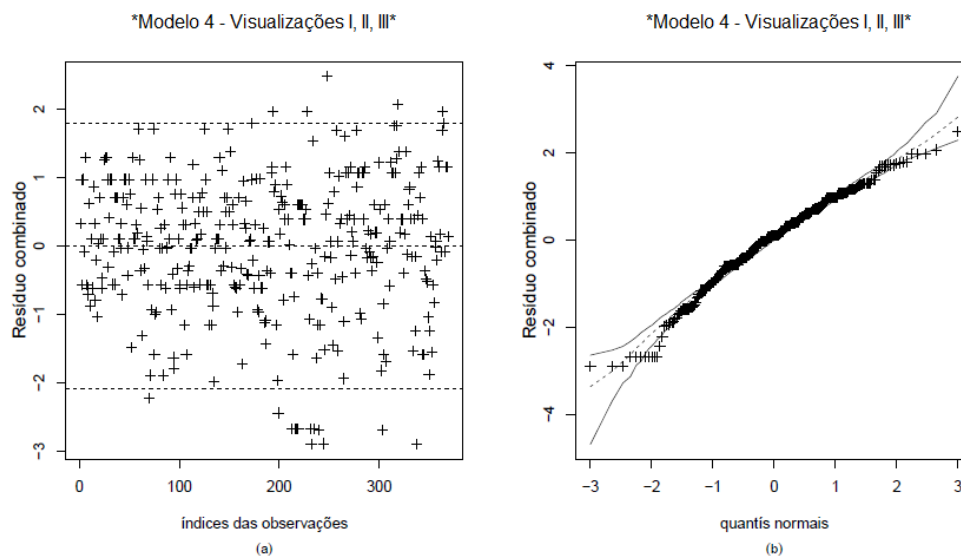
$$\beta_1 + \beta_2 IQ2_i + \beta_3 IQ3_i + \beta_4 QB1_i + \beta_5 QB2_i + \beta_6 QB3_i + \beta_7 QB4_i + \beta_8 QC1_i + \beta_9 QC2_i + \beta_{10} QC3_i$$

e

$$\log(\phi_i) = \gamma_1 + \gamma_2 QB2_i + \gamma_3 QC2_i + \gamma_4 QC3_i, \quad i = 1, \dots, n.$$

Notamos com base na Tabela 7. que as estimativas dos coeficientes das variáveis IQ2 e IQ3 são, respectivamente, $\hat{\beta}_2 = -0.6123$ e $\hat{\beta}_3 = -0.5296$. Isto implica dizer que, a visualização do tipo II apresenta desempenho geral $[1.0 - \exp(-0.6123)] \times 100 = 46\%$ inferior que a do tipo I, enquanto o percentual de inferioridade de desempenho da visualização do tipo III com relação à de tipo I é igual a $[1.0 - \exp(-0.5296)] \times 100 = 42\%$. Para validar a conclusão que a visualização do tipo I tem desempenho superior aos demais tipos realizamos a análise de resíduos do Modelo 4 estimado (Figura 19). Como os resíduos estão aleatoriamente distribuídos em torno de zero e a em sua maioria encontram-se dentro das bandas do envelope, podemos concluir que o Modelo 4 está adequado aos dados, o que valida as conclusões acerca dos desempenhos dos três tipos de visualizações.

Figura 19 – Gráficos de resíduos. Modelo 4 estimado.



Fonte: Elaborado pela autora.

6.5.1 Perfil do avaliador e escores de pontuação dos três tipos de visualizações

Nesta etapa vamos avaliar como as variáveis relacionadas com o perfil dos indivíduos que responderam ao survey influenciou nas assertivas acima. Para alcançar tal objetivo consideramos o Modelo 5. Mais uma vez a variável resposta é o escore de pontuação dos três tipos de visualizações.

Investigamos diversa e intensivamente todas as variáveis de perfil do indivíduo tentando construir um modelo de regressão beta capaz de associar o desempenho do tipo de visualização com as características de seus avaliadores. Finalmente, os Modelos 5, 6 e 7 são os finalistas, respectivamente, para as visualizações dos tipos I, II e III. As versões estimadas destes modelos com os respectivos p-Valores estão apresentadas na tabela 8.

A princípio atentamos para o fato que se o sinal da estimativa do parâmetro é negativo existe uma relação inversa entre o valor da covariada e o escore atribuído ao tipo de visualização. Ocorre o contrário quando o referido sinal é positivo. Adicionalmente, os valores das estimativas dos parâmetros também são importantes, tal que quanto maior esses valores maior é o impacto negativo ou positivo no escore de pontuação do tipo de visualização.

Inicialmente consideremos a visualização do tipo I. Vemos, com base na tabela 8, que a estimativa do coeficiente da variável sexo é negativo, esta variável assume valores 0, caso o entrevistado seja do sexo masculino e 1, caso seja do sexo feminino. Como o sinal negativo implica que quando aumentamos o valor da covariada o escore de pontuação cai, e dado que,

Tabela 8 – Modelo de regressão beta. Modelos 5, 6 e 7: Avaliação das qualificações atribuídas aos três tipos de visualizações de acordo com o perfil do entrevistado.

Efeito do perfil do usuário no escore parcial de pontuação - Segundo tipo de Visualização											
Visualização do tipo I				Visualização do tipo II				Visualização do tipo III			
Modelo 5				Modelo 6				Modelo 7			
Parâmetros	Descrição	Estimativas	p_Valor	Par.	Desc.	Est.	p_Valor	Par.	Descrição	Est.	p_Valor
β_1	Constante	-1.61	0.000***	β_1	Constante	-1.39	0.000***	β_1	Constante	-1.73	0.000***
β_2	Sexo	-0.13	0.011**	β_2	Sexo	-0.15	0.011*	β_2	Sexo	0.07	0.059
β_3	Exp4	0.09	0.093	β_3	Esc1	-0.14	0.047*	β_3	Exp1	0.13	0.070
β_4	Esc2	0.13	0.024*	β_4	Esc2	0.22	0.002***	β_4	Esc1	-0.10	0.037*
β_5	Form2	0.13	0.022*	β_5	Exp3	-0.08	0.290*	β_5	Esc2	0.09	0.046*
β_6	Form5	0.18	0.002***	β_6	Exp4	0.17	0.007	β_6	CurSup1	0.10	0.008***
β_7	Dalt1	0.50	0.005***	β_7	Dalt1	0.58	0.005***	β_7	Form6	-0.07	0.099
β_8	CurSup3	0.26	0.016*	γ_1	Form2	0.25	0.000***	β_8	QC1	0.43	0.000***
β_9	QC2	-0.14	0.006***	γ_2	CurSup2	0.15	0.012*	β_9	QC2	0.43	0.000***
β_{10}	QA1	0.58	0.000***	γ_3	QB1	0.52	0.000***	β_{10}	QC3	0.48	0.000***
β_{11}	QA2	0.34	0.000***	γ_4	QB2	0.48	0.000***	β_{11}	QC4	0.44	0.000***
γ_1	QA3	0.24	0.000***	γ_5	QB3	0.60	0.000***	β_{12}	QC6	0.01	0.000***
γ_2	QA5	0.41	0.000***		Constante	3.45	0.000***	β_{13}	Q6	0.00	0.000***
γ_3	Constante	4.63	0.000***		QC1	0.65	0.009***	β_{14}	QA7	0.00	0.000***
γ_4	QA7	-0.02	0.000***		Exp2	-0.57	0.007***	γ_1	Constante	3.30	0.000***
γ_5	CurSup3	0.90	0.100		Form5	-0.65	0.005***	γ_2	QC5	1.34	0.000***
γ_6	QC1	-0.82	0.001***		Exp1	0.83	0.039*				
γ_7	CurSup1	0.52	0.012*								

· significante a 10%. * significante a 5%. ** significante a 1%.
*** significante a mais de 1%

Fonte: Elaborado pela autora.

neste caso aumentar o valor da covariada é passar do valor 0 (sexo masculino) para o valor 1 (sexo feminino), conclui-se para entrevistados do sexo feminino a pontuação, a compreensão da visualização do tipo I é $[1.0 - \exp(-0.21)] * 100 = 12\%$ mais difícil ou menos favorável, que para o sexo masculino.

Assim, para formar o perfil do indivíduo de acordo com cada tipo de visualização deve-se analisar as estimativas β 's que são os coeficientes das variáveis do modelo que explica os escores de pontuação. Além e analisar o sinal estimativa é necessário avaliar o p-Valor. Quanto menor o p-Valor mais confiança temos na conclusão que a variável de fato deve entrar no modelo. As estimativas dos γ 's estão associados a um modelo auxiliar que determina a variância dos dados, mas que não tem interpretação e não interferem na conclusão do modelo principias definido pelos β 's e respectivas covariadas.

Com estas considerações e com base na tabela 8, podemos formar o perfil do indivíduo de acordo com cada tipo de visualização:

• **Visualização tipo I:** Sexo: masculino. Tempo de experiência: mais de 10 anos. Ensino Médio: instituição pública. Curso Técnico ou profissionalizante: instituição pública. Ensino Superior: instituição pública. Número visto na figura de teste de Daltonismo: visualizou o

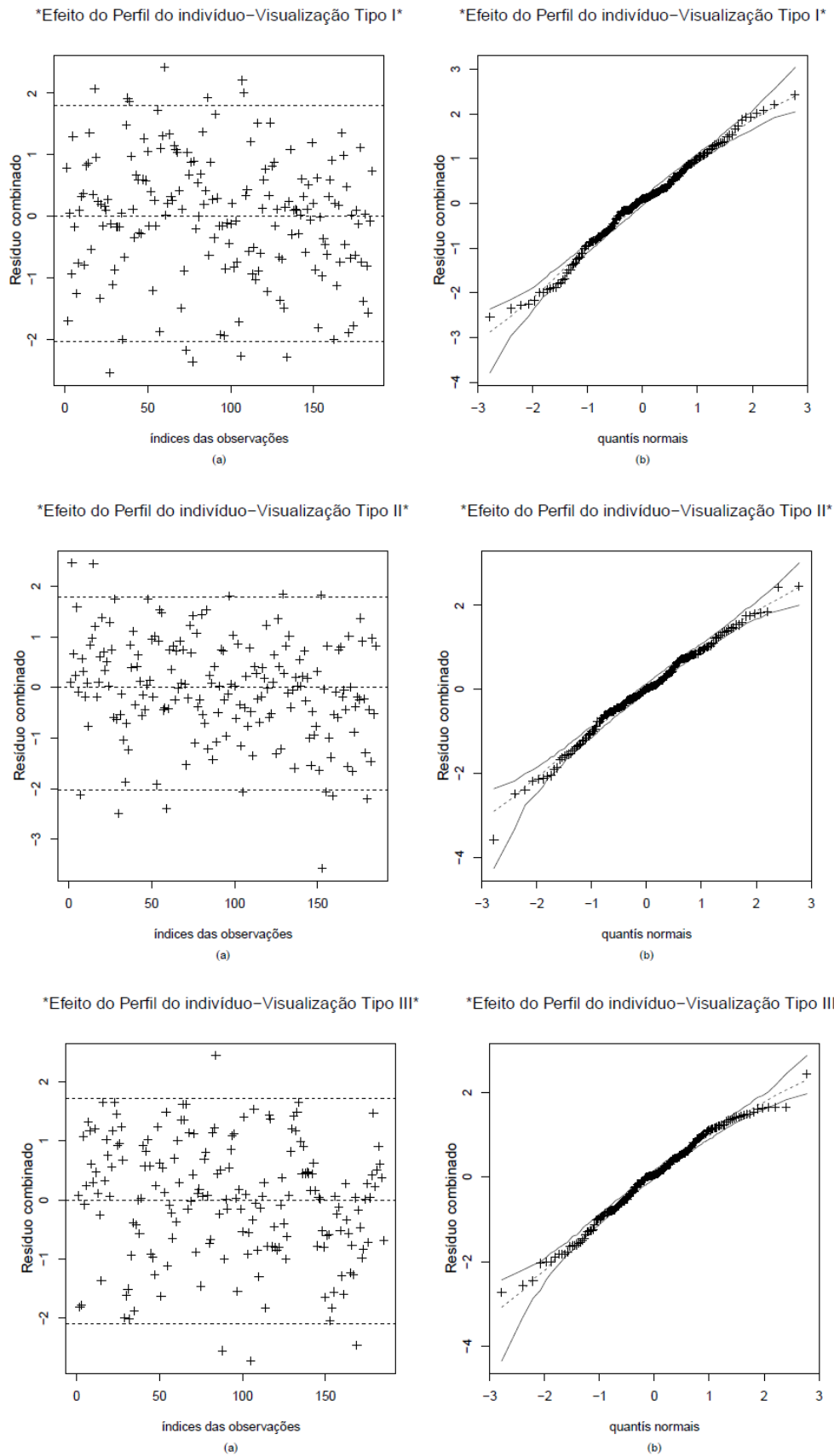
número oito - não daltônico. Curso Superior: ciências exatas.

• **Visualização do tipo II:** Sexo: masculino. Ensino Fundamental: instituição pública. Ensino Médio: instituição privada. Tempo de experiência: mais de 10 anos. Número visto na figura de teste de Daltonismo: visualizou o número oito - não daltônico. Curso Técnico ou Profissionalizante: instituição pública. Curso Superior: ciências humanas.

• **Visualização do tipo III:** Sexo: feminino. Tempo de experiência: menos de 1 ano. Ensino Fundamental: instituição pública. Ensino Médio: instituição privada. Curso Superior: ciências exatas. Ensino superior: instituição privada.

A figura de 20, cujos gráficos de resíduos são apresentados, indicam que os modelos estatísticos representam adequadamente os dados, uma vez que a maioria dos resíduos estão aleatoriamente distribuídos próximo de zero e dentro das bandas do envelope.

Figura 20 – Gráficos de resíduos.



Fonte: Elaborado pela autora.

6.6 MODELO DE CLASSIFICAÇÃO

Após a realização da análise descritiva e dos modelos de regressão beta, com os quais conseguimos formar um perfil dos professores com as variáveis (características) que tem maior impacto na escolha de uma visualização. Nós buscamos classificar todas as variáveis (características) dos professores, relacionando-as umas com as outras, identificando assim que conjuntos de variáveis (características) fazem com que cada professor escolha um determinado tipo de visualização, a fim de se criar uma categorização que aponte o tipo de visualização mais apropriado mediante as suas diversas características pessoais.

O modelo de classificação, então irá modelar todas as informações colhidas do *surveys*, referentes às informações pessoais dos professores e às respostas corretas baseadas nas visualizações. A criação deste modelo inicialmente nos levou a estabelecer algumas etapas. Estas etapas equivalem-se a: (i) preparar dos dados, (ii) criar arquivo contendo a mineração (formato ARFF), (iii) estabelecer o algoritmo de mineração, (iv) minerar os dados utilizando o algoritmo e (v) exibir as informações obtidas.

Durante a preparação dos dados, para a criação da classificação, identificamos a necessidade de tratar *outliers* e valores ausentes, normalizar, agrupar e/ou transformar os dados. Fizemos isso porque alguns tipos de dados podem não ser aceitos, ou diminuem a eficiência dos algoritmos utilizados. Ações pertinentes à etapa (ii) nos levaram a converter os dados que, originalmente estavam no formato XLSX, para o formato CSV, e em seguida convertidos para ARFF. Este último é o formato nativo do WEKA, ferramenta utilizada no processo de mineração.

Tinha-se como objetivo gerar uma classificação, sendo a mesma do tipo árvore de decisão, para que assim, pudéssemos explorar todos os critérios da categorização utilizados pelo WEKA. Para se chegar a classificação, foi utilizando o algoritmo C4, implementado no WEKA através do algoritmo J48. O algoritmo tem os seguintes parâmetros, conforme apresentado na tabela 9

Antes de gerarmos a classificação, com base nos dados e nas variáveis pré-estabelecidas, foram gerados quatro classes de investigação, sendo elas, a classe preferência, a classe percepção, a classe ajuda e classe tempo. Nós tínhamos como pressuposto possibilitar que fossem investigadas quais características pessoais seriam pertinentes, sobre diferentes perspectivas, sendo elas:

- Quando se busca (i) visualizações baseadas nas preferências (gosto dos professores);
- Que (ii) viabilizem uma melhor percepção;
- Que (iii) viabilizem uma melhor ajuda;
- Que (iv) viabilizem uma economia de tempo.

Tabela 9 – Parâmetros para a criação da classificação

Variável	Valor	Descrição
Algoritmo	J48	Implementação do C4
batchSize	100	O número preferido de exemplos para processar se a previsão do lote estiver a ser realizada.
binarySplits	True	Realiza a divisão de atributos nominais.
collapseTree	True	Remover as partes produzem erro no treinamento.
ConfidenceFactor	0.25	Fator de confiança. Quanto menor, maior será a poda.
debug	False	informações adicionais sobre a execução do processo de mineração são exibidas no console de saída de dados.
doNotCheckCapabilities	False	As capacidades de classificadores não são verificados antes do classificador é construído.
doNotMakeSplitPointActualValue	False	A possibilidade de o ponto de divisão não ser atribuída a um valor de dados reais.
minNumObj	2	Número mínimo de objetos por folha.
numDecimlPlaces	2	O número de casas decimais a ser usado para a saída de números no modelo.
numFolds	3	Determina a quantidade de dados utilizados para a poda de erros reduzida.
reducedErrorPruning	False	Maximiza poda da arvore de forma a que evite erros de classificação, gerando arvores mais simples.
saveInstanceData	False	Se pretende guardar os dados de treinamento para visualização.
seed	1	número aleatório de seed a ser utilizado.
subtreeRaising	True	Expandi as sub-arvores, se possível.
unpruned	True	Gera arvore sem poda.
useLaplace	False	Suaviza a quantidade de folhas, considerando funções mais gerias.
useMDLcorrection	True	MDL correção é usado quando encontrar divisão no atributos numéricos.

Fonte: Elaborado pela autora.

Conforme isso, temos nas seções seguintes as descrições das classes e suas arvores correspondentes.

6.6.1 Classe Preferência

Para a classe preferência, foi avaliada a nota que o professor deu para cada visualização em relação à ajuda e ao tempo. Investigamos se a visualização o ajudou a responder as questões e se a visualização reduziu o tempo do professor para entender o problema e responder à questão.

Para alcançarmos um valor para tanto, subtraímos da pontuação ajuda, o valor da pontuação tempo, uma vez que ambos eram inversamente proporcionais, em que quanto maior a ajuda, menor seria o tempo gasto. Assim:

- Se o resultado da subtração das respostas que o professor forneceu para a visualização do tipo I em relação à ajuda e ao tempo, for maior que esse mesmo cálculo para visualização do tipo II e para a visualização III, então, a preferência deste professor é equivalente à visualização do tipo I, logo na arvore será representado pelo o valor 1.

- Se o resultado da subtração das respostas que o professor forneceu para a visualização do tipo II em relação à ajuda e ao tempo, for maior que esse mesmo cálculo para visualização do tipo I e para a visualização III, então, a preferência deste professor é equivalente à visualização do tipo II, logo na arvore será representado pelo o valor 2.

- Se o resultado da subtração das respostas que o professor forneceu para a visualização do tipo III em relação à ajuda e ao tempo, for maior que esse mesmo cálculo para visualização do tipo I e para a visualização II, então, a preferência deste professor é equivalente à visualização do tipo III, logo na arvore será representado pelo o valor 3.

- Se o resultado da subtração de algumas as respostas que o professor forneceu para os três tipos de visualizações em relação à ajuda e ao tempo forem iguais, não apresentando nenhum valor superior, então o professor não demonstrou preferência por nenhum tipo de visualização, sendo que qualquer visualização exposta a ele será indiferente, logo na arvore será representado pelo o valor 4.

Como resultado dos parâmetros estabelecidos e da organização da classe preferência, a figura 20, apresenta o modelo de classificação de acordo com as preferências dos professores.

6.6.2 Classe Percepção

Para a classe percepção, foi avaliada a quantidade de acertos do professor, de acordo com cada visualização exposta, levando em consideração a ajuda fornecida pela visualização e o tempo gasto para responder. Para tanto, formamos três subclasses, a percepção I, a percepção II e a percepção III da classe percepção, as quais correspondem à percepção do tipo de visualização I, II e III. Para obtermos os valores destas subclasses, seguimos quatro passos:

- 1º somamos os pontos obtidos pelo professor por cada acerto;
- 2º dividimos a soma dos pontos pela quantidade total de questões;
- 3º multiplicamos o valor obtido no 2º passo, pelo valor da pontuação ajuda que o professor forneceu;
- 4º subtraímos do valor obtido no 3º passo, o valor da pontuação tempo que o professor forneceu.

Após estes passos, obtivemos um valor para cada subclasse. A classe percepção será dada mediante aos valores de suas subclasses, em que:

- Se o valor da subclasse percepção I, for maior que o valor da subclasse percepção II, e maior que o valor da subclasse III, então percepção deste professor é maior na visualização I, logo, na árvore será representado pelo o valor 1.

- Se o valor da subclasse percepção II, for maior que o valor da subclasse percepção I, e maior que o valor da subclasse III, então, a percepção deste professor é maior na visualização II, logo na árvore será representado pelo o valor 2.

- Se o valor da subclasse percepção III, for maior que o valor da subclasse percepção I, e maior que o valor da subclasse II, então, a percepção deste professor é maior na visualização III, logo, na árvore será representado pelo o valor 3.

- Se o valor de alguma destas subclasses da percepção for igual, não apresentado nenhum valor superior, então, este professor não demonstrou uma percepção significativa com nenhum tipo de visualização, visto que, qualquer visualização exposta a ele será indiferente, logo, na árvore será representado pelo o valor 4.

Como resultado dos parâmetros estabelecidos e da organização da classe percepção, a figura 21, apresenta o modelo de classificação de acordo com percepção dos professores.

6.6.3 Classe Ajuda

Para a classe ajuda, foi avaliada a opinião do professor sobre o quanto as visualizações o ajudaram a compreender questões sobre situações a ele expostas. Como cálculo para tanto, foram comparadas as pontuações dadas pelos professores para cada visualização, em que:

- Se o valor da ajuda fornecida pela visualização I, for maior que o valor da ajuda fornecida pela visualização II, e maior que o valor da ajuda fornecida pela visualização III, então, para este professor, a visualização que mais ajudou foi a visualização I, logo, na árvore será representado pelo o valor 1.

- Se o valor da ajuda fornecida pela visualização II, for maior que o valor da ajuda fornecida pela visualização I, e maior que o valor da ajuda fornecida pela visualização III, então, para este professor, a visualização que mais ajudou foi a visualização II, logo, na árvore será representado pelo o valor 2.

- Se o valor da ajuda fornecida pela visualização III, for maior que o valor da ajuda fornecida pela visualização I, e maior que o valor da ajuda fornecida pela visualização II, então, para este professor, a visualização que mais ajudou foi a visualização III, logo, na árvore será representado pelo o valor 3

- Se o valor de ajuda de alguma das visualizações for igual, então, nenhuma visualização demonstrou ajuda significativa a este professor, assim, qualquer visualização exposta a ele será indiferente, logo, na árvore será representado pelo o valor 4.

Como resultado dos parâmetros estabelecidos e da organização da classe ajuda, a figura 22, apresenta o modelo de classificação de acordo com opinião dos professores a respeito ajuda fornecida pelas visualizações.

6.6.4 Classe Tempo

Para classe tempo, foi avaliada a opinião do professor sobre se a visualização reduziu o tempo do professor para entender o problema exposto e responder as questões. Como cálculo para tanto, foram comparadas as pontuações dadas pelos professores para cada visualização, em que:

- Se o valor do tempo gasto respondendo as questões vinculadas à visualização I, for menor que valor do tempo gasto respondendo as questões vinculadas à visualização II, e menor que valor do tempo gasto respondendo as questões vinculadas à visualização III, então, para este professor, a visualização que proporciona responder em menos tempo é visualização I, logo, na árvore será representado pelo o valor 1.

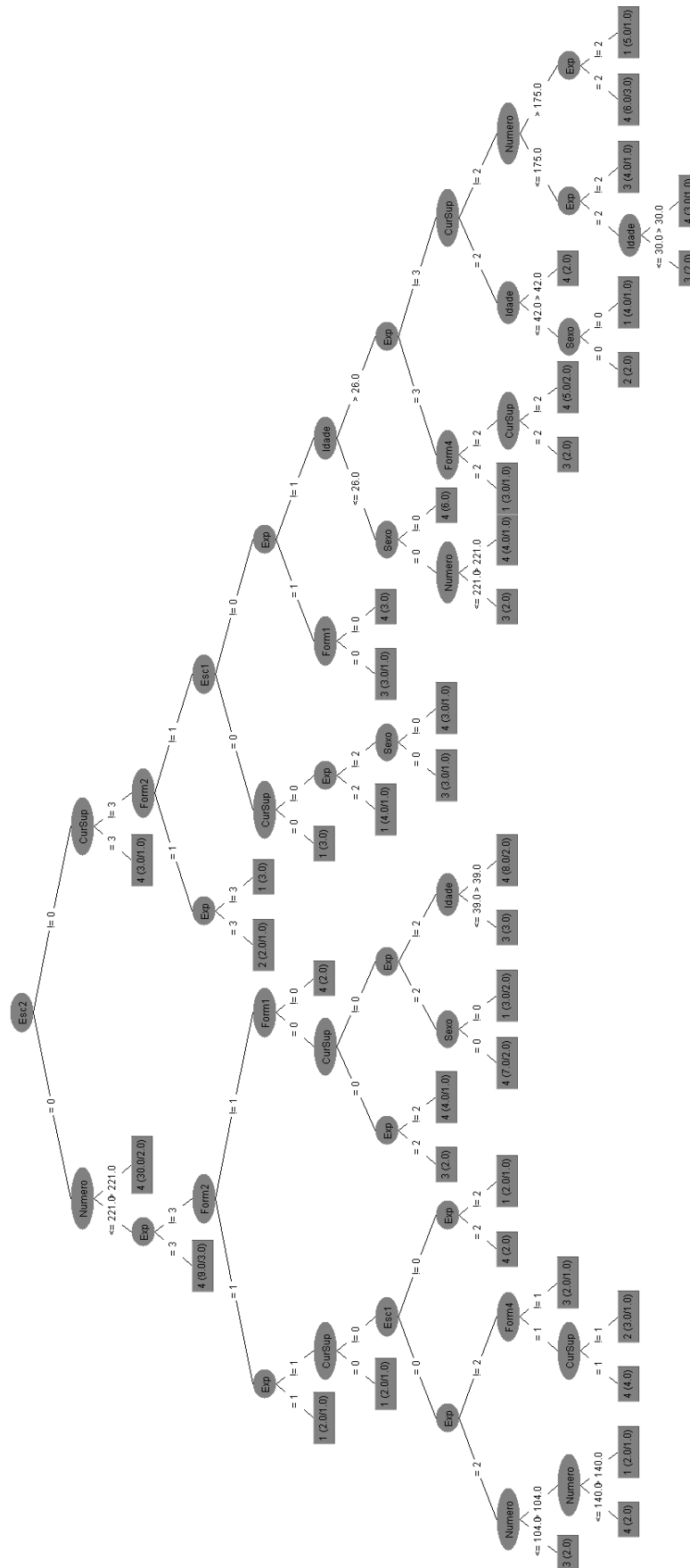
- Se o valor do tempo gasto respondendo as questões vinculadas à visualização II, for menor que valor do tempo gasto respondendo as questões vinculadas à visualização I, e menor que valor do tempo gasto respondendo as questões vinculadas à visualização III, então, para este professor, a visualização que proporciona responder em menos tempo é visualização II, logo, na árvore será representado pelo o valor 2.

- Se o valor do tempo gasto respondendo as questões vinculadas à visualização III, for menor que valor do tempo gasto respondendo as questões vinculadas à visualização I, e menor que valor do tempo gasto respondendo as questões vinculadas à visualização II, então, para este professor, a visualização que proporciona responder em menos tempo é visualização III, logo, na árvore será representado pelo o valor 3.

- Se o valor do tempo gasto respondendo a alguma questão vinculada às visualizações for igual, então, nenhuma visualização demonstrou uma redução de tempo significativa a este professor, uma vez que, qualquer visualização exposta a ele será indiferente, logo, na árvore será representado pelo o valor 4.

Como resultado dos parâmetros estabelecidos e da organização da classe preferência, a figura 23, apresenta o modelo de Classificação de acordo com opinião dos professores em relação ao tempo gasto para analisar as visualizações.

Figura 21 – Arvore de Decisão - Classe Preferência



Fonte: Elaborado pela autora.

7 DISCUSSÕES

O trabalho conduzido investigou a possibilidade de prover auxílio ao professor, ajudando-o a tomar decisões com base em visualizações de dados. Para se prover essa ajuda, buscou-se analisar o professor e suas características para que assim, fosse possível desenvolver a proposta.

Durante essa pesquisa, aplicação do survey e análise dos dados, notou-se que uma das visualizações expostas no questionário apresentado nesse trabalho, visualização tradicional (gráfico de barras), possibilitou um maior número de acertos e eventualmente uma leve preferência dos professores por ela. Isto pode levantar suspeitas de que os professores que a preferiram, estariam condicionados a esta escolha e sua melhor interpretação, uma vez que ela está presente cotidianamente nos livros, jornais e trabalhos acadêmicos.

Algo importante a ser descrito, é que mesmo os professores que acertaram mais com base nas visualizações não tradicionais, eles não compartilhavam de todas as características semelhantes, eles apresentavam características diferentes entre si, como podemos observar na figura 25.

Logo, ainda que superficialmente pode-se inferir que nenhuma visualização é totalmente boa, nem totalmente ruim para diferentes tipos de usuários. A visualização pode ser adequada ou não as características do usuário, dependendo da perspectiva sobre a qual está sendo empregada e analisada.

Outro ponto importante observado sobre as características dos professores, é que foram levantados indícios de que o sexo, a formação acadêmica (os ensinos fundamental, médio, técnico e superior, oferecido por instituição pública ou privada e curso de formação) experiência (anos trabalhados como professor) e fato de ser daltônico ou não, são características que influenciam na preferência entre visualizações tradicionais e não tradicionais e que fortificam a ideia que as visualizações precisam ser personalizadas de acordo com características individuais de seus utilizadores, entretanto, é preciso realizar novos estudos sobre estas características.

Em relação a variável sexo, que se comportou de forma diferente, nos três modelos de perfis, esta pesquisa corrobora e é subsidiada com pesquisas que investigam o tema no campo da percepção visual, como a pesquisa de (SKARF; KUPERSMITH; MCNUSSEN, 1996) que investiga que os cérebros de homens e mulheres são anatomicamente, fisiologicamente e funcionalmente diferentes. Nisto, este trabalho pode trazer novos questionamentos e apontamentos sobre a influência do sexo nas escolhas e preferências de cada ser de forma individual. Possibilitando-nos assim estudarmos com mais profundidade não só a variável sexo, mas também outras características predominantes observadas.

Figura 25 – Características dos professores que escolheram os três diferentes tipos de visualização de dados..



Fonte: Elaborado pela autora

De forma geral, esse trabalho apresenta novos questionamentos sobre quais características ou fatores individuais influenciam na escolha de uma visualização, isto pode abrir novos caminhos para descobertas e alicerces que viabilizem a complementação de trabalhos anteriores e dê novos rumos à área de visualização de dados e áreas correlatas.

7.1 RESPONDENDO AS HIPÓTESES

De acordo com as análises realizadas e a criação dos modelos, buscou-se responder as hipóteses levantadas em nessa dissertação. As primeiras verificações de hipóteses foram feitas com relação às questões de pesquisa 1 e 2, respectivamente, que investigava se o perfil do professor influencia na escolha de uma visualização específica a ser vista e que características podem estar presente na preferência de uma determinada visualização?

Durante as análises realizadas foi possível identificar que a visualização do tipo I (tradicional), apresentou escores de pontuações superiores aos escores relativos os dois outros tipos de visualização (não tradicionais), apesar de todos os escores de pontuações serem muito similares. Isto nos mostrou indícios que de a preferência dos professores eram

diferentes.

Visando confirmar as conjecturas, foram realizadas diversas investigações no intuito de se criar um modelo de regressão beta que fosse capaz de confirmar nossas suspeitas. Dentre as tentativas foi possível chegar aos modelos cinco, seis e sete (tabela 8). Estes modelos expressam as qualificações atribuídas aos três tipos de visualizações de acordo com o perfil do entrevistado. Com isso constou-se, com base nos dados analisados, que os perfis dos professores influenciam na escolha entre visualizações tradicionais e não tradicionais, isso refuta a primeira hipótese, já que a preferência dos professores é diferente para cada visualização, sendo ela tradicional ou não tradicional.

Na segunda questão de pesquisa era dependente da primeira, uma vez que, se de fato o perfil dos professores influencia na escolha da visualização, então quais são as características pertencentes a este perfil. Assim a segunda era: Que características podem estar presente na preferência de uma determinada visualização?

Com base nos os modelos cinco, seis e sete (tabela 8 e figura 25), pôde-se observar que as características dos professores são diferentes para cada tipo de visualização. Sendo que, professores que preferiram a visualização tradicional (tipo I), têm características diferentes dos que preferiram as visualizações não tradicionais (tipo II, III). Desta forma refuta-se a primeira hipótese, pois é possível identificar com base nas análises, que as características dos professores são diferentes na preferência de uma visualização.

Com a refutação da segunda questão de pesquisa, surgia então a terceira e última questão de pesquisa, questiona se é possível fazer uma previsão computacional de preferência de visualização de dados a partir de características.

Mediante a criação dos modelos de classificação, em que é possível identificar que conjuntos de variáveis (características) fazem com que cada professor escolha um determinado tipo de visualização, mostrou-se que é possível prever a preferência dos professores a partir de características.

8 CONCLUSÕES

Esta dissertação teve como objetivo a criação de modelos de personalização de visualização de dados para o professor, uma vez que buscamos criar uma modelagem das características pertinentes aos professores e a forma com o eles interagiram com três tipos de visualização de dados, sendo elas uma tradicional e duas não tradicionais.

Para essa modelagem, é preciso realizar uma adaptação do Processo de Recomendação Pedagógica de (PAIVA; BITTENCOURT; SILVA, 2013) (PAIVA et al., 2014), para que se possa alcançar os objetivos. Nesta adaptação foi criado o Processo de Recomendação de Visualização de Dados para o Professor, este que serviu como guia durante as etapas realizadas.

Para que fosse possível coletar dados dos professores e da interação deles com as visualizações, foi realizada a aplicação de um *survey*. Este *survey* foi criado para suprir as necessidades e perspectivas que surgiram durante a pesquisa. Após a aplicação do *survey*, foram realizadas diversas análises estatísticas e a criação de modelos de regressão, estes que continham quais características tem maior impacto na escolha de uma visualização. Em seguida foi realizada a criação de modelos de classificações, os quais foram identificados assim os conjuntos de características que levam os professores a escolher um determinado tipo de visualização.

Por fim, através deste estudo conseguiu-se demonstrar que as características dos professores influenciam na escolha de uma visualização de dados, que conjunto de características podem estar diretamente relacionadas à escolha, sendo possível classificar as características e a escolha da visualização de dados

8.1 CONTRIBUIÇÕES DO TRABALHO

O presente trabalho contribui com a comunidade de Informática na Educação (i) com indícios de características pessoais podem influenciar na preferência de uma visualização de dados, (ii) com a extração de características de professores para a personalização da tomada de decisão baseada em visualização de dados e (iii) com classificações de conjuntos de características pessoais dos professores personalização da tomada de decisão baseada em visualização de dados sobre diferentes perspectivas.

8.2 PERSPECTIVAS FUTURAS

De uma forma geral pretende-se realizar posteriormente as etapas do processo de recomendação de visualização de dados que não foram aqui realizadas, no intuito de se

aplicar os modelos em ambientes reais e levantar subsídios para a criação de uma ferramenta de recomendação de visualização de dados que seja atuante e efetiva no auxílio os professores.

Pretende-se ainda, estudar de forma mais contundente as características identificadas, uma vez que queremos buscar maiores informações sobre qual o real efeito delas sobre a escolha e preferência entre visualizações, sendo estas tradicionais ou não tradicionais.

REFERÊNCIAS

- ALEXANDER, M.; YUK, M.; DIAMOND, S. **Data visualization for dummies**. [S.l.]: John Wiley & Sons, 2014.
- ALMEIDA, P. M.; SOUZA, T. C. Estimativas de votos de Dilma Rousseff nas eleições presidenciais de 2010 sob o âmbito do Bolsa Família. **Ciencia & Natura**, v. 37, n. 1, p. 12–22, 2015.
- ANNIE, B. P. F. **The teacher as a decision maker**. [S.l.]: Methods for Effective Teaching: Meeting the Needs of All Students, 2014.
- BIENKOWSKI, M.; FENG, M.; MEANS, B. Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. 2012.
- BRASIL. Declaração universal dos direitos humanos. **Adotada e proclamada pela resolução 217 A (III) da Assembleia Geral das Nações Unidas**, 1948.
- BRASIL. Lei de diretrizes e bases da educação: Lei nº 9.394/96 - 24 de dez. 1996. **Estabelece as diretrizes e bases da educação nacional**. Brasília, 1996.
- BRASIL. Educação brasileira: Indicadores e desafios. documento de consulta. **Conferência Nacional de Educação - CONAE. Ministério da Educação - MEC**, 2013.
- CONATI, C.; MACLAREN, H. Exploring the role of individual differences in information visualization. In: ACM. **Proceedings of the working conference on Advanced visual interfaces**. [S.l.], 2008. p. 199–206.
- CURY, A. **Pais brilhantes, Professores fascinantes**. [S.l.]: Editora Sextante / GMT Editores Ltda, 2003.
- DEFANTI, T. A.; BROWN, M. D. Visualization in scientific computing. **Advances in Computers**, v. 33, n. 1, p. 247–305, 1991.
- ESPINHEIRA, P. et al. Resíduos combinados e baseados na diferença. 2015.
- FREITAS, H. et al. O método de pesquisa survey. **Revista de Administração da Universidade de São Paulo**, v. 35, n. 3, 2000.
- FRIENDLY, M.; DENIS, D. J. Milestones in the history of thematic cartography, statistical graphics, and data visualization. **URL [http://www. datavis. ca/milestones](http://www.datavis.ca/milestones)**, 2001.
- FRY, B. **Visualizing data: Exploring and explaining data with the processing environment**. [S.l.]: "O'Reilly Media, Inc.", 2007.
- GARCÍA, E. et al. A collaborative educational association rule mining tool. **The Internet and Higher Education**, Elsevier, v. 14, n. 2, p. 77–88, 2011.
- INEP. Metodologia da concepção do IDEB (nota técnica). **Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira**, 2011.

- INEP. O que é o ideb. **Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira**, <http://portal.inep.gov.br/web/portal-ideb/o-que-e-o-ideb>, 2011.
- JOUNG, Y.; ZARKI, M. E.; JAIN, R. A user model for personalization services. p. 1–6, 2009.
- KHAN, M.; KHAN, S. S. Data and information visualization methods, and interactive mechanisms: A survey. **International Journal of Computer Applications**, Citeseer, v. 34, n. 1, p. 1–14, 2011.
- KHATAEI, A.; LAU, D. Recommender narrative visualization. In: IBM CORP. **Proceedings of the 2013 Conference of the Center for Advanced Studies on Collaborative Research**. [S.l.], 2013. p. 415–421.
- KIRK, A. **Data Visualization: a successful design process**. [S.l.]: Packt Publishing Ltd, 2012.
- KRUM, R. **Cool infographics: Effective communication with data visualization and design**. [S.l.]: John Wiley & Sons, 2013.
- LIU, Z.; STASKO, J. T. Mental models, visual reasoning and interaction in information visualization: A top-down perspective. **Visualization and Computer Graphics, IEEE Transactions on**, IEEE, v. 16, n. 6, p. 999–1008, 2010.
- MADDEN, J. Teachers as researchers: One school's journey thus far. 2014.
- MONTGOMERY, A. L.; SMITH, M. D. Prospects for personalization on the internet. **Journal of Interactive Marketing**, Elsevier, v. 23, n. 2, p. 130–137, 2009.
- NETO, O. Joint-de: Sistema de mapeamento objeto-ontologia com suporte a objetos desconectados. In: . [S.l.: s.n.], 2014.
- OLIVEIRA, D. A. A reestruturação do trabalho docente: precarização e flexibilização. **Educação & Sociedade**, SciELO Brasil, v. 25, n. 89, p. 1127–1144, 2004.
- PAIVA, R. et al. Mineração de dados e a gestão inteligente da aprendizagem: Desafios e direcionamentos. In: **Anais do Workshop de Desafios da Computação Aplicada à Educação**. [S.l.: s.n.], 2012. p. 158–167.
- PAIVA, R.; BITTENCOURT, I. I.; SILVA, A. P. da. Uma ferramenta para recomendação pedagógica baseada em mineração de dados educacionais. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**. [S.l.: s.n.], 2013. v. 2, n. 1.
- PAIVA, R. et al. Lessons learned from an online open course: a brazilian case study. In: ACM. **Proceedings of the 29th Annual ACM Symposium on Applied Computing**. [S.l.], 2014. p. 229–234.
- RIBEIRO, V. G. et al. Uma análise conceitual sobre métodos de pesquisa utilizados em design. **Revista D.: Design, Educação, Sociedade e Sustentabilidade.**, v. 3, n. 1, p. 97–112, 2012.
- RIZZI, E.; GONZALEZ, M.; XIMENES, S. Direito humano à educação. Ação Educativa. Projeto Ação na Justiça, 2009.
- ROBERTS, M.; ZAHAY, D. **Internet marketing: Integrating online and offline strategies**. [S.l.]: Cengage Learning, 2012.

SCHILDKAMP, K.; LAI, M. K.; EARL, L. **Data-based decision making in education: Challenges and opportunities**. [S.l.]: Springer Science & Business Media, 2012.

SIMON, P. **The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions**. [S.l.]: John Wiley & Sons, 2013.

SKARF, B.; KUPERSMITH, M. J.; MCNUSSSEN, P. J. Visual perceptual differences between men and women. 1996.

SMITHSON, M.; VERKUILEN, J. Beta regression: Practical issues in estimation. **Disponível em <http://www.anu.edu.au/psychology/people/smithson/details/betareg/betareg.html>**, 2005.

TOKER, D. et al. Individual user characteristics and information visualization: connecting the dots through eye tracking. In: ACM. **proceedings of the SIGCHI Conference on Human Factors in Computing Systems**. [S.l.], 2013. p. 295–304.

WARD, M. O.; GRINSTEIN, G.; KEIM, D. **Interactive data visualization: foundations, techniques, and applications**. [S.l.]: CRC Press, 2010.

WARE, C. **Information visualization: perception for design**. [S.l.]: Elsevier, 2012.

ZIEMKIEWICZ, C. et al. Understanding visualization by understanding individual users. **Computer Graphics and Applications, IEEE, IEEE**, v. 32, n. 6, p. 88–94, 2012.

Anexos

Questionário de pesquisa em Informática na Educação



A resolução desse questionário o classificará para participação de um sorteio de 15 livros durante o CBIE & LACLO 2015. Para participar, entregue esse questionário respondido até quinta-feira (29/10/15) ao meio-dia ao pessoal de apoio do evento. O sorteio será realizado na sexta-feira (30/10/15).

Os dados desta pesquisa serão mantidos em sigilo, bem como a identificação pessoal dos respondentes; os dados serão usados para análise apenas coletivamente. Os seguintes gráficos mostram a utilização de recursos de aprendizagem, de um ambiente online de aprendizagem, utilizados por 231 estudantes do ensino básico. Desses, 147 estudantes pertencem ao município de São Miguel dos Campos (Grupo 1), Alagoas; 55 estudantes pertencem ao município de São Sebastião (Grupo 2), Alagoas; 29 estudantes pertencem a outros municípios (Grupo 3).

Coletamos os dados para identificar o desempenho desses estudantes, e que recursos de aprendizagem estão influenciando no referido desempenho.

Para o cálculo do desempenho consideramos o total de questões respondidas e a quantidade de questões respondidas corretamente através do ambiente online de aprendizagem. Com o resultado, categorizamos os estudantes, de cada um dos grupos, como: (Vermelho) desempenho inferior; (Amarelo) desempenho médio; (Verde) desempenho superior.

Os recursos de aprendizagem utilizados para a criação dos gráficos, foram: (1) quantidade de vídeos assistidos; (2) quantidade de questões respondidas corretamente; (3) quantidade de questões respondidas incorretamente; e (4) a quantidade de acessos realizados ao ambiente. Consideramos as três categorias (Vermelho, Amarelo e Verde) para cada um dos grupos de estudantes (1, 2 e 3).

Este questionário tem por objetivo descobrir se os gráficos apresentados auxiliam os professores na tomada de decisões pedagógicas que podem melhorar o desempenho de seus estudantes.

Os dados desta pesquisa serão mantidos em sigilo, bem como a identificação pessoal dos respondentes; os dados serão usados para análise apenas coletivamente.

Informações pessoais

Nome: _____

Sexo: () Feminino () Masculino

Idade: _____ anos

A quanto tempo você trabalha ou trabalhou como professor?

() Menos de 1 ano () De 1 a 5 anos () De 6 a 10 anos () Mais de 10 anos

() Não trabalhei como professor

Escolaridade

Você concluiu o ensino fundamental em uma instituição: () Pública () Particular

Você concluiu o ensino médio em uma instituição: () Pública () Particular

Formação

Você está cursando ou concluiu algum curso técnico ou profissionalizante? () Sim () Não

Se sim, qual(is)? _____

Se sim, esse curso é oferecido por uma instituição: () Pública () Privada

Você está cursando ou concluiu o ensino superior em uma instituição:

() Pública () Privada () Não estou cursando e não concluí o ensino superior

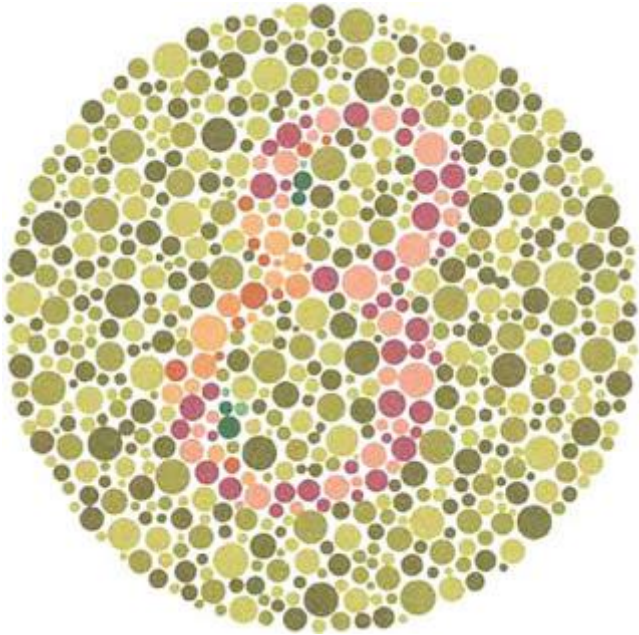
Caso tenha concluído ou esteja cursando, qual(is) curso(s) de ensino superior você está cursando ou concluiu? _____

() Declaro que li os detalhes descritos neste documento. Entendo que eu sou livre para aceitar ou recusar e que eu posso interromper minha participação a qualquer momento. Eu concordo que os dados coletados para o estudo sejam usados para os propósitos acima descritos. Para participar da pesquisa, é necessário que você concorde com o termo de Consentimento Livre e Esclarecido. Você concorda em participar desta pesquisa?

Você é daltônico(a)?

Não Sim

Qual número você enxerga na imagem abaixo?



Número: _____



Você consegue diferenciar as cores abaixo? Se sim, qual o nome de cada uma delas?

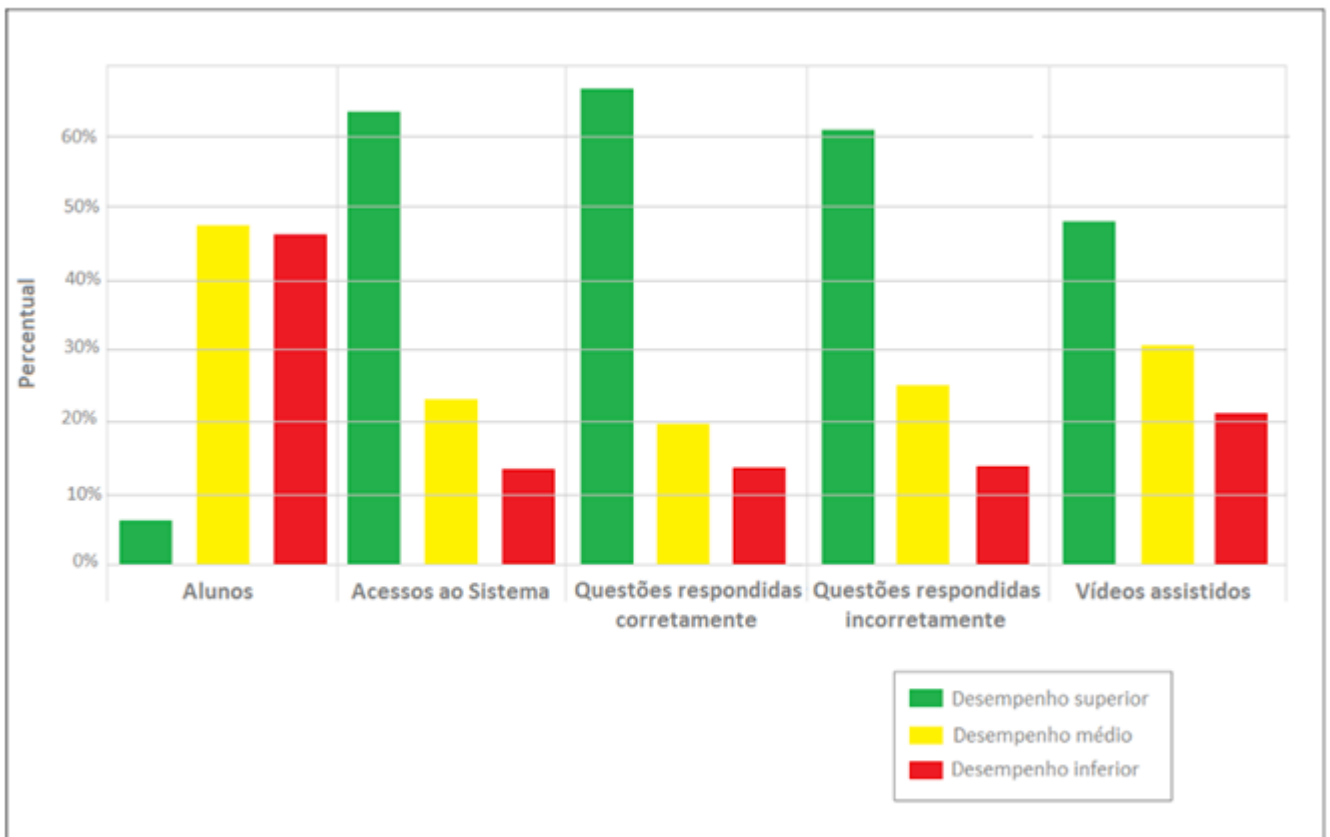
Nome: _____

Nome: _____

Nome: _____

Não consigo diferenciar

Gráficos



Considerando o gráfico acima, assinale a opção que representa sua opinião:

1. Qual recurso de aprendizagem impactou **MAIS** no desempenho dos estudantes?
 Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos
 2. Qual recurso de aprendizagem impactou **MENOS** no desempenho dos estudantes?
 Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos
 3. Para os casos com desempenho **INFERIOR (VERMELHO)**, a interação com qual recurso deveria ser melhorada em primeiro lugar?
 Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos
 4. Para os casos com desempenho **MÉDIO (AMARELO)**, a interação com qual recurso deveria ser melhorada em primeiro lugar?
 Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos
 5. Para os casos com desempenho **SUPERIOR (VERDE)**, a interação com qual recurso deveria ser melhorada em primeiro lugar?
 Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos
- O quanto o gráfico acima o **AJUDOU** a responder as perguntas propostas.
 Muito pouco Pouco Indiferente Muito MUITÍSSIMO
 - Em relação ao **TEMPO** que você gastou para responder as questões de 1 a 5, marque a opção mais apropriada.
 Muito pouco Pouco Indiferente Muito MUITÍSSIMO



Considerando o gráfico acima, assinale a opção que representa sua opinião:

1. Qual recurso de aprendizagem impactou MAIS no desempenho dos estudantes?

Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos

2. Qual recurso de aprendizagem impactou MENOS no desempenho dos estudantes?

Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos

3. Para os casos com desempenho INFERIOR (VERMELHO), a interação com qual recurso deveria ser melhorada em primeiro lugar?

Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos

4. Para os casos com desempenho MÉDIO (AMARELO), a interação com qual recurso deveria ser melhorada em primeiro lugar?

Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos

5. Para os casos com desempenho SUPERIOR (VERDE), a interação com qual recurso deveria ser melhorada em primeiro lugar?

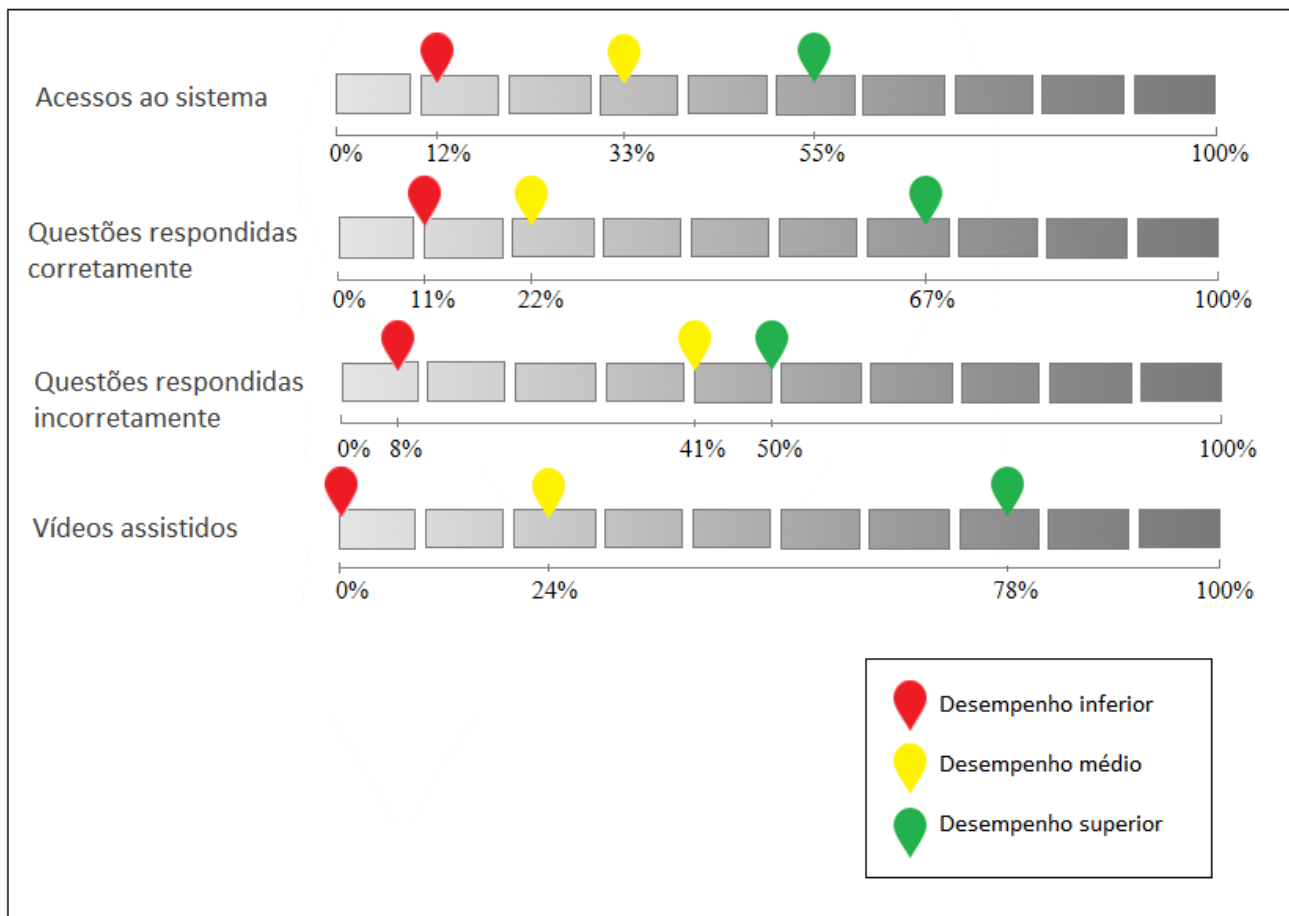
Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos

• **O quanto o gráfico acima o AJUDOU a responder as perguntas propostas.**

Muito pouco Pouco Indiferente Muito MUITÍSSIMO

• **Em relação ao TEMPO que você gastou para responder as questões de 1 a 5, marque a opção mais apropriada.**

Muito pouco Pouco Indiferente Muito MUITÍSSIMO



Considerando o gráfico acima, assinale a opção que representa sua opinião:

1. Qual recurso de aprendizagem impactou **MAIS** no desempenho dos estudantes?
 Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos

 2. Qual recurso de aprendizagem impactou **MENOS** no desempenho dos estudantes?
 Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos
 3. Para os casos com desempenho **INFERIOR (VERMELHO)**, a interação com qual recurso deveria ser melhorada em primeiro lugar?
 Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos
 4. Para os casos com desempenho **MÉDIO (AMARELO)**, a interação com qual recurso deveria ser melhorada em primeiro lugar?
 Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos
 5. Para os casos com desempenho **SUPERIOR (VERDE)**, a interação com qual recurso deveria ser melhorada em primeiro lugar?
 Quantidade de vídeos assistidos Questões corretas Questões incorretas Acessos
- O quanto o gráfico acima o **AJUDOU** a responder as perguntas propostas.
 Muito pouco Pouco Indiferente Muito MUITÍSSIMO
 - Em relação ao **TEMPO** que você gastou para responder as questões de 1 a 5, marque a opção mais apropriada.
 Muito pouco Pouco Indiferente Muito MUITÍSSIMO