

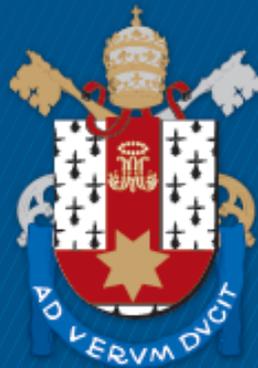
ESCOLA POLITÉCNICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO  
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO

DAIANE RODRIGUES BALDO

***VISUAL ANALYTICS PARA O ACOMPANHAMENTO DE  
MODELOS DE CREDIT SCORING***

Porto Alegre  
2022

PÓS-GRADUAÇÃO - *STRICTO SENSU*



Pontifícia Universidade Católica  
do Rio Grande do Sul

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL  
ESCOLA POLITÉCNICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

***VISUAL ANALYTICS PARA O  
ACOMPANHAMENTO DE  
MODELOS DE CREDIT  
SCORING***

**DAIANE RODRIGUES BALDO**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciência da Computação na Pontifícia Universidade Católica do Rio Grande do Sul.

Orientadora: Prof. Isabel Harb Manssour

**Porto Alegre  
2022**



## Ficha Catalográfica

B178v Baldo, Daiane Rodrigues

Visual analytics para o acompanhamento de modelos de credit scoring / Daiane Rodrigues Baldo. – 2022.

110 f.

Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Computação, PUCRS.

Orientadora: Profa. Dra. Isabel Harb Manssour.

1. Credit Scoring. 2. Visual Analytics. 3. Acompanhamento. 4. Modelagem. I. Manssour, Isabel Harb. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da PUCRS  
com os dados fornecidos pelo(a) autor(a).

Bibliotecária responsável: Clarissa Jesinska Selbach CRB-10/2051



**DAIANE RODRIGUES BALDO**

***VISUAL ANALYTICS PARA O  
ACOMPANHAMENTO DE MODELOS DE CREDIT  
SCORING***

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Ciência da Computação do Programa de Pós-Graduação em Ciência da Computação, Escola Politécnica da Pontifícia Universidade Católica do Rio Grande do Sul.

Aprovado(a) em 21 de Março de 2022.

**BANCA EXAMINADORA:**

Prof. Dra. Maria Cristina Ferreira de Oliveira (ICMC/USP)

Prof. Dr. Afonso Henrique Correa de Sales (PPGCC/PUCRS)

Prof. Isabel Harb Manssour (PPGCC/PUCRS - Orientadora)



## **DEDICATÓRIA**

À memória de minha mãe Maria Jurema Rodrigues que sempre me incentivou e comemorou todas as minhas conquistas.

Uma dedicação especial ao meu marido Vinicius Baldo pela paciência e cuidado nesses 2,5 anos.



“Seja a mudança que você quer ver no mundo.”  
(Mahatma Gandhi)



## AGRADECIMENTOS

À minha orientadora prof.<sup>a</sup> Isabel Harb Manssour pela oportunidade, confiança e paciência durante o projeto. A dedicação que tens com a pesquisa científica é inspiradora.

Ao Prof. Afonso Henrique Correa de Sales por toda colaboração durante o trabalho, fundamental para concretização desta pesquisa.

Ao Programa de Pós-Graduação em Ciência da Computação e a todos os professores incríveis que pude conviver durante esse período.

Aos colegas de mestrado por sempre estarem disponíveis para ajudar, tirar dúvidas e compartilhar seus conhecimentos.

A todos os amigos que me ajudaram e entenderam a minha ausência, especialmente a Mariana Nolde Detoni e Cristiano Sulzbach pela parceria e apoio de sempre.

Um agradecimento especial aos meus sogros Maria Claudete Vicari Baldo e a Celso Roque Baldo por todo apoio.

Agradeço também aos participantes das entrevistas, pois foram fundamentais para a conclusão desta pesquisa.



# **VISUAL ANALYTICS PARA O ACOMPANHAMENTO DE MODELOS DE CREDIT SCORING**

## **RESUMO**

Modelos de *Credit Scoring* são utilizados por instituições financeiras com o objetivo de prever a inadimplência de seus clientes e auxiliar na tomada de decisão sobre a concessão de crédito. Como há um grande volume de transações de crédito sendo gerado diariamente e um potencial aumento dessas informações com o advento do *open finance*, existe o desafio de conseguir monitorar essas informações de forma rápida para que se possa atuar em caso de perda de desempenho desses modelos. Diversos trabalhos encontrados na literatura visam aperfeiçoar as técnicas utilizadas na etapa de construção do modelo. No entanto, não encontramos trabalhos relacionados ao acompanhamento destes modelos. Considerando este contexto, o principal objetivo desta pesquisa foi criar uma abordagem de *Visual Analytics* para auxiliar na gestão de modelos de crédito. Para isto, inicialmente, realizamos uma revisão sistemática da literatura sobre o tema e conduzimos entrevistas semiestruturadas com 13 profissionais que possuem experiência na área. Considerando as necessidades levantadas com este estudo, criamos um protótipo chamado VACS (*Visual Analytics para o Acompanhamento de Modelos de Credit Scoring*). As principais contribuições deste trabalho são: (a) Os resultados obtidos com a revisão sistemática da literatura que mostram que há uma lacuna sobre o tema e que permitiram identificar *insights* sobre o uso de *Visual Analytics* e análise de cenários no acompanhamento destes modelos. (b) O levantamento de requisitos realizado por meio das entrevistas com especialistas, que permitiu o registro de como os modelos são acompanhados dentro das instituições financeiras, algo que não é divulgado devido ao sigilo dos dados e que pode ajudar na padronização dos monitoramentos; (c) O VACS, que foi avaliado por quatro especialistas de domínio que o classificaram como uma ferramenta muito completa e fácil de usar; (d) As

sugestões coletadas na etapa de *feedbacks*, que contribuirão no aprimoramento do VACS e em trabalhos futuros.

**Palavras-Chave:** Credit Scoring, Visual Analytics, Acompanhamento, Modelagem de dados, Risco de Crédito.

# VISUAL ANALYTICS FOR MONITORING CREDIT SCORING MODELS

## ABSTRACT

Financial institutions use credit Scoring models to predict the default of their customers and assist in decision-making about the granting of credit. As there is a large volume of credit transactions being generated daily and a potential increase in this information with the advent of Open Finance, there is the challenge of being able to monitor this information quickly so we can act in case of loss of performance of these models. Several works found in the literature aim to improve the techniques used in the model construction stage. However, we did not find studies related to monitoring these models. Considering this context, the main objective of this research was to create a Visual Analytics approach to assist in the management of credit models. For this, initially, we carried out a systematic review of the literature on the subject and conducted semi-structured interviews with 13 professionals who have experience in the area. Considering the needs raised with this study, we created a prototype called VACS (Visual Analytics for Tracking Credit Scoring Models). The main contributions of this work are: (a) The results obtained from the systematic review of the literature shows that there is a gap on the subject and allowed us to identify insights into the use of visual analytics and scenario analysis in monitoring those models; (b) The survey of requirements carried out through interviews with specialists, which allowed the recording of how the models are monitored within financial institutions, something that is not disclosed and that can help in the standardization; (c) VACS, which was evaluated by four domain experts who rated it as a very complete and easy-to-use tool; (d) The suggestions collected in the feedback stage, which will contribute to the improvement of VACS and future work.

**Keywords:** Credit Scoring, Visual Analytics, Performance, Data Modeling, Credit Risk.

## LISTA DE FIGURAS

Figura 2.1 – Fluxo de concessão de crédito com <i>Credit Scoring</i> [SC10]. . . . .	31
Figura 2.2 – Modelo não linear de um neurônio [Hay07]. . . . .	33
Figura 3.1 – Etapas de pesquisa: Principais atividades realizadas na ordem de execução. . . . .	36
Figura 4.1 – Etapas do processo de revisão sistemática da literatura. . . . .	40
Figura 4.2 – Matriz de confusão: (A) matriz com os erros e acertos totais do modelo [WPB <sup>+</sup> 19]; (B) matriz com os erros e acertos totais do modelo de acordo com a categoria da variável [CHH <sup>+</sup> 19]. . . . .	41
Figura 4.3 – Matriz de confusão temporal: a evolução da matriz de confusão e os dados fluindo entre as diferentes células da matriz [WZWY21]. . . . .	42
Figura 4.4 – Gráfico de barras: (A) barras empilhadas com a ocorrência de erros e acertos de acordo com a faixa de probabilidade gerada pelo modelo; (B) barras empilhadas e não empilhadas com a quantidade de alarmes tratados e falsos positivos [SBMK18]. . . . .	43
Figura 4.5 – Erros de classificação: (A) <i>error dependence plots</i> na visualização do erro do modelo entre as categorias de variáveis na amostra de treino e teste [AT19]; (B) <i>parallel error plots</i> na visualização do erro do modelo entre as variáveis [AT19]. . . . .	43
Figura 4.6 – Análise de justiça: resultados do modelo analisados separadamente para homens e mulheres [WPB <sup>+</sup> 19]. . . . .	44
Figura 4.7 – Análise de justiça: (A) <i>Scores</i> do modelo analisados separadamente para homens e mulheres em um modelo sem alteração; (B) <i>Scores</i> do modelo corrigido para ser justo com homens e mulheres [SL21]. . . . .	45
Figura 4.8 – Análise de variáveis: (A) histograma interativo com distribuição da variável aceleração [CHH <sup>+</sup> 19]; (B) visualização de <i>outliers</i> [JAC15]; (C) probabilidades preditas pelo modelo separadas por sexo [WPB <sup>+</sup> 19]. . . . .	46
Figura 4.9 – Análise de variáveis: o <i>radial chart</i> evidencia a importância e a relação entre variáveis. Cada arco representa o modelo de aprendizado de máquina baseado em cada uma das variáveis (F1, F2, F3, F4). As linhas representam o modelo baseado em múltiplas variáveis [ZHC21]. . . . .	47
Figura 4.10 – Exemplos de acompanhamento: (A) resultados de um ciclo de sono de acordo com cada estágio [GWGv19]; (B) gráficos de linha com limites de controle de acordo com seu desvio [SS14]. . . . .	49
Figura 4.11 – Resultados da RSL. . . . .	51
Figura 5.1 – Perfil dos respondentes. . . . .	55

Figura 5.2 – Informações adicionais sobre ferramentas e dados. . . . .	56
Figura 5.3 – Importância do monitoramento dos modelos nas instituições financeiras. . . . .	60
Figura 5.4 – Preferência pelo monitoramento na mesma interface. . . . .	62
Figura 5.5 – Indicadores estatísticos listados pelos participantes. . . . .	63
Figura 6.1 – Visão geral sobre o uso do VACS no gerenciamento de modelos. . . . .	67
Figura 6.2 – Técnicas de visualização e indicadores levantados na RSL e nas entrevistas. . . . .	69
Figura 6.3 – Layout de entrada de banco de dados sumarizados. . . . .	70
Figura 6.4 – Layout de entrada de banco de dados analíticos. . . . .	70
Figura 6.5 – Resumo dos modelos. . . . .	75
Figura 6.6 – Análise do <i>Score</i> - Distribuição do <i>Score</i> . . . . .	76
Figura 6.7 – Gráfico de KS2 e matriz de Confusão - Curva ROC. . . . .	77
Figura 6.8 – Análise cenários dispersão e análise de tabelas. . . . .	79
Figura 6.9 – Variáveis preditoras - Análise geral. . . . .	80
Figura 6.10 – Variáveis preditoras - Análise geral - Exemplo de interação. . . . .	81
Figura 6.11 – Variáveis preditoras - Análise individual. . . . .	81
Figura 6.12 – Indicadores - Variáveis. . . . .	82
Figura 6.13 – Indicadores - <i>Score</i> . . . . .	83
Figura 6.14 – Catálogo - Variáveis modelo A. . . . .	84
Figura 6.15 – Catálogo - Variáveis modelo B. . . . .	84
Figura 6.16 – Catálogo - Descrição dos indicadores. . . . .	85
Figura 7.1 – Perfil dos respondentes, P1 e P4 participaram do levantamento de requisitos. . . . .	88
Figura C.1 – Termo de consentimento utilizado nas entrevistas - Página 1/2. . . . .	109
Figura C.2 – Termo de consentimento utilizado nas entrevistas - Página 2/2. . . . .	110

## LISTA DE TABELAS

Tabela 4.1 – Palavras-chave por assunto. . . . .	39
Tabela 4.2 – Técnicas de visualização por artigo. . . . .	41
Tabela 5.1 – Empresas que oferecem soluções específicas. . . . .	56
Tabela 5.2 – Número de citações. . . . .	58
Tabela 5.3 – Indicadores no acompanhamento de curto prazo. . . . .	63
Tabela 6.1 – Variáveis do modelo preditivo A. . . . .	72
Tabela 6.2 – Variáveis do modelo preditivo B. . . . .	73
Tabela 6.3 – Situações simuladas nos bancos de dados do modelo A. . . . .	74
Tabela 7.1 – Opiniões sobre utilização do VACS. . . . .	92
Tabela 7.2 – Sugestões para aprimoramento do VACS. . . . .	93



## LISTA DE SIGLAS

AUC – Area Under the Curve

CEP – Comitê de Ética e Saúde

EMA – Exploratory Model Analysis

IFS – Instituições Financeiras

KS – Kolmogorov-Smirnov

MLP – Multilayer Perceptrons

RSL – Revisão Sistemática da Literatura

SBA – U.S. Small Business Administration

SVM – Support Vector Machine

TCLE – Termo de Consentimento Livre e Esclarecido

VA – Visual Analytics



# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>25</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>29</b>
2.1	MODELOS DE <i>CREDIT SCORING</i> .....	29
2.1.1	<i>SCORE</i> DE CRÉDITO .....	30
2.1.2	APLICAÇÃO DO <i>SCORE</i> .....	30
2.2	REGRESSÃO LOGÍSTICA .....	32
2.3	REDES NEURAS ARTIFICIAIS .....	32
<b>3</b>	<b>METODOLOGIA DE PESQUISA</b> .....	<b>35</b>
<b>4</b>	<b>TRABALHOS RELACIONADOS</b> .....	<b>39</b>
4.1	METODOLOGIA .....	39
4.2	FERRAMENTAS E TÉCNICAS DE VISUALIZAÇÃO NO ACOMPANHAMENTO DE MODELOS .....	40
4.2.1	ERROS DE CLASSIFICAÇÃO .....	41
4.2.2	ANÁLISE DE JUSTIÇA .....	44
4.2.3	ANÁLISE DE VARIÁVEIS .....	45
4.2.4	NOVAS FERRAMENTAS .....	46
4.3	MÉTRICAS DE IDENTIFICAÇÃO DE DESVIO .....	47
4.3.1	MONITORAMENTO DE MODELOS DE <i>CREDIT SCORING</i> .....	48
4.3.2	MONITORAMENTO APLICADO A ÁREAS FORA DO CRÉDITO .....	48
4.4	ACOMPANHAMENTO VISUAL NO GERENCIAMENTO DE MODELOS DE CRÉDITO .....	50
4.5	DISCUSSÃO .....	50
<b>5</b>	<b>ENTREVISTAS COM ESPECIALISTAS</b> .....	<b>53</b>
5.1	APLICAÇÃO E ANÁLISE DO QUESTIONÁRIO .....	53
5.2	PERFIL DOS PARTICIPANTES .....	54
5.3	RESULTADO PARA DADOS E FERRAMENTAS .....	55
5.4	RESULTADO PARA TÉCNICAS DE VISUALIZAÇÃO E VA .....	57
5.5	RESULTADO PARA ATIVIDADES NO MONITORAMENTO DE MODELOS DE <i>CREDIT SCORING</i> .....	59

5.6	RESULTADO PARA INDICADORES .....	61
5.7	NECESSIDADES DOS USUÁRIOS .....	63
5.7.1	PRATICIDADE .....	64
5.7.2	FACILIDADE DE USO .....	64
5.7.3	EXPLORAÇÃO DE CENÁRIOS E TESTE DE HIPÓTESES .....	64
5.7.4	EMISSÃO DE ALERTAS .....	65
5.7.5	RESUMO DOS MODELOS .....	65
5.8	CONCLUSÃO DAS ENTREVISTAS .....	65
<b>6</b>	<b>DESCRIÇÃO DO PROTÓTIPO .....</b>	<b>67</b>
6.1	VISUALIZAÇÕES E INDICADORES .....	68
6.2	ENTRADA DE DADOS .....	68
6.3	MODELOS PREDITIVOS .....	71
6.4	FUNCIONALIDADES DO VACS .....	74
6.4.1	RESUMO DOS MODELOS .....	74
6.4.2	MODELOS .....	75
6.4.3	CATÁLOGO .....	83
<b>7</b>	<b>VALIDAÇÃO DO PROTÓTIPO VACS .....</b>	<b>87</b>
7.1	METODOLOGIA E PERFIL DOS PARTICIPANTES .....	87
7.2	RESULTADOS .....	88
7.3	CONTRIBUIÇÕES E LIMITAÇÕES .....	92
<b>8</b>	<b>CONCLUSÃO .....</b>	<b>97</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>99</b>
	<b>APÊNDICE A – Levantamento de Requisitos: Questionário aplicado .....</b>	<b>105</b>
	<b>APÊNDICE B – <i>Feedback</i> VACS: Questionário aplicado .....</b>	<b>107</b>
	<b>APÊNDICE C – Termo de Consentimento Entrevistas .....</b>	<b>109</b>

## 1. INTRODUÇÃO

O crédito no sistema financeiro é o contrato entre consumidores e credores que ocorre devido à confiança entre as partes. O crédito faz parte do dia a dia de praticamente todas as pessoas, sendo um elemento tão difundido e essencial que empresas comerciais deixam de fazer negócios caso não tenham os recursos necessários para aceitar, por exemplo, um cartão de crédito. Cada transação representa para as Instituições Financeiras (IFs) receitas e despesas, impactando diretamente em seus resultados [MESB18].

IFs são responsáveis por conceder crédito a pessoas físicas e jurídicas para a compra de bens e custeio de atividades que não seriam possíveis caso não houvesse esse recurso [MT12]. Nos últimos anos, a concorrência entre IFs que concedem esse crédito cresceu rapidamente e a oferta do crédito passou a ser abundante e mais acessível à população. Em 2017, segundo pesquisa realizada pelo Banco Mundial, 70% dos adultos no Brasil possuíam conta bancária [DKKSA18].

Dessa forma, cresceu também a disputa por clientes, tanto para manter os que já integram a sua carteira de crédito, como também para atrair novos públicos. Ao mesmo tempo, ao intermediar essas transações financeiras e conceder o crédito, torna-se necessário que as IFs façam a gestão do risco inerente a essas operações, pois isso representa a garantia de sustentabilidade do negócio.

A crise mundial ocorrida em 2008, causada pela concessão de crédito hipotecário para credores sem capacidade de pagamento, é uma amostra das consequências de uma gestão de crédito irresponsável [Eva11]. Por isso, órgãos reguladores passaram a ser mais rígidos com relação aos riscos tomados por IFs, inclusive, o 3º acordo de Basiléia, chamado Basiléia III, criado como resposta a crise de 2008, obriga IFs a aumentarem a liquidez e as exigências de capital dos bancos, além de controlarem a exposição a ativos de baixa qualidade [CCDM19]. Para muitos especialistas a crise de 2008 poderia ter sido evitada se a regulamentação financeira fosse mais rígida [AH14].

Nesse cenário, uma ferramenta utilizada para a gestão do risco de crédito é o desenvolvimento de modelos de *Credit Scoring*, que consistem em modelos preditivos que possuem como objetivo mensurar o risco de inadimplência de potenciais contratantes no momento da solicitação do crédito. Estes modelos baseiam-se nas informações sobre o perfil do proponente de crédito que são coletadas no momento da solicitação. A saída dos modelos consiste em um *Score* que mensura a probabilidade de inadimplência futura de cada cliente. Além disso, esses modelos também agilizam a concessão de crédito, já que o cálculo do modelo leva alguns segundos, enquanto uma análise manual faria com que o processo fosse consideravelmente mais demorado [Gon05].

As IFs que concedem crédito utilizam modelos de predição que auxiliam na tomada de decisão [DGP20]. Com o avanço no estudo das técnicas e estratégias de construção

de modelos de *Credit Scoring*, as empresas estão mais conscientes de que uma decisão baseada em dados pode ser mais acurada que uma decisão subjetiva.

Há ainda a grande expectativa da introdução dos dados de *open finance* nesses modelos. O *open finance* é o compartilhamento de informações financeiras e serviços autorizados pelo Banco Central do Brasil entre as instituições [VGB20]. Com isso cresce exponencialmente o número de informações sobre cada cliente que podem ser testadas e usadas no modelo. A expectativa é que esse aumento de informações aumente a concorrência entre as instituições e ajude na inclusão financeira [VGB20].

No entanto, além de construir modelos potencialmente mais precisos que utilizem diversas informações, é necessário acompanhar o desempenho dos mesmos após sua implementação, garantindo que não houveram erros no momento da implantação e que eles seguem calibrados ao longo do tempo [BCGL17]. O desempenho dos modelos, independente da técnica aplicada na construção, pode ser afetado por mudanças nas informações que os compõem, ou mesmo por fatores econômicos externos, como, por exemplo, o aumento do desemprego que irá afetar diretamente os índices de inadimplência das IFs [TH19].

Sendo assim, há a necessidade que esses modelos sejam acompanhados para que no caso de perda de desempenho se possa atuar de maneira rápida. Ao mesmo tempo, há o desafio de minerar os dados utilizados pelos modelos, construindo indicadores de controle de desempenho que auxiliem em seu acompanhamento. Nesse sentido, a área de *Visual Analytics* (VA) desempenha papel relevante na era chamada *data-driven*, quando empresas buscam tomar decisões baseadas em dados [TH19].

A análise baseada em visualizações interativas e mineração de dados permite o entendimento de dados complexos de forma simples e rápida, fornecendo *insights* que auxiliam na tomada de decisão [TMGL20]. Ferramentas como Tableau, Microsoft Power BI e SAS Visual Analytics são utilizadas na construção de *dashboards* que auxiliam no entendimento e comunicação de grandes volumes de dados [BWTK19].

No entanto, entendemos que o acompanhamento de modelos exige o uso de indicadores estatísticos que não são encontrados de forma direta nessas ferramentas. Além disso, existem tópicos distintos dentro destes acompanhamentos, como por exemplo estabilidade das variáveis e avaliação do modelo, que trariam a necessidade da criação de diferentes *dashboards* para cada um deles. Logo, para avaliação dos modelos, seria necessário acessar diferentes locais, o que não é tão prático quando comparado a termos todos eles em uma mesma abordagem. Ao mesmo tempo, algumas ferramentas que utilizam VA na interpretação e investigação de modelos de aprendizado de máquina foram criadas nos últimos anos [WPB<sup>+</sup>19]. O objetivo é que de maneira simples e intuitiva, sem nenhum código necessário, seja possível compreender o que está acontecendo nesses modelos muitas vezes considerados caixa preta.

Portanto, surge a seguinte questão de pesquisa: Como o uso de VA pode contribuir no acompanhamento de modelos de *Credit Scoring*? Visando responder esta pergunta o objetivo deste trabalho é o projeto e desenvolvimento de uma abordagem de VA para o acompanhamento de modelos de *Credit Scoring*.

Para atingir o objetivo proposto, inicialmente foi feita uma revisão sistemática da literatura sobre o tema e entrevistas semiestruturadas com especialistas da área. A partir dos resultados coletados, foi desenvolvido o protótipo VACS (*Visual Analytics* para o Acompanhamento de *Credit Scoring*) que possibilita acompanhar dois modelos de crédito criados previamente. O protótipo utiliza visualizações e interações para: verificar a estabilidade das variáveis contidas no modelo e os indicadores relacionados ao *Score*; apresentar de forma resumida os principais alertas; e fazer a análise de justiça do *Score*. Além disso, todos os menus são interativos, permitindo a seleção do período de análise e mudanças nas principais *features* dos gráficos.

Como as informações sobre os acompanhamentos dos modelos são restritas a cada instituição financeira e não é de conhecimento público, acreditamos que os resultados possam ajudar na padronização sobre a metodologia de acompanhamento, auxiliando principalmente IFs que estejam iniciando o uso de modelos preditivos na gestão de risco de crédito e tomada de decisão. Realizamos ainda uma pesquisa qualitativa com especialistas de domínio, para avaliar todas as funcionalidades criadas no VACS, bem como coletar novas sugestões de aprimoramento para trabalhos futuros.

As principais contribuições deste trabalho são:

- A revisão sistemática da literatura, pois como existia uma lacuna de trabalhos relacionados ao acompanhamento de modelos de *Credit Scoring*, os resultados obtidos permitiram a identificação de *insights* sobre o uso de VA no acompanhamento desses modelos.
- O levantamento de requisitos baseado nas entrevistas com especialistas de domínio, que permitiu detalhar como os modelos são acompanhados nas IFs e identificar as necessidades que os especialistas possuem em uma ferramenta de acompanhamento.
- A construção do protótipo VACS que contemplou as necessidades identificadas através dos trabalhos relacionados e das entrevistas com especialistas.
- A etapa de *feedbacks* com especialistas, pois conseguimos verificar quais foram as necessidades atendidas e levantar uma lista de sugestões para o aprimoramento do VACS.

O restante deste trabalho está organizado da seguinte maneira: no Capítulo 2 introduzimos alguns conceitos sobre crédito, *Scores* de crédito e modelos preditivos. Apresentamos a metodologia de pesquisa que utilizamos no Capítulo 3, e no Capítulo 4 descre-

vemos trabalhos relacionados sobre visualização e métricas aplicadas em modelos preditivos. No Capítulo 5, apresentamos os resultados obtidos nas entrevistas com especialistas e a lista de necessidades levantadas após as entrevistas. Apresentamos todas as funcionalidades e como funciona o VACS no Capítulo 6 e detalhamos os *feedbacks* dos especialistas sobre o protótipo no Capítulo 7. Por fim, no Capítulo 8, apresentamos as principais contribuições da pesquisa e quais são as limitações e os planos para trabalhos futuros.

## 2. FUNDAMENTAÇÃO TEÓRICA

Existem diversos métodos, estatísticos e computacionais, para construção de modelos de *Credit Scoring*, tais como regressão logística, análise discriminante, *support vector machine* e árvores de decisão. No trabalho de Louzada et al. [LAF16] foi feita uma revisão sistemática sobre os métodos utilizados na construção de modelos de *Credit Scoring*, considerando publicações entre 1992 e 2015. Os autores mostram que o número de trabalhos relacionados ao tema cresceu exponencialmente entre 2012 e 2015, sendo que os métodos mais utilizados foram a regressão logística e redes neurais artificiais. Discutiremos neste capítulo alguns conceitos importantes sobre modelos de *Credit Scoring* e exploraremos as duas técnicas mais utilizadas nos últimos anos para este tipo de modelo.

### 2.1 Modelos de *Credit Scoring*

Em um mercado cada vez mais competitivo, IFs precisam se esforçar para atrair novos clientes e também para conseguir atender as expectativas dos já existentes. A oferta de produtos financeiros e serviços aprimorados são fundamentais nesse processo. No entanto, as IFs precisam gerenciar os riscos inerentes à essa oferta de crédito. Sendo assim, a gestão do crédito passa por conseguir distinguir o risco de inadimplência de cada cliente [TCE17].

A decisão sobre a concessão do crédito por muitos anos foi baseada unicamente na opinião de analistas, sendo que a análise de informações dos clientes, preenchidas na solicitação do crédito, era feita manualmente. O crescimento do mercado de cartões de crédito na década de 1960 impulsionou a automatização do processo de decisão, pois não era viável analisar o grande número de solicitações manualmente. A partir dos anos 1980, com o avanço computacional, técnicas estatísticas como regressão logística foram introduzidas na criação de modelos preditivos [Tho00]. O uso de modelos de *Credit Scoring* cresceu nos últimos anos não somente para a oferta do crédito, mas também para o gerenciamento do risco das carteiras de Crédito [LAF16].

Modelos de *Credit Scoring* avaliam o risco que IFs tomam ao conceder crédito aos seus clientes. Tais modelos utilizam dados passados para mensurar a probabilidade de inadimplência futura de cada cliente, também conhecido como “Credit Score” e “probabilidade de *Default*” [TSR+21]. Como todo modelo preditivo, o objetivo é prever um evento futuro (uma variável resposta). O tempo em que se deseja observar o evento varia de acordo com a instituição. Por exemplo, um possível evento de resposta pode ser o cliente ter mais de 90 dias de atraso após 12 meses de observação. Assim, é comum que clientes que

tenham tido o evento de resposta sejam classificados pelo modelo como “1” ou como “mau” e clientes que não tiveram o evento sejam classificados como “0” ou “bom” [GVBB+21a].

### 2.1.1 *Score* de Crédito

Para o cálculo do *Score* são utilizadas diversas informações sobre o cliente que entram como variáveis preditoras no modelo. As fontes de dados são variadas, tais como informações demográficas (por exemplo, CEP, idade, escolaridade), de relacionamento com a instituição financeira (por exemplo, tempo de relacionamento, atraso, histórico de crédito), externas de bureaus de crédito (por exemplo, restritivos, números de consultas do CPF, cadastro positivo), entre outras [Sid12].

Esse *Score*, em conjunto com outras políticas de crédito da instituição, auxilia na tomada de decisão sobre a concessão do crédito ou mesmo sobre a taxa de juros adequada para o risco tomado na operação. Assim, clientes que tenham maior probabilidade de inadimplência terão taxas mais altas que compensem a concessão por parte da instituição. Iniciativas como o *open finance* ajudam a diminuir as taxas ofertadas no mercado, pois quanto mais informações as instituições possuem sobre o cliente que está solicitando o crédito, mais acurado e capaz de encontrar grupos menos arriscados será o modelo [DGR+21].

O ponto de corte do modelo, ou seja, a partir de qual *Score* será concedido ou negado o crédito, varia bastante de acordo com a instituição e o tipo de produto. Por exemplo, para um produto como cheque especial é esperada uma inadimplência alta. Por isso, a taxa é alta e o ponto de corte pode ser menos conservador. Dois tipos de erro podem ser cometidos ao determinar o ponto de corte: conceder o crédito para um cliente que não irá pagar a dívida (I) ou recusar uma operação que seria paga (II). Ao aumentar o ponto de corte, o risco de a instituição cometer o erro II é reduzido (mais conservador). No entanto, ao reduzir o ponto de corte (menos conservador) o risco de ocorrer o erro I aumenta [Sic10].

Apesar de diversas técnicas serem comparadas e aprimoradas, é natural que os modelos preditivos errem. Isso é inerente aos modelos preditivos, ou seja, o modelo pode indicar que o crédito seja concedido a um mau pagador ou que não seja concedido para um bom pagador [SC10]. A Figura 2.1 ilustra o processo de tomada de decisão sobre a concessão de crédito utilizando modelos de *Credit Scoring*.

### 2.1.2 Aplicação do *Score*

Após o desenvolvimento dos modelos de crédito existe a etapa de implantação do mesmo para que ele seja utilizado nas decisões sobre concessão. Logo, o cálculo do modelo final é computado nos motores de decisão da instituição ou mesmo em processos

internos de atualização do *Score*. Ou seja, o modelo passa a ser calculado com base em novas informações e os *Scores* utilizados efetivamente em concessões reais.

É esperado que após a implantação o modelo permaneça com o mesmo desempenho do desenvolvimento. Mas, para avaliar se isso acontece é necessário que as informações utilizadas no cálculo do *Score* sejam armazenadas. Além disso, as informações podem também ajudar a avaliar se ocorreram erros na implantação, o que também impactaria o seu desempenho.

As informações armazenadas são comparadas ao desenvolvimento, para determinar se o modelo segue com um desempenho aderente. Para isso, um ponto importante na avaliação é a manutenção da discriminação nas variáveis preditoras do modelo [Sic10]. É esperado que esse desempenho seja estável, ou seja, siga no mesmo patamar de inadimplência apresentado no desenvolvimento e que não ocorra inversões nas categorias das variáveis. Por exemplo, supondo que uma das variáveis do modelo seja restritivo no mercado, onde, no desenvolvimento do modelo, o grupo com restritivo possuía 15% de inadimplência e o grupo sem restritivo 4%. Esperamos que essa diferença se mantenha após o início da utilização do modelo, pois caso esse comportamento se inverta, ou seja, quem tem restritivo passe a ser melhor, o modelo estaria atribuindo *Scores* equivocados, o que poderia gerar concessões erradas.

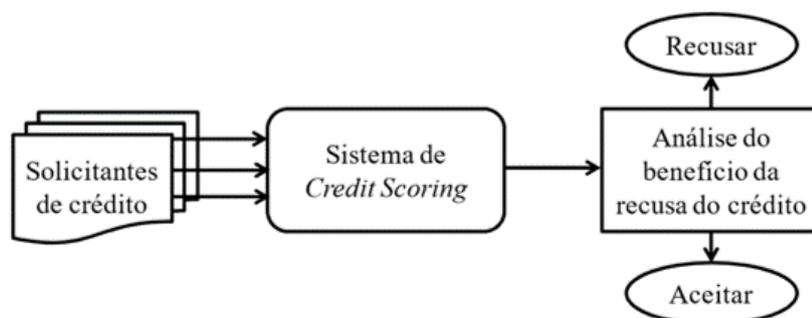


Figura 2.1 – Fluxo de concessão de crédito com *Credit Scoring* [SC10].

Uma das formas de verificar mudanças significativas no *Score* em comparação com o desenvolvimento é analisar se ocorreram mudanças na distribuição após o início da utilização do modelo. A função de densidade de probabilidade (FDP) é uma equação que representa a distribuição de probabilidade de uma variável aleatória contínua [RR11]. Assim, observando a curva de FDP do *Score* conheceremos as regiões de maior e menor probabilidade para os valores de *Score* e poderemos analisar se houve deslocamento da curva ao longo tempo.

## 2.2 Regressão Logística

A regressão logística é uma técnica multivariada adequada para diferentes situações, pois a partir de variáveis explicativas (contínuas ou discretas), é possível analisar o efeito dessas na presença (1) ou ausência (0) de uma característica [HL00]. Assim, por meio do modelo de regressão logística, calculamos a probabilidade de ocorrência de um evento, como mostra a equação 2.1 [HL00].

$$\pi = \frac{e^{(\beta_0 + \sum_{j=1}^k \beta_j X_j)}}{1 + e^{(\beta_0 + \sum_{j=1}^k \beta_j X_j)}}. \quad (2.1)$$

Nesta equação,  $\pi$  é a probabilidade de ocorrência do evento, e  $\beta_j$  são os coeficientes associados a cada variável  $X_j$ .

Os coeficientes de regressão e seus erros padrões são estimados com método de máxima verossimilhança, que maximiza a probabilidade de obter o grupo observado de dados, por meio do modelo estimado. Podemos obter a solução da função por meio de métodos numéricos, como o de Newton-Raphson [CB02], que irão criar uma série de soluções que convergem para o ponto máximo da função de verossimilhança. O modelo de regressão logística deve atender ao pressuposto de baixa correlação entre variáveis preditoras, pois o modelo é sensível a multicolinearidade [Hai11].

## 2.3 Redes Neurais Artificiais

As redes neurais artificiais são constituídas por unidades de processamento que tentam simular o cérebro humano. Baseado no envio de sinais eletrônicos entre os neurônios, ela combina variáveis de entrada com interações lineares e não lineares através de camadas, resultando nas variáveis de saída. As redes neurais se diferenciam em relação à estrutura utilizada, podendo variar o número de camadas ocultas e a função de ativação aplicada.

A rede neural MLP (*Multilayer Perceptrons*) é uma das técnicas comparadas na literatura para a construção de modelos de *Credit Scoring* [NCC18]. A MLP é composta por três tipos de camadas, uma de entrada, uma ou mais camadas escondidas e uma de saída. Os neurônios processam a entrada e geram uma saída que é transmitida para os neurônios das camadas subsequentes [Hay07], como podemos observar na Figura 2.2. Esse valor é computado por uma função de ativação, que decide se o neurônio deve ser ativado ou não, como a função sigmoide. Para casos de classificação binária recomenda-se o uso da função de ativação logística, já que as saídas geram valores entre [0,1], o que permite que sejam interpretadas como uma probabilidade [Hay07]. A função aditiva faz a soma dos

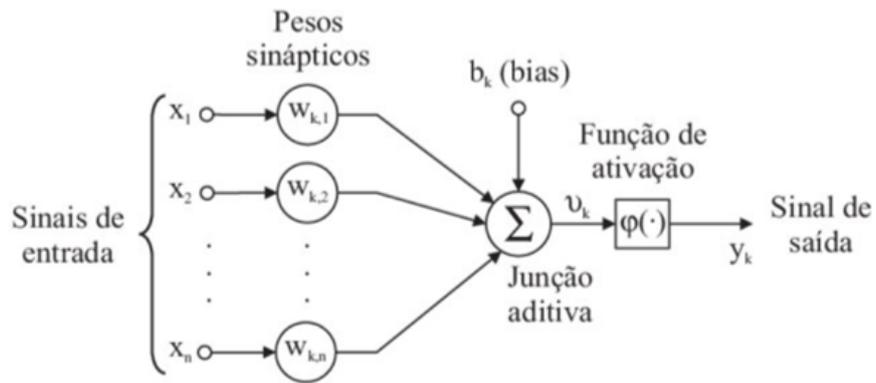


Figura 2.2 – Modelo não linear de um neurônio [Hay07].

sinais de entrada ponderado pelos respectivos pesos e o termo de *bias* é uma constante da função aditiva.



### 3. METODOLOGIA DE PESQUISA

O objetivo desta pesquisa é o projeto e desenvolvimento de uma abordagem de VA que possua os requisitos necessários para um acompanhamento adequado dos modelos de *Credit Scoring*, gerando informações que ajudem a instituição financeira na decisão sobre a manutenção dos modelos utilizados na concessão de crédito de seus clientes. Nesse sentido, construímos um protótipo para a abordagem de VA em Shiny<sup>1</sup>, pacote desenvolvido para usuários do *software* estatístico R<sup>2</sup> que permite criar aplicações interativas em HTML. Escolhemos esse pacote porque ele suporta todos os cálculos estatísticos necessários para o acompanhamento dos modelos.

De maneira a guiar este estudo, definimos a principal questão de pesquisa: Como o uso de VA pode contribuir no acompanhamento de modelos de *Credit Scoring*? Com o intuito de responder essa questão e auxiliar no desenvolvimento do trabalho, também formulamos as seguintes questões secundárias (QS) de pesquisa:

- QS1: Como é feito o acompanhamento visual de modelos de *Credit Scoring* em IFs?
- QS2: Quais ferramentas e técnicas visuais podem ser utilizadas no acompanhamento de modelos?
- QS3: Quais métricas são utilizadas na identificação de possíveis desvios após a implementação de modelos?
- QS4: O acompanhamento visual pode ajudar no gerenciamento dos modelos de crédito?

Para responder estas questões de pesquisa e alcançar o resultado esperado, foi necessário seguir uma série de etapas. A Figura 3 ilustra a metodologia que utilizamos durante a execução da pesquisa. Após a definição do tema, passamos para a elaboração sobre a questão principal de pesquisa, criando também questões secundárias que nos ajudassem a respondê-la. Logo depois, buscamos responder às questões secundárias QS1, QS2, QS3 e QS4, revisando a literatura em busca de trabalhos relacionados que auxiliassem nessa discussão. Para isso, conduzimos uma Revisão Sistemática da Literatura (RSL), através da metodologia proposta por Kitchenham et al. (2007), cujos resultados podem ser verificados no Capítulo 4. Após a aplicação dos critérios de inclusão e exclusão, chegamos a 15 artigos finais que ajudaram a responder as questões QS2, QS3 e QS4. Entretanto, não encontramos trabalhos que respondessem a QS1 sobre como são feitos os acompanhamentos de modelos nas IFs. Acreditamos que isso seja devido ao sigilo nos processos e dados das IFs, o que dificulta o compartilhamento de conhecimento na área.

---

<sup>1</sup><https://shiny.rstudio.com/>

<sup>2</sup><https://www.r-project.org/>

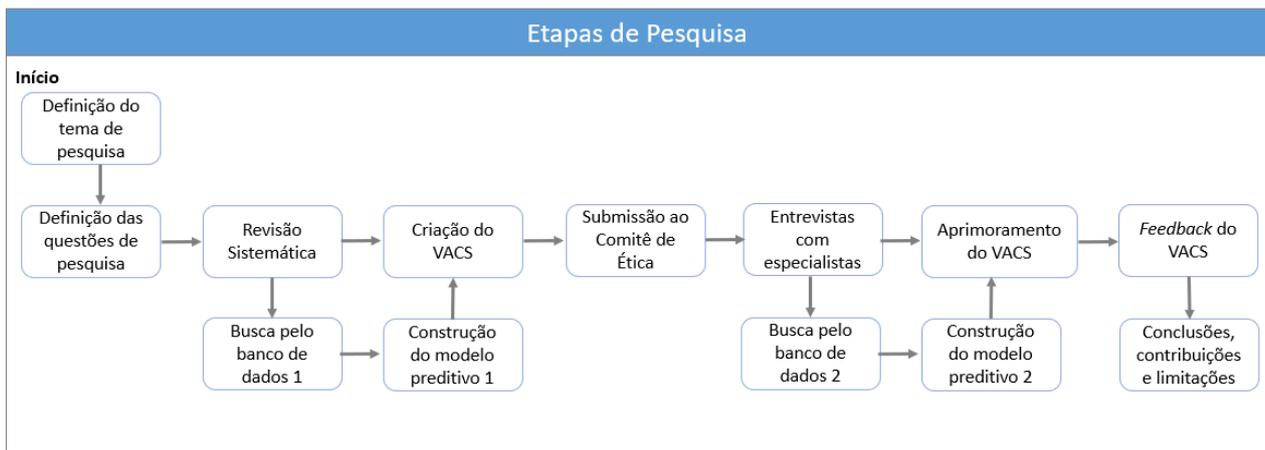


Figura 3.1 – Etapas de pesquisa: Principais atividades realizadas na ordem de execução.

Durante a execução da RSL ficou claro que precisaríamos construir um modelo preditivo para ser utilizado como entrada no protótipo. Assim, em paralelo a revisão sistemática, já iniciamos a busca na literatura por um banco de dados que conseguisse exemplificar dados de concessão de crédito reais e construímos um modelo preditivo de *Credit Scoring*. Após a revisão e a construção do modelo, já foi possível iniciar o desenvolvimento do protótipo.

Para responder a QS1 e com a intenção de aumentar a credibilidade das evidências encontradas, ampliando o conhecimento obtido com a RSL, conduzimos entrevistas semiestruturadas com profissionais que já trabalharam ou trabalham atualmente com modelos de *Credit Scoring*. Para condução das entrevistas, seguimos o protocolo estabelecido pelo Comitê de Ética e Pesquisa e obtivemos a aprovação para a realização da pesquisa, ou seja, para a aplicação do questionário com o intuito de validar o nosso estudo. A documentação completa com o protocolo de pesquisa pode ser encontrada no site da Plataforma Brasil<sup>3</sup>, (número CAAE 37086420.2.0000.5336).

Durante a criação do questionário, levamos em consideração o sigilo sobre os processos das IFs, por isso não fizemos nenhuma questão que fosse capaz de identificar em qual instituição o participante trabalhava. Além disso, customizamos cada questão para que ela focasse nas experiências gerais do participante. Sendo assim, ele responde sobre o que conhece de forma geral, considerando todas as experiências que já teve e sem especificar a qual instituição ele se refere. Caso isso tenha acontecido, não foi incluído nos resultados. Incluímos o questionário aplicado no Apêndice A e o TCLE (Termo de Consentimento Livre e Esclarecido) no Apêndice C.

Com os resultados obtidos nas entrevistas, apresentados no Capítulo 5, foi possível responder a QS1 sobre como são feitos os acompanhamentos de modelos de *Credit Scoring*, bem como aprofundar quais visualizações e funcionalidades os especialistas que

<sup>3</sup><https://plataformabrasil.saude.gov.br/login.jsf>

trabalham na área acreditam serem necessárias em uma abordagem de VA para o acompanhamento de modelos de *Credit Scoring*. Ainda, sintetizamos os resultados das entrevistas em cinco necessidades dos usuários (ver Seção 5.7) que foram usadas no desenvolvimento do protótipo.

Após a pesquisa com especialistas, entendemos que seriam necessários pelo menos dois modelos para que fosse exemplificado como a abordagem de VA funcionaria em situações reais, pois as instituições tem sempre mais de um modelo. Assim, selecionamos mais um banco de dados e construímos um novo modelo preditivo de *Credit Scoring*. Na Seção 6.2 apresentamos os bancos de dados selecionados e na Seção 6.3 explicamos como construímos cada um dos modelos preditivos e as variáveis que os compõem. Ainda com o objetivo de tornar o protótipo próximo de um cenário real, foi necessária uma etapa de simulação utilizando os dados de saída do modelo. Na Seção 6.3 detalhamos quais cenários foram simulados.

A partir dos resultados obtidos na RSL com relação as técnicas de visualizações e indicadores encontrados e das necessidades levantadas nas entrevistas com especialistas, criamos o VACS que utilizou como entrada os resultados de ambos os modelos construídos. Após a etapa de entrevistas, foi possível aprimorar significativamente o protótipo, levando em consideração as sugestões dos especialistas. Na Seção 6.4 apresentamos detalhadamente todas as funcionalidades do protótipo, explicando as funcionalidades, indicadores e visualizações construídos.

Por fim, realizamos ainda uma etapa de avaliação do protótipo. Para isso, elaboramos outro questionário para ser aplicado com especialistas de domínio, para tentar entender se as funcionalidades criadas atendiam as necessidades levantadas na primeira parte das entrevistas. O questionário aplicado encontra-se no Apêndice B e o TCLE no Apêndice C. Como forma de mitigar algum viés na avaliação, não entrevistamos apenas os participantes da primeira etapa das entrevistas, convidamos também pessoas que não participaram dessa primeira etapa, respeitando os mesmos requisitos. No Capítulo 7 apresentamos os resultados obtidos na avaliação do protótipo, bem como as suas contribuições e limitações.



## 4. TRABALHOS RELACIONADOS

Fizemos uma RSL sobre técnicas de visualização utilizadas na modelagem de dados e métricas para o acompanhamento do desempenho de modelos de *Credit Scoring*. Além disso, buscamos também por ferramentas utilizadas nesse acompanhamento. O resultado desta pesquisa é apresentado nas seções a seguir, ressaltando as principais contribuições dos trabalhos relacionados. Também listamos os desafios e considerações identificados durante o processo de revisão.

### 4.1 Metodologia

A Figura 4.1 sintetiza todas as etapas desta RSL. Na primeira etapa do processo de RSL, durante a escolha da *string* de busca, percebemos que não havia trabalhos que tratassem especificamente do acompanhamento visual de modelos de *Credit Scoring*. Porém, identificamos trabalhos que exploravam visualizações durante a construção de modelos preditivos, bem como trabalhos com sugestões de métricas para acompanhamento de modelos de forma geral ou especificamente para modelos de *Credit Scoring*. Por isso, ampliamos a nossa busca para incluir qualquer trabalho relacionado com modelagem no contexto de visualização ou no contexto de acompanhamento.

Optamos pela construção de uma *string* de busca genérica, o que levou a um número elevado de artigos iniciais. As palavras-chave utilizadas podem ser observadas na Tabela 4.1. Devido as diferenças entre os portais de busca, a *string* construída foi personalizada para cada um deles.

Tabela 4.1 – Palavras-chave por assunto.

Assunto	Palavras-Chave
Visualização de Dados	data AND visual*
Modelos	model* OR "machine learning"OR "logistic regression"
Credit Scoring	"credit scor*"OR "credit risk"
Monitoramento	monitoring OR performance OR evaluation
Apenas visualização ou acompanhamento de modelos	NOT("All Metadata":video OR "computer simulation"OR image OR robot* OR "virtual reality"OR "social media"OR health)

Utilizamos as bases *IEEEExplore*, *ACM Digital Library*, *Science Direct* e *Scopus* para extração dos estudos primários. Com relação aos critérios de inclusão utilizados, consideramos os artigos com publicação disponível *online*, escritos em inglês e publicados em conferências e revistas no período entre janeiro de 2011 a dezembro de 2021. Os critérios de exclusão foram aplicados em três momentos: primeiro retiramos os trabalhos

repetidos através do uso da ferramenta StArt (*State of the Art through Systematic Review* [FHDT<sup>+</sup>12]); após a leitura do título, resumo e palavras-chave retiramos os trabalhos sem relação com modelagem de dados no contexto de visualização ou acompanhamento de modelos; por fim, após a leitura na íntegra, retiramos os trabalhos não relacionados com o tema de pesquisa. Depois da aplicação dos critérios de inclusão e exclusão, chegamos a 15 artigos finais.

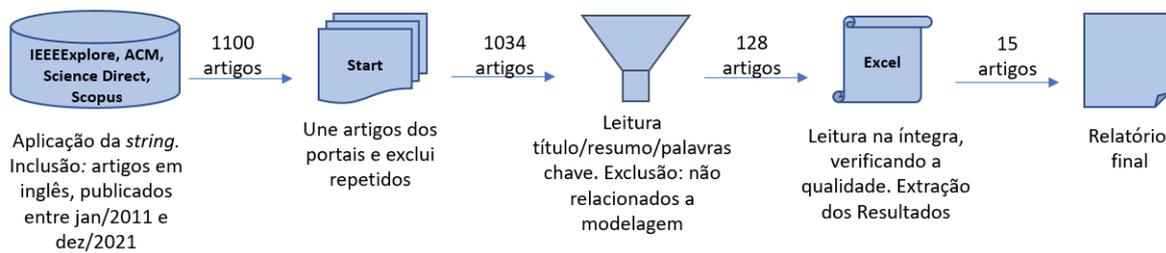


Figura 4.1 – Etapas do processo de revisão sistemática da literatura.

As contribuições dos 15 trabalhos finais estão dispostas nas seguintes seções: 4.2 Ferramentas e técnicas de visualização para o acompanhamento de modelos; 4.3 Métricas de identificação de desvio; 4.4 Acompanhamento visual no gerenciamento de modelos de crédito. No último capítulo, fazemos uma análise sobre os resultados obtidos com o processo de revisão.

## 4.2 Ferramentas e Técnicas de Visualização no Acompanhamento de Modelos

Os artigos selecionados mostram que técnicas de visualização têm sido utilizadas na explicação de modelos de aprendizado de máquina. Como esse tipo de modelo tem a desvantagem de ter difícil explicabilidade, nos últimos anos foram criadas visualizações e novas ferramentas visuais que facilitam seu entendimento [ZHC21]. A maior parte destas ferramentas também pode ser utilizada no acompanhamento de modelos de forma geral, pois não focam em uma técnica de modelagem específica. As visualizações partem da saída gerada pelo modelo, sendo adaptáveis a vários contextos de modelagem [WPB<sup>+</sup>19]. Neste capítulo, analisamos 10 dos trabalhos selecionados na RSL que contribuem nas diferentes etapas da modelagem de dados. Consideramos as contribuições que poderiam ser utilizadas no monitoramento de modelos, separando as visualizações criadas nos seguintes tópicos descritos nas seções a seguir: erros de classificação, análise de justiça e análise de variáveis. Separamos nessas categorias por entender que são visualizações que seriam utilizadas de forma separada em um acompanhamento. Além disso, criamos um tópico sobre as novas ferramentas criadas nesses trabalhos.

A Tabela 4.2 resume as técnicas de visualização utilizadas em cada artigo e quais deles criaram uma nova ferramenta de visualização de dados.

Tabela 4.2 – Técnicas de visualização por artigo.

Referência	Categoria	Técnicas de Visualização	Nova Ferramenta
[ZWM <sup>+</sup> 18]	Erro de classificação, análise de justiça	Matriz de confusão, curva ROC	Não
[LGL <sup>+</sup> 19]	Erro de classificação	Gráfico de área	Sim
[WPB <sup>+</sup> 19]	Erro de classificação, análise de variáveis	Matriz de confusão, curva ROC, gráfico de dispersão	Sim
[SBMK18]	Erro de classificação, análise de variáveis	Gráfico de barras normal e empilhado	Não
[AT19]	Erro de classificação	<i>Error dependence plots, parallel error plots</i>	Não
[CHH <sup>+</sup> 19]	Erro de classificação, análise de variáveis	Matriz de confusão, histograma	Sim
[JAC15]	Análise de variáveis	<i>3-D firework plot</i>	Não
[ZHC21]	Análise de variáveis	<i>Radial chart</i>	Sim
[SL21]	Análise de justiça	Dependência parcial	Não
[WZWY21]	Erro de classificação	Matriz de confusão temporal	Não

#### 4.2.1 Erros de Classificação

Um dos objetivos dos trabalhos desta categoria é o detalhamento de como ocorrem as classificações erradas do modelo, sendo que há diferentes formas de representá-los.

A matriz de confusão é bastante utilizada, pois ela permite comparar o cruzamento entre as classificações consideradas como verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos (Figura 4.2-A) [WPB<sup>+</sup>19]. No trabalho de Cashman et al. [CHH<sup>+</sup>19] ela foi utilizada de forma diferente, mostrando não apenas os totais, mas a distribuição esperada e a ocorrida, para cada categoria da variável, e sendo colorida de forma intuitiva em cada quadrado (Figura 4.2-B).

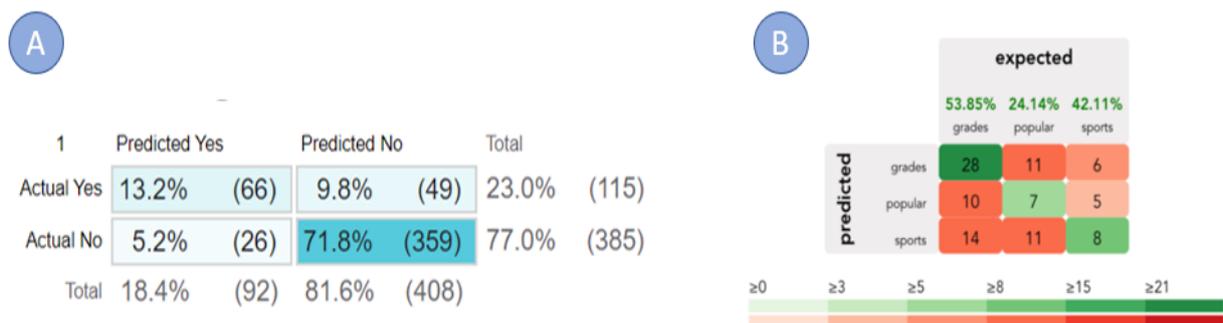


Figura 4.2 – Matriz de confusão: (A) matriz com os erros e acertos totais do modelo [WPB<sup>+</sup>19]; (B) matriz com os erros e acertos totais do modelo de acordo com a categoria da variável [CHH<sup>+</sup>19].

Já no trabalho de Wang et al. [WZWY21] a matriz de confusão é chamada de matriz de confusão temporal e é baseada no diagrama de Sankey [KS98], conforme ilustra a Figura 4.3. O objetivo é rastrear o fluxo de dados durante as interações dos modelos de aprendizado de máquina, sendo que os usuários podem selecionar quais observações desejam examinar na visualização.

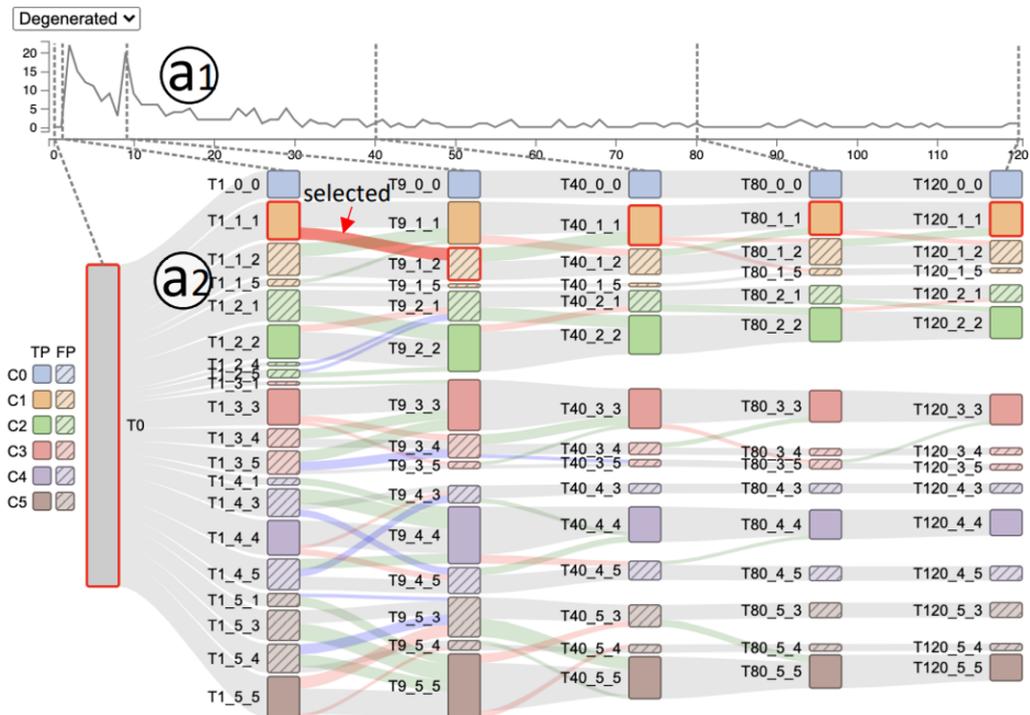


Figura 4.3 – Matriz de confusão temporal: a evolução da matriz de confusão e os dados fluindo entre as diferentes células da matriz [WZWY21].

Outra forma de visualização utilizada é o gráfico de dispersão, no qual cada ponto de predição é colorido de acordo com a variável resposta [ZWM<sup>+</sup>18, WPB<sup>+</sup>19]. No trabalho de Zhang et al. [ZWM<sup>+</sup>18] é possível comparar modelos distintos de acordo com o tipo de erro cometido.

Os gráficos de barras regular e empilhado apresentam a vantagem de fácil investigação sobre a ocorrência de erros, de acordo com faixas de probabilidade específicas, como podemos observar na Figura 4.4-A. No trabalho de Sopan et al. [SBMK18] esses dois tipos de gráficos são utilizados para esse fim, sendo que no gráfico de barras regular podemos verificar quantidade de erros cometidos e no gráfico de barras empilhado o percentual total (Figura 4.4-B). Além disso, as barras são coloridas de forma intuitiva de acordo com o acerto ou erro cometido.

O trabalho de Areosa e Torgo [AT19] utiliza visualizações como *error dependence plots* e *parallel error plots* para visualizar os erros do modelo entre as variáveis. A primeira visualização utiliza o gráfico *box plot*, separando os erros de acordo com as categorias das variáveis (Figura 4.5-A). A segunda mostra como ocorrem os erros entre as variáveis utili-

zando o gráfico de coordenadas paralelas (Figura 4.5-B). Dessa forma, é possível encontrar padrões e identificar a combinação de variáveis que possuem o maior erro.

Leonel et al. [LGL<sup>+</sup>19] utilizam um gráfico 2D separado em áreas, coloridas de acordo com o seu verdadeiro valor, e pontos utilizando símbolos que mostram qual foi a previsão do modelo. Dessa forma, a localização dos erros e as categorias de migração ficam evidentes.

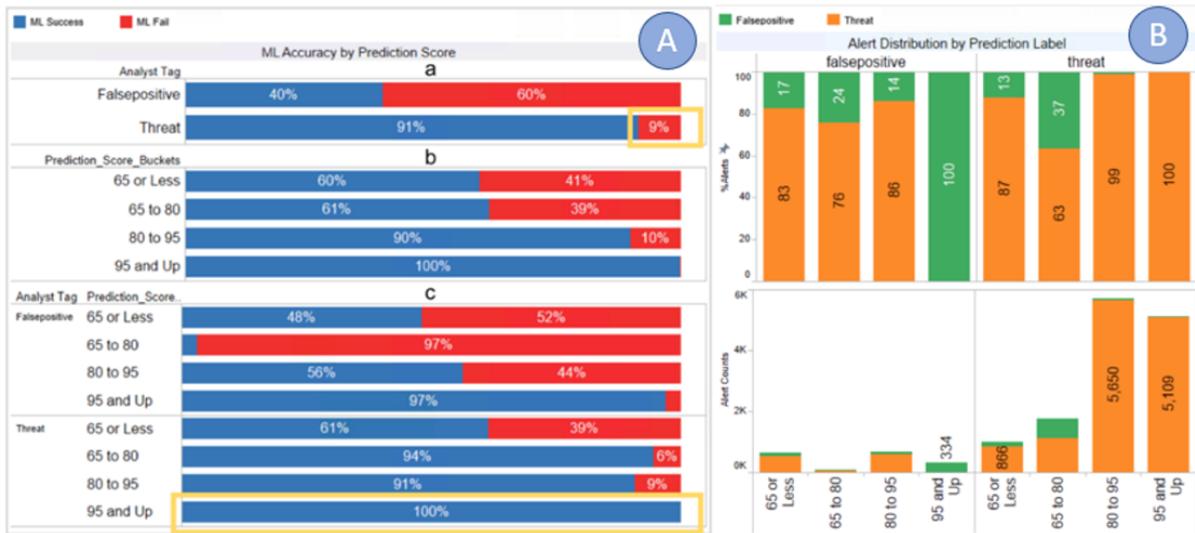


Figura 4.4 – Gráfico de barras: (A) barras empilhadas com a ocorrência de erros e acertos de acordo com a faixa de probabilidade gerada pelo modelo; (B) barras empilhadas e não empilhadas com a quantidade de alarmes tratados e falsos positivos [SBMK18].

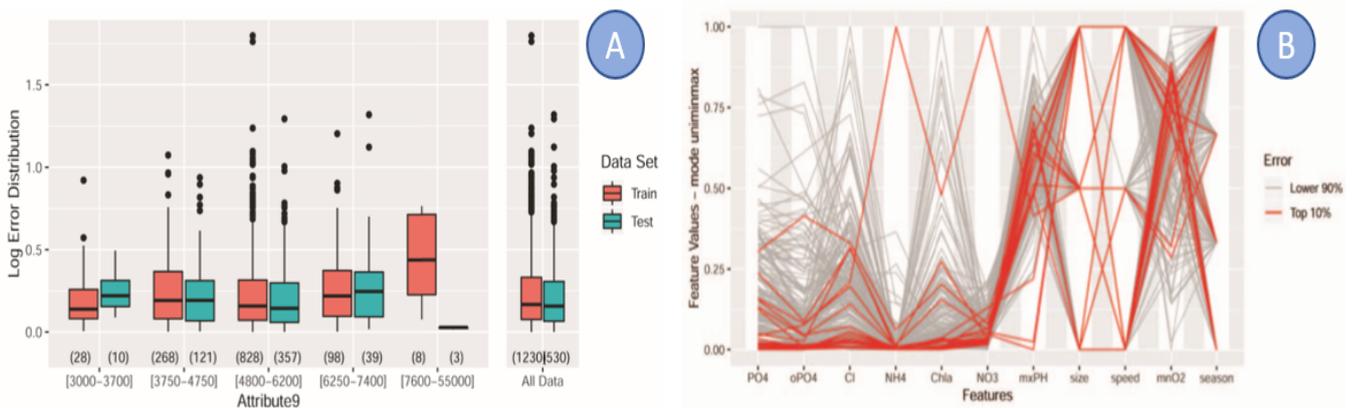


Figura 4.5 – Erros de classificação: (A) *error dependence plots* na visualização do erro do modelo entre as categorias de variáveis na amostra de treino e teste [AT19]; (B) *parallel error plots* na visualização do erro do modelo entre as variáveis [AT19].

#### 4.2.2 Análise de Justiça

O trabalho de Wexler et al. [WPB<sup>+</sup>19] constrói visualizações que ajudam o usuário a entender se o modelo está sendo justo com os indivíduos classificados, ou seja, se o modelo não utiliza variáveis que podem ser discriminatórias. Para isso, utiliza visualizações de desempenho como curva ROC e matriz de confusão, com simulações de ponto de corte de acordo com variáveis selecionadas pelo usuário, como podemos observar na Figura 4.6. De acordo com o exemplo citado no artigo, assim é possível verificar que quando o mesmo ponto de corte é usado para homens e mulheres em um modelo que desejava prever renda acima de 50 mil dólares, a distribuição de probabilidades é bem diferente, pois homens são mais propensos a ter a previsão positiva. Este é um exemplo didático, porém no contexto de crédito há a possibilidade que modelos tratem homens e mulheres de forma distinta com relação a concessão de crédito, algo bastante sensível. Com o crescente uso de modelos de aprendizado de máquina, a análise de justiça sobre os resultados de modelos vem sendo bastante investigada [DLZ<sup>+</sup>19] e esses tipos de visualizações contribuem para diminuir a discriminação que esses modelos podem gerar.



Figura 4.6 – Análise de justiça: resultados do modelo analisados separadamente para homens e mulheres [WPB<sup>+</sup>19].

Szepannek e Lübke [SL21], por sua vez, verificam se o modelo está sendo justo e propõe um algoritmo de correção nos casos onde isso não acontece. Além disso, gráficos de dependência parcial são propostos para visualização da justiça do modelo de acordo com algum atributo, como podemos verificar na Figura 4.7.

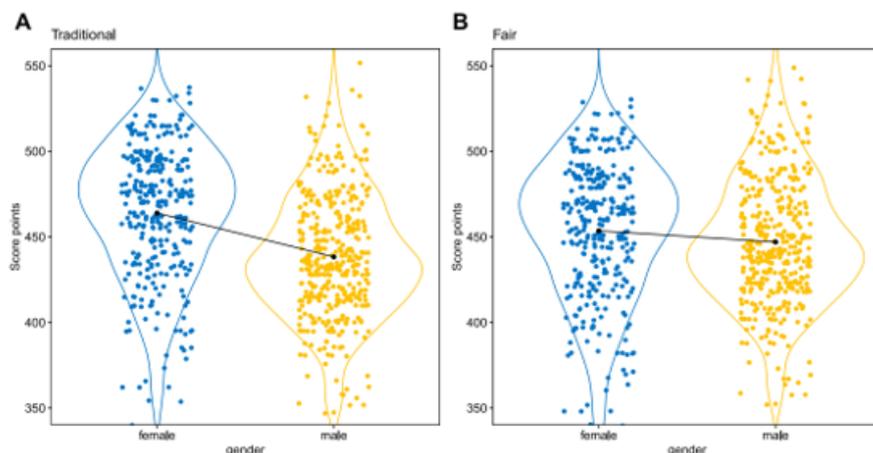


Figura 4.7 – Análise de justiça: (A) Scores do modelo analisados separadamente para homens e mulheres em um modelo sem alteração; (B) Scores do modelo corrigido para ser justo com homens e mulheres [SL21].

#### 4.2.3 Análise de Variáveis

A análise de variáveis preditoras também é abordada nos trabalhos. Foram construídas visualizações que mostram o comportamento individual e cruzado das variáveis e sua relação com o resultado do modelo.

Um gráfico de barras ordenado de forma decrescente, por exemplo, permite verificar a importância de cada variável para o modelo [ZWM<sup>+</sup>18]. No trabalho de Sopan et al. [SBMK18] o modelo criado é utilizado por analistas de segurança no processo decisório de atendimento de chamados de alarmes residenciais. Para isso, eles utilizam gráficos de barras que auxiliam na explicação das categorias das variáveis em conjunto com o resultado gerado pelo modelo.

Análises descritivas das variáveis são exploradas por Cashman et al. [CHH<sup>+</sup>19]. Eles utilizam histogramas interativos que mostram a distribuição dos dados, sendo que cada barra é colorida de acordo com a quantidade apresentada na categoria da variável resposta (Figura 4.8-A). Já Jang e Cook [JAC15] apresentam uma visualização chamada *3-D firework plot*, que mostra qual o efeito de *outliers* no resultado do modelo (Figura 4.8-B).

Um grande número de visualizações para análise de variáveis foi disponibilizado por Wexler et al. [WPB<sup>+</sup>19], permitindo seu entendimento e simulação de hipóteses do usuário. Com um gráfico de dispersão interativo é possível plotar cada ponto de predição

colorida de acordo com a variável resposta, cruzando com outras variáveis do banco de dados, como a própria probabilidade gerada pelo modelo. É possível também simular como ficariam essas distribuições caso houvesse mudança em algum ponto. A Figura 4.8-C mostra cada probabilidade predita pelo modelo, colorida de acordo com a variável resposta. Este trabalho foi o mais completo em analisar as variáveis individualmente ou em conjunto com a variável resposta.

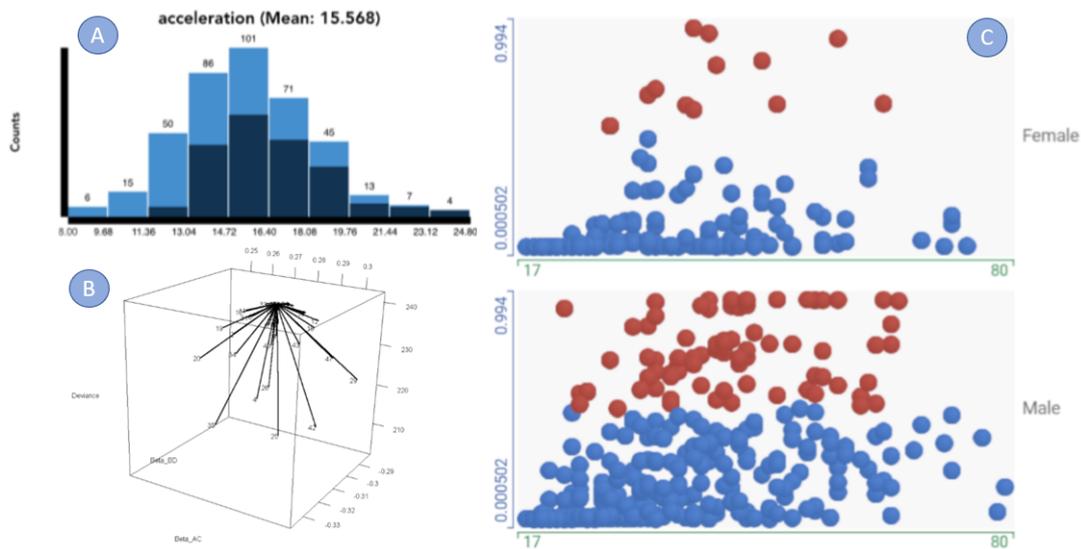


Figura 4.8 – Análise de variáveis: (A) histograma interativo com distribuição da variável aceleração [CHH<sup>+</sup>19]; (B) visualização de *outliers* [JAC15]; (C) probabilidades preditas pelo modelo separadas por sexo [WPB<sup>+</sup>19].

No trabalho de Zhou et al. [ZHC21] a visualização *radial chart* é proposta para identificar, comparar e verificar a importância de variáveis, auxiliando na explicação e avaliação de performance de modelos de aprendizado de máquina. Além disso, também compacta várias informações em uma única visualização, o que ajuda a simplificar o entendimento do modelo. As cores e o tamanho dos arcos ajudam na comparação, pois quanto maior o arco e de acordo com a cor é possível verificar o desempenho dos modelos (Figura 4.9).

#### 4.2.4 Novas Ferramentas

Para gerar as visualizações apresentadas até aqui, alguns trabalhos utilizaram ferramentas pagas, como Tableau, ou livres, como R, e outros desenvolveram novas ferramentas e as disponibilizaram para acesso gratuito. Citaremos aqui as novas ferramentas criadas e suas vantagens.

Zhang et al. [ZWM<sup>+</sup>18] criaram a ferramenta de visualização interativa Manifold, que permite o diagnóstico e comparação entre modelos, sendo que não visa uma técnica específica de modelagem. No trabalho de Wexler et al. [WPB<sup>+</sup>19] foi desenvolvida a fer-

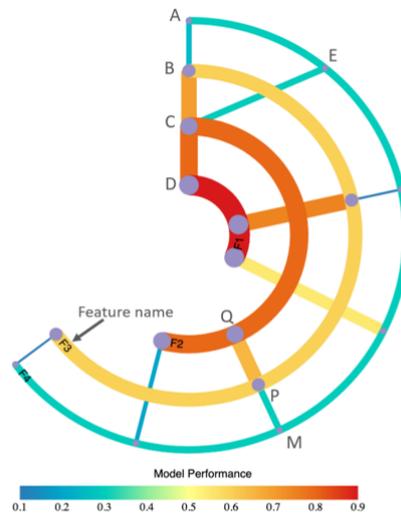


Figura 4.9 – Análise de variáveis: o *radial chart* evidencia a importância e a relação entre variáveis. Cada arco representa o modelo de aprendizado de máquina baseado em cada uma das variáveis (F1, F2, F3, F4). As linhas representam o modelo baseado em múltiplas variáveis [ZHC21].

ramenta *What-If Tool* que também não foca em uma técnica específica e permite verificar a relação entre modelos, porém abrange uma gama maior de possibilidades de visualizações e comparações que o trabalho de Zhang et al. [ZWM<sup>+</sup>18]. O trabalho de Cashman et al. [CHH<sup>+</sup>19] introduz o termo EMA (*Exploratory Model Analysis*) como o processo de explorar o conjunto de modelos potenciais que podem ser treinados em um determinado conjunto de dados e cria um *Visual Analytics Workflow* para o EMA. A análise sugere qual modelo poderia ser utilizado e o executa, sendo possível comparar os resultados gerados. Como essa ferramenta calcula o modelo no momento da execução, isso a torna menos flexível para ser utilizada no acompanhamento de modelos.

O trabalho de Wang et al. [WZWY21] tem como objetivo utilizar *Visual Analytics* para melhorar o entendimento de algoritmos de árvores de decisão como *Extreme Gradient Boosting* (XGBoost) [CG16] e *Light Gradient Boosting Machine* (LightGBM) [KMF<sup>+</sup>17]. A ideia é entender como árvores recém-adicionadas mudam as previsões individuais e os impactos de diferentes variáveis nas interações durante a criação do modelo. Para isso a ferramenta *Visual Interpretation System for Tree Boosting Models* (VISTB) foi criada.

### 4.3 Métricas de Identificação de Desvio

Após o desenvolvimento de um modelo existem as etapas de implantação e acompanhamento [Sid12]. As etapas pós-desenvolvimento são fundamentais para garantir que o modelo atenderá o que foi planejado, pois é possível que a leitura dos dados não acon-

teça corretamente ou mesmo que o cenário atual seja diferente quando comparado ao de desenvolvimento [TCE17].

Na etapa de seleção dos artigos encontramos um trabalho que trata especificamente sobre a fase pós-implantação de modelos de *Credit Scoring* e dois trabalhos que tratam sobre essa etapa em outras áreas. Incluímos trabalhos de outras áreas porque entendemos que eles também podem fornecer métricas e *insights* aplicáveis ao monitoramento de modelos de *Credit Scoring*.

#### 4.3.1 Monitoramento de Modelos de *Credit Scoring*

O trabalho de Bravo e Maldonado [BM15] desenvolve métricas que apoiam na tomada de decisão sobre a troca de modelos de *Credit Scoring* que estejam em produção. A contribuição do trabalho foi construir novas métricas que identifiquem quando há mudança crítica no conjunto de dados.

Uma das métricas criada é o *dimensionless confidence interval*. No cálculo são utilizados os limites do intervalo de confiança dos coeficientes atribuídos à cada variável, ou seja, é construída uma zona de confiança do que se aceita errar para cada variável. É estipulado o que é significativo de acordo com o tamanho de amostra utilizado, pois quando o banco de dados é pequeno se tem menos segurança em se dizer o que é significativo. A métrica compara variáveis preditoras do desenvolvimento do modelo com a pós-implantação. A nova métrica é comparada a outras já existentes, como *Area Under the Curve* (AUC) e *Kolmogorov-Smirnov* (KS). Outra nova medida proposta é o *mean ratio* que calcula a divisão entre a média da variável na amostra de desenvolvimento e na pós-implantação, sendo que é calculado o intervalo de confiança para essa medida.

O trabalho apresenta métricas que se mostram acuradas em quantificar qual foi a mudança pós-implantação do modelo. Um diferencial é que as medidas propostas não necessitam da variável resposta do modelo, ou seja, elas podem ser utilizadas em acompanhamentos de modelos em tempo real.

#### 4.3.2 Monitoramento Aplicado a Áreas Fora do Crédito

Dois trabalhos com aplicação fora da área de crédito foram selecionados, o objetivo de ambos é o monitoramento das falhas na predição dos modelos. São criadas métricas e visualizações que identificam um padrão de erro.

Em um dos trabalhos [GWGv19], os modelos são utilizados na predição dos estágios do sono de pacientes. O seu objetivo é identificar e corrigir predições erradas após

a etapa de desenvolvimento, para que a predição ajude o médico na tomada de decisão sobre o diagnóstico. Foi criada uma ferramenta visual que permite explorar a saída do modelo e entender o que determinou a predição. Uma das formas de detecção abordadas foi o mapeamento de padrões que eram improváveis de acontecer, sendo criadas visualizações, chamadas *cycles view*, que ilustram como esse tipo de desvio ocorre. Por exemplo, a Figura 4.10-A ilustra os resultados de um ciclo de sono de acordo com cada estágio, através do qual especialistas podem detectar desvios de padrões ocorridos. Outra forma de detectar a predição errada foi calcular a entropia, que evidencia casos nos quais as variáveis predictoras são muito diferentes e a probabilidade atribuída pelo modelo é parecida. Foram identificadas 87% das classificações erradas. Além disso, as análises permitiram a identificação de erros nos sensores que eram conectados aos pacientes. Os métodos utilizados no trabalho se mostraram efetivos na identificação de erros, sendo capazes também de identificar erros de *input* dos dados, algo que é comum em situações reais.

Schlechtingen e Santos [SS14] construíram um modelo para predição de falhas em turbinas eólicas. Foram desenvolvidas formas de identificar quando aconteceram erros na predição, para isso foram elaborados gráficos de linha, com relação ao erro, com limites de controle de acordo com seu desvio padrão (Figura 4.10-B). A gravidade do erro é sinalizada com cores, de acordo com limites pré-estabelecidos. O erro de predição é mensurado conforme os dados reais são obtidos. Assim, é possível verificar o desempenho do modelo, retreiná-lo, caso necessário, e verificar a aderência das predições. Como se espera um patamar de erro conhecido, quando algo acontece fora dos limites é possível identificar a falha rapidamente. O trabalho mostrou que utilizar gráficos de linhas com limites é uma alternativa para encontrar padrões fora do esperado e, assim, identificar situações ocorridas na variável resposta do modelo que podem não estar relacionadas com as variáveis utilizadas como *input*. Isso é relevante, pois nem sempre o erro do modelo estará relacionado com as variáveis de *input*, mas sim a uma mudança inesperada no ambiente.

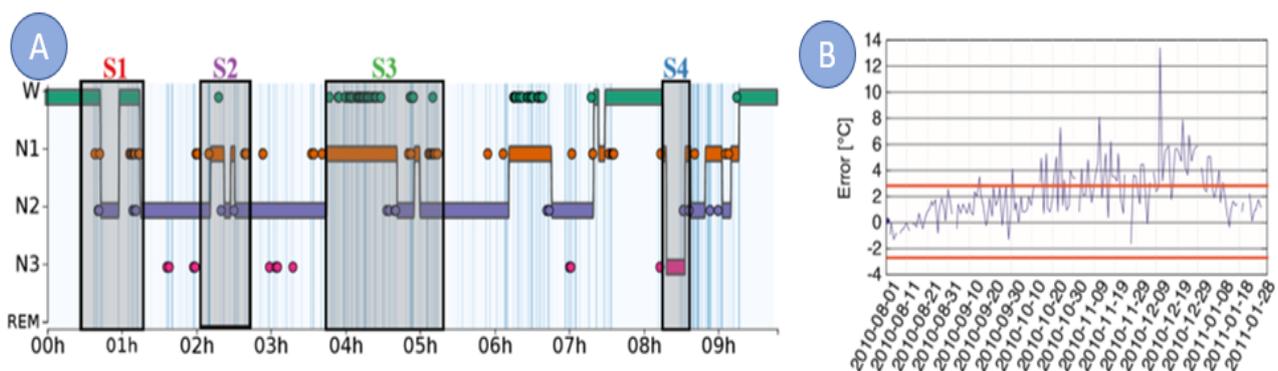


Figura 4.10 – Exemplos de acompanhamento: (A) resultados de um ciclo de sono de acordo com cada estágio [GWGv19]; (B) gráficos de linha com limites de controle de acordo com seu desvio [SS14].

Os dois trabalhos utilizaram visualizações para o monitoramento das falhas nas predições, sendo que, em ambos, o mapeamento dos padrões esperados após a implantação do modelo se mostrou efetivo na descoberta de falhas.

#### 4.4 Acompanhamento Visual no Gerenciamento de Modelos de Crédito

O uso de visualizações no gerenciamento de modelos de crédito é abordado nessa seção. Foram encontrados dois trabalhos que mostram que quando visualizações são introduzidas no processo decisório, no que diz respeito a modelagem, houve mudança em decisões tomadas sem esses recursos, tornando as escolhas mais acuradas.

Um dos objetivos do trabalho de Gingerich e Conati [GC15] é investigar a eficácia da visualização personalizada com *highlights* na acurácia de modelos. Eles concluem que modelos nos quais os usuários empregam esse tipo de visualização podem ser melhores que modelos feitos por usuários que não utilizaram. Assim, o trabalho mostra que decisões tomadas durante o desenvolvimento do modelo que utilizaram visualizações personalizadas para destacar oportunidades foram importantes na tomada de decisão da melhor estratégia da modelagem.

Já no trabalho de Baets e Harvey [DBH20] um experimento controlado foi realizado com o objetivo de comparar o modelo desenvolvido por dois grupos. Um dos grupos tinha a introdução de uma visualização antes da escolha do modelo final, mudando sua decisão inicial, e o outro grupo não. Os resultados do experimento mostram que o grupo com a opção de utilizar visualizações antes da decisão do modelo final obtiveram melhores resultados quando comparado àquele que não tinha essa opção. Com isso, é possível inferir que quando a visualização é introduzida no processo de decisão, os usuários tendem a ajustar a suas escolhas de acordo com o que é visualizado.

Ambos os trabalhos mostram que quando visualizações são introduzidas no processo decisório elas apoiam a tomada de decisão. Como o acompanhamento do modelo busca facilitar o processo de decisão de gestores com relação a mudanças, a introdução de visualizações poderia apoiar esse processo.

#### 4.5 Discussão

Durante a execução da RSL, não encontramos artigos que abordassem especificamente o acompanhamento visual de modelos de *Credit Scoring*. No entanto, há trabalhos que tratam sobre visualização na construção de modelos, principalmente modelos de aprendizado de máquina, e sobre métricas de acompanhamento de modelos de *Credit Scoring* ou modelos preditivos de forma geral. A Figura 4.11 sintetiza o conteúdo apresentado nas

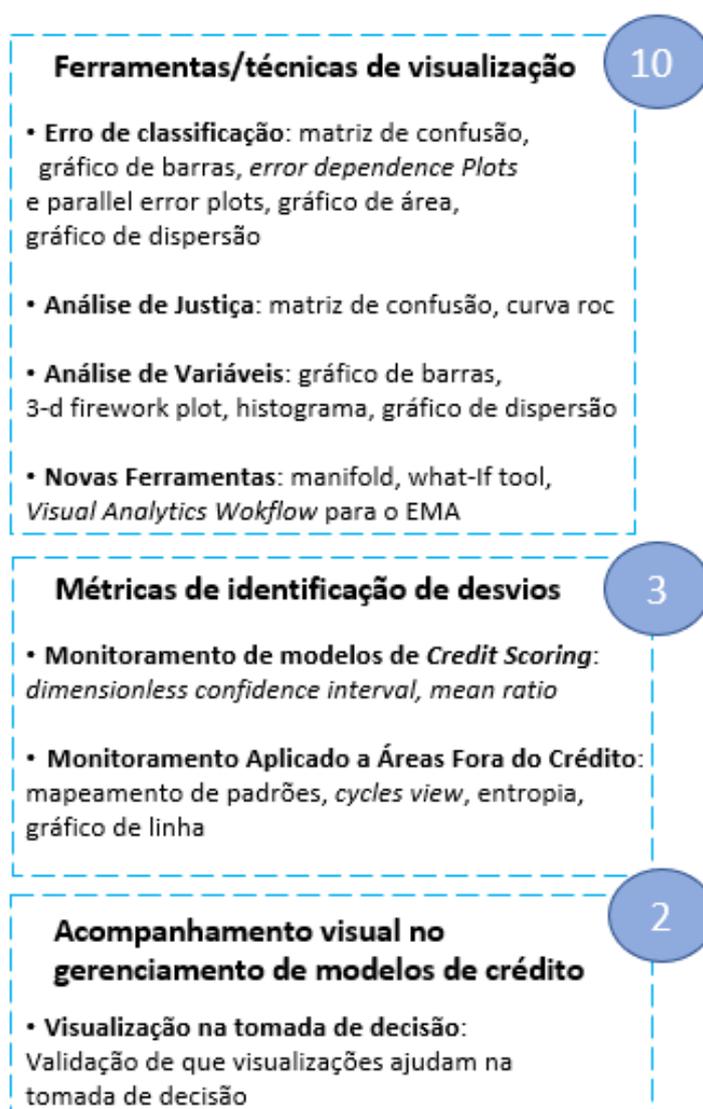


Figura 4.11 – Resultados da RSL.

seções anteriores e mostra quantos artigos foram utilizados na sua construção. Podemos observar que 10 dos 15 trabalhos finais são relacionados a ferramentas e técnicas de visualização, três trabalhos apresentam métricas e identificação de desvios e dois abordam o acompanhamento visual no gerenciamento de modelos de crédito.

Com essa revisão foi possível perceber que a dificuldade da explicabilidade de modelos de aprendizado de máquina propiciou diversas pesquisas que mostraram valor no uso de técnicas de visualização na etapa de modelagem dos dados. Foram criados diferentes gráficos que explicam como o resultado do modelo se relaciona com as variáveis preditoras e onde e como ocorrem os erros do modelo. Quando os modelos de *Credit Scoring* foram criados eles utilizavam majoritariamente técnicas estatísticas, como a regressão logística, na qual a relação entre as variáveis do modelo e o *Score* é de fácil interpretabi-

lidade [B+10]. Isso pode ter motivado a falta de trabalhos relacionados à exploração visual desses modelos.

Assim, há uma lacuna que pode ser preenchida aplicando as diferentes técnicas de visualização utilizadas na explicação de modelos de aprendizado de máquina, assim como métricas de acompanhamento utilizadas em outros tipos de modelo, no acompanhamento visual de modelos de *Credit Scoring*.

## 5. ENTREVISTAS COM ESPECIALISTAS

Segundo Longhurst [Lon03] entrevistas semiestruturadas permitem obter uma maior interação com o participante, pois são realizadas de forma aberta e ele fica livre para esclarecer dúvidas e expor suas opiniões. Portanto, com objetivo principal de realizar um levantamento de requisitos, para entender as principais necessidades dos especialistas com relação a abordagem de VA que seria desenvolvida, conduzimos entrevistas com especialistas da área de modelagem de crédito. Além disso, buscamos responder a QS1 sobre como é feito o acompanhamento visual de modelos de *Credit Scoring*. Logo, conforme descrito por Creswell [CC17], nossa pesquisa se enquadra como um estudo qualitativo com metodologia *survey*. Este formato de pesquisa nos permitiu um levantamento de requisitos mais completo para construirmos uma abordagem de VA que conseguisse atender as reais necessidades dos usuários.

Como conduzimos uma pesquisa qualitativa, cujo objetivo não é generalizar os resultados, é comum entrevistarmos um número reduzido de participantes [BR08]. Nós entramos em contato, principalmente via LinkedIn<sup>1</sup>, com 44 profissionais que atuaram pelo menos dois anos com modelos de crédito. Destes, 14 aceitaram participar da pesquisa. Algum tempo após as entrevistas, um dos participantes pediu para que as respostas fossem retiradas da pesquisa, pois teve receio de ferir alguma questão de confidencialidade. Assim, os resultados foram compilados novamente, desconsiderando as respostas deste participante.

O questionário no formato de perguntas abertas que foi aplicado está no Apêndice A. As perguntas referentes a primeira etapa de entrevistas foram divididas em quatro seções relacionadas aos objetivos da pesquisa, são elas: (1) Levantamento do perfil do participante; (2) Ferramentas e dados; (3) Atividades do monitoramento de modelos; (4) Técnicas de visualização e indicadores. Após finalizarmos a abordagem, foi realizada uma segunda rodada de entrevistas, na qual coletamos o *feedback* de alguns participantes (ver Capítulo 7).

### 5.1 Aplicação e Análise do Questionário

Os participantes responderam ao questionário individualmente, via vídeo chamada, e a duração das entrevistas variou entre 18 e 57 minutos. O TCLE para pesquisa foi enviado previamente para os participantes. Inicialmente, contextualizamos os participantes sobre a motivação e objetivos da pesquisa. Em seguida, perguntamos se havia alguma dúvida adicional sobre a pesquisa ou sobre o TCLE. Por fim, antes de dar início a aplica-

---

<sup>1</sup><https://www.linkedin.com/>

ção do questionário, lembramos os participantes que a entrevista seria gravada e deixamos a seu critério desligar a câmera no momento de responder as perguntas. As entrevistas foram guiadas por um questionário semiestruturado com 28 perguntas, composto por quatro blocos (ver Apêndice A). Deixamos os participantes livres para comentar tópicos sobre o acompanhamento que não fossem estritamente ligados a pergunta, para, assim, conseguirmos capturar como é feito o acompanhamento de forma geral. Anotamos os principais tópicos levantados pelos participantes e, como as entrevistas foram gravadas sempre que eles consentiram, posteriormente fizemos sua transcrição.

Iniciamos as entrevistas com uma pesquisa piloto, coletando o *feedback* do participante sobre o questionário aplicado. Após essa etapa, de acordo com as sugestões coletadas, alteramos algumas perguntas, incluindo exemplos que ajudassem na contextualização das questões. Como houve apenas pequenas alterações no questionário, incluímos a entrevista piloto na análise final dos resultados.

Nas seções a seguir detalhamos como conduzimos as entrevistas e os resultados que obtivemos com as respostas coletadas. Os resultados foram divididos de acordo com o assunto em cada bloco de questões. Iniciamos o questionário com o levantamento do perfil dos respondentes. Em seguida, passamos para os tópicos relacionados às experiências do participante com acompanhamento de modelos. Pedimos para que os participantes nos relatassem suas experiências como um todo, sem focar apenas na última, para que assim conseguíssemos ampliar os resultados e o entendimento de como é feito o acompanhamento em um maior número de instituições e para não ferir nenhum compromisso de confidencialidade do participante junto a empresa em que ele trabalha.

## 5.2 Perfil dos Participantes

Nós entrevistamos 13 profissionais (P1 a P13), nove mulheres e quatro homens que trabalham ou trabalharam com modelagem de crédito há pelo menos dois anos. Apenas dois deles não trabalham atualmente com modelagem de crédito, mas ambos seguem trabalhando em IFs. Os profissionais ocupam diferentes posições na empresa, sendo que sete deles atuam em cargos de gestão. A maior parte dos participantes atua em Porto Alegre (sete), cinco atuam em SP e um atua em Londres. No entanto, a UF onde a maior parte dos participantes já atuou (sete) foi SP. Todos os participantes possuem graduação na área de exatas, sendo que nove deles são formados em Estatística, dois em Matemática, um em Engenharia Civil e um em Física. Além disso, todos citaram ter realizado alguma pós-graduação, sendo que três deles possuem mestrado e um deles possui doutorado. A Figura 5.2 mostra informações adicionais sobre os participantes.

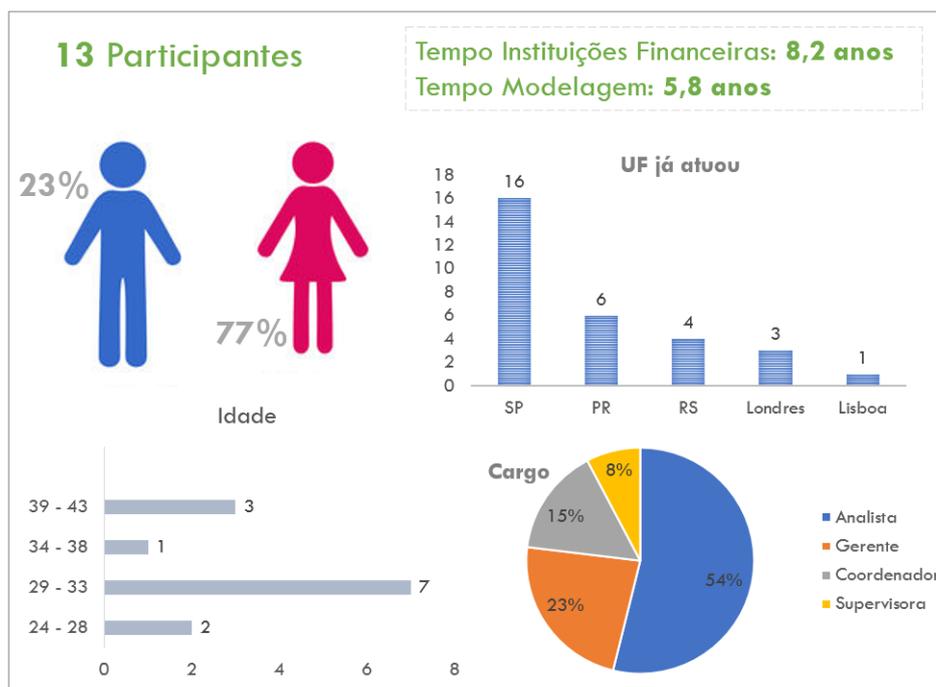


Figura 5.1 – Perfil dos respondentes.

### 5.3 Resultado para Dados e Ferramentas

Em relação ao processamento de dados, os participantes comentaram sobre uma etapa de pré-processamento, na qual ocorre o agrupamento dos dados e posteriormente a exportação para alguma ferramenta que gera as visualizações. O tratamento dos dados inclui a categorização das variáveis do modelo, utilizando as mesmas categorias construídas no desenvolvimento, acompanhando se elas se mantêm estáveis ao longo do tempo, tanto com relação as quantidades, como de discriminação da variável. Além disso, ocorre também a agregação dos dados por faixas de decis de *Score* e sua discriminação de acordo a faixa, ou seja, o percentual de clientes inadimplentes de acordo com a faixa de *Score* atribuída no momento da concessão do crédito.

O formato mais utilizado na importação dos dados agregados são arquivos gerados pelo “Excel”, que foi citado por seis participantes, e os demais formatos citados são apresentados na Figura 5.3. O mesmo número de participantes mencionou que as agregações são feitas com conexões no banco de dados e que as ferramentas não específicas utilizam essa conexão para gerar as visualizações. Adicionalmente, os especialistas pontuaram que a construção do monitoramento busca realizar uma comparação com o desenvolvimento do modelo e que isso é utilizado como base para entender se o modelo está se “comportando” de maneira esperada.

Nenhum participante trabalha ou trabalhou com ferramentas que fossem específicas para o acompanhamento de modelos. No entanto, seis deles afirmaram conhecer

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
<b>Ferramentas</b>													
EXCEL/VBA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SAS Visual Analytics					✓				✓	✓	✓	✓	
TABLEAU		✓		✓			✓						
Power BI	✓		✓	✓				✓					
R									✓				
Cognos								✓					
Python								✓					
MicroStrategy		✓											
<b>Extensão</b>													
Conectar banco		✓	✓				✓		✓			✓	✓
Excel				✓	✓		✓			✓	✓	✓	
csv	✓	✓			✓								
sas7bdat					✓	✓			✓	✓			
txt						✓		✓					

Figura 5.2 – Informações adicionais sobre ferramentas e dados.

ferramentas para essa finalidade. Alguns participantes relataram que fornecedores já lhes apresentaram tais ferramentas, mas elas não foram adquiridas nas IFs. A Tabela 5.1 lista as empresas que apresentaram essas soluções e o número de citações. Quando questionamos sobre o motivo das ferramentas específicas não terem sido adquiridas, alguns participantes comentaram sobre o alto preço e outros sobre o tempo que seria gasto na adaptação do que já possuem hoje para a nova ferramenta. A SERASA foi citada por um participante como tendo uma ferramenta específica, mas entendemos que essa é uma solução disponibilizada exclusivamente para os clientes que compram os modelos desenvolvidos pela empresa.

Tabela 5.1 – Empresas que oferecem soluções específicas.

<b>Empresa</b>	<b>nº</b>	<b>site</b>
SAS	3	<a href="https://www.sas.com/pt_br/solutions.html">https://www.sas.com/pt_br/solutions.html</a>
Data Robot	2	<a href="https://www.datarobot.com/solutions">https://www.datarobot.com/solutions</a>
FICO	2	<a href="https://www.fico.com/br/node/78">https://www.fico.com/br/node/78</a>
Data Risk	1	<a href="https://datarisk.io/solucoes/">https://datarisk.io/solucoes/</a>

Todos os participantes já utilizaram ferramentas não específicas para o acompanhamento de modelos, como mostra a Figura 5.3. Sobre o uso dessas ferramentas, questionamos se elas atendiam as necessidades para um monitoramento apropriado, oito disseram que elas atendem, quatro disseram que elas atendem parcialmente e um disse que elas não atendem. Os participantes que responderam que as ferramentas atendem, comentaram que conseguiam criar todos os indicadores e visualizações que julgavam necessários, mas que a sua construção demanda muito trabalho. Os que responderam que as ferramentas atendiam parcialmente, comentaram que conseguiam apenas construir os indicadores

básicos e que os acompanhamentos eram feitos de acordo com as demandas em um processo evolutivo. Além disso, comentaram que a adaptação demanda muito trabalho, pois o processo é muito manual. O participante que respondeu que as ferramentas não atendem citou que elas são muito limitadas. Por exemplo, alguns cálculos não são possíveis no Tableau e o R não atendia devido a sua dificuldade de manutenção e necessidade de muito treinamento. Foi apontado também que esse tipo de ferramenta depende de muita interpretação e não gera alertas quando há algum problema nos modelos. Um dos participantes, que respondeu que essas ferramentas atendem, comentou também que quando alguns modelos passaram a ser desenvolvidos com aprendizado de máquina e não com regressão logística, as ferramentas não específicas deixaram de atender.

#### 5.4 Resultado para Técnicas de Visualização e VA

Fizemos perguntas relacionadas as visualizações que os participantes utilizam em seus monitoramentos e sobre a contribuição de uma abordagem específica de VA para o acompanhamento de modelos de *Credit Scoring*.

Todos os participantes comentaram que são utilizadas visualizações na interpretação dos dados, e as mais citadas foram as seguintes: barras, linhas, histograma e tabelas. Podemos observar o número de citações na Tabela 5.2. A combinação do gráfico de barras e gráfico de linhas na mesma visualização foi bastante mencionada, principalmente para o acompanhamento do *Score*, com as barras mostrando a quantidade em faixas de *Score* e as linhas mostrando o percentual de inadimplência das faixas. Além disso, são utilizadas tabelas para o acompanhamento dos indicadores utilizados pela empresa e cores para informar caso não estejam dentro dos limites esperados. O gráfico de linhas foi mencionado principalmente para o acompanhamento das séries históricas de dados, para verificar possíveis mudanças de padrão. Alguns participantes não conseguiram lembrar quais visualizações utilizavam, mas citaram que os objetivos delas eram os seguintes:

- Verificar a distribuição do *Score*;
- Acompanhar os indicadores;
- Comparar os resultados do modelo atual com o modelo do desenvolvimento (*Score* e variáveis);
- Avaliar tendências.

Quando perguntamos sobre quais análises e visualizações uma abordagem de VA para o acompanhamento de modelos de *Credit Scoring* deveria conter, os participantes, em geral, repetiram as visualizações que já são utilizadas por eles, citados na Tabela 5.2. Outras considerações feitas pelos participantes são descritas a seguir.

Tabela 5.2 – Número de citações.

Visualização	citações
Barras	7
Linhas	7
Histograma	4
Tabelas	3

- **Sinalização com cores** - Cores sinalizando se os indicadores estão dentro do aceitável ou perto dos limites facilitaria a interpretação dos indicadores que são mostrados em tabelas.
- **Análise de cenários** - Visões estratificadas para que seja possível identificar o impacto nos modelos quando ocorre alguma mudança inesperada. A quarentena que o Brasil viveu em 2020 foi citada por quatro participantes como exemplo de mudanças nos padrões esperados pela empresa e com grande impacto nos modelos.
- **Novos indicadores** - Indicadores que sinalizem mudanças bruscas no modelo.
- **Glossário** - Glossário esclarecendo os limites estabelecidos e uma breve explicação sobre os indicadores utilizados. Além disso, informações sobre os modelos da IF, tais como número de modelos, data em que foram calibrados, quem desenvolveu, etc.
- **Cálculo do Score** - Visualizações relacionadas ao cálculo do *Score* nos motores de crédito, por exemplo, falha no cálculo, tempo médio de resposta e análise de preenchimento das variáveis utilizadas pelo modelo.
- **Visão resumo e visão detalhada** - A maior parte dos participantes, em diferentes momentos do questionário, comentou que gostaria de uma página que resumisse o status de todos os modelos utilizados pelas IFs, apontando quais modelos estão mostrando sinais de deterioração e deveriam ser observados no detalhe. Porém, que fosse possível entrar no detalhe da performance de cada variável caso fosse necessário, ou seja, separar o nível macro do micro.

Perguntamos aos participantes se uma abordagem de VA contribuiria para o acompanhamento dos modelos das IFs, 12 participantes disseram que contribuiria e um disse que dependeria da abordagem. Um dos pontos citados como motivador para as respostas foi a relevância do tema, pois as IFs utilizam os modelos nas decisões sobre concessão do crédito, assim, caso o modelo não esteja desempenhando como o esperado, as decisões erradas acarretarão perdas financeiras. P10 comentou que: *“O modelo sempre tem o seu círculo de vida e se não estivermos monitorando isso é um prejuízo muito grande para quem está utilizando o modelo. Ter isso de forma automatizada é muito interessante.”*

Outro ponto é a agilidade que se ganha em ter um monitoramento visual, o que facilitaria a gestão e tomada de decisão sobre a continuidade do modelo. A economia

de tempo também foi mencionada, pois, segundo os participantes, o desenvolvimento das visualizações e cálculos dos indicadores demora muito tempo. P4 comentou que: *“Hoje não temos ferramentas voltadas pra isso, então a princípio quem precisa fazer o monitoramento precisa fazer todos os cálculos dos indicadores. Caso existisse uma ferramenta de Visual Analytics específica pra isso facilitaria bastante.”*

Foi citado também que o acompanhamento de modelos é um tema com poucos materiais de consulta, então ter uma abordagem de VA facilitaria. P1 apontou que: *“É o ponto mais crítico, porque para desenvolver têm muito curso, ferramenta, etc. Mas, monitoramento de forma geral é pobre.”*

O participante P6 que respondeu que dependeria da abordagem e justificou que: *“Se for uma ferramenta como Excel que entraria com a informação meio que pronta e apenas ajustaria a visualização seria legal. Mas se for uma ferramenta para o desenvolvimento, daí é mais complicado por que teria que ter todo o input preparado para ela e nem sempre ela comporta, por que até o Excel é limitado em número de linhas. As IFs estão cada vez mais com um volume maior de dados para armazenar e processar. Outra limitação que eu vejo é estatística, pois, por exemplo, cálculo de capital precisa de uma ferramenta com transformações matemática e estatística que não são tão simples de implantar. Então dependendo do que quero reportar tu precisa ter uma ferramenta que faça todo esse processo matemático e estatístico que precisamos. Então se for muito simples ou limitado talvez não sirva para a IFs.”*

## **5.5 Resultado para Atividades no Monitoramento de Modelos de Credit Scoring**

Além das perguntas sobre a construção do monitoramento e as principais atividades executadas, questionamos sobre a relevância do tema e dificuldades encontradas. Quando questionamos se de forma geral os modelos são monitorados nas IFs, 11 participantes responderam que sim; um participante respondeu que a maioria sim, mas que havia modelos sem monitoramento e um participante respondeu que em uma instituição financeira onde trabalhou sim e em outra não. Entretanto, todos os participantes acreditam que é importante que os modelos desenvolvidos sejam monitorados. Os principais motivos que justificam essa relevância são apresentados na Figura 5.5.

Ainda sobre a importância do monitoramento, o principal ponto comentado foi “Tomada de decisão”, dado que os modelos auxiliam as IFs na decisão do crédito e isso pode comprometer sua saúde financeira. P2 comentou que:

*“Por que o Score é o que toma a decisão toda na empresa e se tiver ruim colocamos a saúde financeira da IF em risco.”*

A “Deterioração” também foi sinalizada, pois o modelo perde performance ao longo do tempo e é necessário identificar o momento em que isso ocorre. No mesmo sentido, podem ocorrer “Mudanças” no perfil dos clientes do banco devido as estratégias de crédito, ou mesmo por cenários econômicos diferentes do desenvolvimento do modelo. P5 comentou que:

*“Os modelos são criados como um retrato do passado, baseado no comportamento da carteira de crédito e a ideia é que no futuro esse perfil se manterá, mas isso não é necessariamente verdade. Pode ser que o público mude, por estratégia da empresa ou mesmo por que o cenário macro econômico mudou. O modelo construído pode perder aderência com o passar do tempo, inclusive isso é comum então precisamos monitorar em tempo integral, para que caso não esteja se possa tomar as medidas necessárias.”*

Logo, é relevante que a instituição consiga identificar quando isso ocorre e o impacto disso nos modelos. “Tomada de ação” aparece porque o acompanhamento auxilia a decidir se o modelo precisa ser trocado, recalibrado ou ajustado. Por meio do acompanhamento também é possível identificar algum “Erro” no cálculo do modelo, devido, por exemplo, a falta de preenchimento de alguma variável. P7 comentou que:

*“Podemos identificar inicialmente casos em que o modelo não está funcionando como deveria e precisamos redesenvolver ou aplicar alguma transformação no modelo para que ele volte a performar da forma que esperamos. Se deixamos de fazer essa análise o banco pode estar perdendo dinheiro, deixando clientes que são ruins entrando dentro da carteira.”*

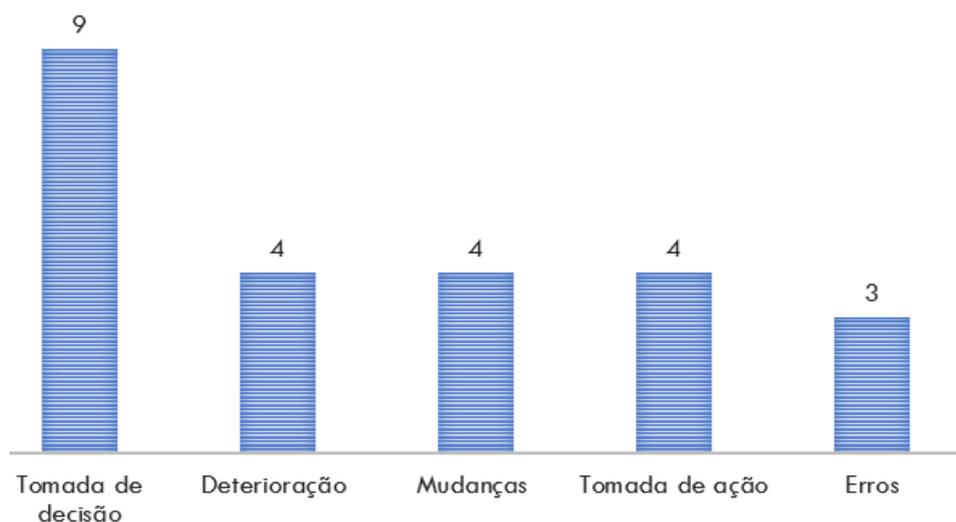


Figura 5.3 – Importância do monitoramento dos modelos nas instituições financeiras.

Perguntamos aos participantes sobre as dificuldades na construção do monitoramento, e eles responderam que, apesar da importância, é difícil conseguir que essa atividade seja priorizada dentre as demais atividades da área. Isso se dá principalmente por

outro ponto citado como dificuldade que é o tempo gasto na construção, pois ela é muito manual e os indicadores precisam ser construídos desde o início, o que faz com que não se busque novos indicadores e que se use apenas os indicadores básicos. Um dos participantes citou preocupação neste processo manual, pois é suscetível a erros. Além disso, foi citada também a dificuldade em disseminar a importância da atividade para os demais membros da instituição financeira. Os participantes citaram que o grande número de modelos e informações dificultam a construção. Ainda, o que se monitora fica em um ambiente diferente do desenvolvimento e há grande dificuldade em consolidar essas informações em um único lugar. Adicionalmente, a equipe que constrói o modelo pode não ser a mesma que monitora o que dificulta ainda mais a construção do monitoramento.

Com relação a periodicidade em que monitoramento ocorre, quatro participantes responderam que ele ocorre mensalmente, um respondeu que diariamente ou semanalmente e oito responderam que depende do modelo. Os que responderam que depende do modelo, justificaram que ele tem que acontecer de acordo com a sua atualização, ou seja, modelos com atualização mensal devem ser atualizados mensalmente, mas modelos online deveriam ser atualizados em tempo real ou diariamente. Um dos participantes comentou que não acredita ser necessário que a atualização seja feita em uma periodicidade menor que mensal.

Perguntamos aos participantes se o monitoramento dos diferentes modelos é apresentado na mesma interface, e verificamos, como ilustra a Figura 5.5, que a maior parte respondeu que isso não ocorre. No entanto, 11 acreditam que seria importante, pois seria uma forma eficiente de fazer a gestão dos modelos. Outro participante respondeu que seria interessante, mas não fundamental, já que os modelos são independentes, mas que gostaria de um resumo sobre todos os modelos. Por fim, um respondeu que não, mas que gostaria de uma visão resumindo todos os modelos.

Todos os participantes acham importante acompanhar a distribuição dos dados antes que ocorra o desfecho da variável resposta do modelo. Os principais pontos citados como importantes foram: a variável resposta do modelo pode demorar muito para acontecer; é preciso saber se o modelo apresenta sinais de mudança de público, instabilidade em alguma variável preditiva, ou até mesmo alguma falha no cálculo da variável e, conseqüentemente, do modelo, para que se possa tomar uma ação rápida.

## **5.6 Resultado para Indicadores**

No bloco de questões sobre indicadores, perguntamos sobre o uso de indicadores de crédito como proxy para variável resposta do modelo. Além disso, questionamos sobre quais indicadores estatísticos são utilizados para mensurar a performance do modelo.

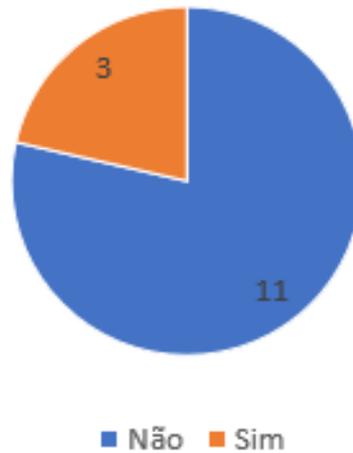


Figura 5.4 – Preferência pelo monitoramento na mesma interface.

Sobre o uso de indicadores de curto prazo como proxy para variável resposta, nove participantes responderam que utilizam e quatro responderam que não. Os indicadores citados como utilizados estão na Tabela 5.3. O indicador mais utilizado foi o *First Payment Default* (FPD) que indica se o cliente pagou a primeira parcela do contrato com a instituição financeira. De forma geral, os indicadores utilizados verificam se os clientes atrasaram o pagamento em certos períodos de tempo. Os períodos citados variaram bastante, sendo que o segundo indicador mais utilizado foi o *Over 30 MOB* (*Month on Book*) quatro, que verifica se o cliente possui mais de 30 dias de atraso depois de quatro meses. Um dos respondentes comentou que a instituição financeira não decide sobre refazer o modelo com base nessas métricas, mas sim acompanham mais atentamente possíveis mudanças nos patamares de inadimplência esperada. Questionamos os participantes que não utilizam esses indicadores sobre a relevância de tê-los no acompanhamento, três deles responderam que seria importante, sendo que um disse que não tinha pensado nessa possibilidade, mas que era interessante e um respondeu que não, pois seria uma avaliação do modelo com outra variável resposta.

Com relação a identificação de possíveis problemas após a implementação dos modelos, de forma geral, os especialistas comentaram que analisam os indicadores de crédito, verificando se algo está diferente do esperado, assim como a distribuição das variáveis e do *Score* calculado, sempre comparando com o desenvolvimento. A interação com a equipe de vendas também foi citada como forma de validação, pois eles alertam a equipe técnica quando existe uma taxa de reprovação ou aprovação de propostas de crédito maiores que o esperado.

Sobre os indicadores estatísticos utilizados nos acompanhamentos, o indicador mais citado foi o Kolmogorov Smirnov (KS), que verifica a discriminação dos modelos. Todos os participantes utilizam mais de um indicador, e todos que foram mencionados estão

Tabela 5.3 – Indicadores no acompanhamento de curto prazo.

Indicador	nº de citações
<i>First Payment Default (FPD)</i>	5
<i>Over 30 MOB 3</i>	3
<i>Ever 30 MOB 4</i>	3
<i>Ever 30 MOB 3</i>	2
<i>Over 60 MOB 6</i>	1
<i>Over 15 MOB 3</i>	1
<i>Second Payment Default</i>	1
<i>Third Payment Default</i>	1

apresentados na Figura 5.6. Apenas um participante não lembrou o nome dos indicadores utilizados, mas comentou que os utiliza para comparar o estimado versus o realizado.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
<b>Indicador Estatístico</b>													
KS2	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
GINI	✓	✓	✓				✓		✓	✓		✓	✓
Estatística Descritiva	✓		✓	✓				✓	✓	✓	✓		✓
ROC	✓		✓		✓			✓	✓			✓	✓
PSI			✓		✓		✓			✓	✓		
KS1		✓	✓		✓					✓		✓	
IEP		✓											
Intervalo de Confiança				✓									
R2					✓								
WOE							✓						
VDI									✓				
AUROC										✓			
Acurácia							✓						
Coefficiente de Correlação							✓			✓			
Matriz de confusão								✓					

Figura 5.5 – Indicadores estatísticos listados pelos participantes.

## 5.7 Necessidades dos Usuários

Como fechamento da entrevista, pedimos para que os participantes citassem quais funcionalidades uma abordagem de VA para o acompanhamento de modelos deveria conter. Os participantes foram encorajados a citar quaisquer funcionalidades ou análises que achassem necessárias. Com base nas respostas elencamos cinco necessidades dos usuários descritas a seguir, utilizando como critério para inclusão a necessidade ter sido citado por pelo menos quatro participantes.

### 5.7.1 Praticidade

A abordagem deve conter os gráficos e os indicadores mais utilizados prontos, assim como uma breve explicação sobre o conceito de cada um, pois os usuários não querem ter que construí-los manualmente. Os gráficos e indicadores devem ser comparáveis ao desenvolvimento. Um dos participantes citou também que a abordagem não pode se fixar em uma técnica, ela deve ser flexível a diferentes tipos de modelos. Outro participante citou que não gostaria de visualizar todas as variáveis do modelo ao mesmo tempo e que gostaria de poder ir trocando de forma interativa.

### 5.7.2 Facilidade de Uso

A abordagem deve ser fácil de ser utilizada e a entrada de dados deve ser simples. Além disso, a resposta aos comandos deve ser rápida, e para isso a abordagem deve agregar os dados e não lidar com os casos analíticos. Por exemplo, ao invés do banco de dados conter uma linha para cada *Score* recebido pelo cliente, que se crie faixas de *Score* com as suas quantidades. Ainda, os participantes gostariam de flexibilidade para trocar os limites dos indicadores. A abordagem deve ser bonita e amigável, com ajuste de rótulos e legendas. Deve ser interativa com o usuário para troca de cores e títulos, escolha de indicadores a serem visualizados e comparação com o desenvolvimento. Ela também deve permitir filtros de tempo para períodos específicos.

### 5.7.3 Exploração de Cenários e Teste de Hipóteses

Os usuários querem poder testar hipóteses e cenários, verificando como o modelo se adere a subpopulações específicas. A abordagem deve poder filtrar a aplicação dos modelos de acordo com alguma variável que estratifique os públicos. Foi citado que os modelos podem ter *Scores* chamados de desafiantes, ou seja, está sendo utilizado um modelo, mas outro também é calculado, assim devia ser possível compará-los. Os usuários gostariam de testar o que acontece caso haja uma mudança no público, sendo bastante citado como exemplo a pandemia da COVID 19 que alterou o comportamento dos clientes, e uma análise deste impacto seria muito interessante.

#### 5.7.4 Emissão de Alertas

O uso de alertas foi citado tanto no fechamento da entrevista como em vários momentos do questionário. Os usuários gostariam de definir níveis de alertas para os indicadores, para que quando eles estejam fora do que eles considerem adequados, de acordo com a governança dos modelos, eles possam fazer uma investigação mais detalhada. Além disso, seria interessante que a abordagem também enviasse e-mails caso houvesse algum problema, para não ser necessário que algum analista faça o constante acompanhamento.

#### 5.7.5 Resumo dos Modelos

Os participantes gostariam que houvesse uma página com um resumo dos indicadores dos diferentes modelos, apresentando tudo de forma gerencial. Assim, seria necessário fazer uma análise detalhada apenas se essa página contivesse algum alerta sobre um problema específico. Neste caso, a abordagem também deveria permitir uma verificação detalhada do que pode ter acontecido. Além disso, ela poderia conter uma outra página com informações sobre o modelo, como data de implantação e nome de quem desenvolveu.

### 5.8 Conclusão das Entrevistas

Por meio das entrevistas foi possível verificar que as IFs acompanham os modelos de *Credit Scoring*, no entanto, não são utilizadas ferramentas específicas para fazer este acompanhamento. Ao mesmo tempo, a maior parte dos participantes tem conhecimento de sua existência, mesmo que nenhum deles as utilize. Conforme comentado pelos participantes, o custo elevado e o tempo que seria gasto na adaptação dos acompanhamentos já utilizados pela instituição financeira são os principais ofensores no uso dessas soluções.

Os participantes veem extrema importância no acompanhamento dos modelos, mas, devido a outras atividades de trabalho, encontram dificuldade em conseguir priorizar sua construção. Por consequência, quando são construídos novos modelos, não se tem tempo de incluir novos indicadores ou novas visualizações.

Sobre as visualizações e indicadores utilizados nos acompanhamentos, podemos perceber que as visualizações citadas nas entrevistas foram as mesmas encontradas na RSL (ver Capítulo 4). Porém, com as entrevistas foi possível coletar um maior número de indicadores estatísticos aplicados. Além disso, podemos incluir os indicadores de crédito utilizados em uma visão de curto prazo para uma proxy da variável resposta do modelo.

As entrevistas nos permitiram entender como acontece o acompanhamento dos modelos em IFs, respondendo a QS1 e contribuindo nos resultados encontrados na RSL que respondem as QS2, QS3 e QS4 (ver Capítulo 4). Ainda, levantamos as principais dificuldades encontradas em sua construção, o perfil dos profissionais que atuam nessa área e as principais necessidades dos participantes para uma abordagem de VA para o acompanhamento de modelos de *Credit Scoring*.

## 6. DESCRIÇÃO DO PROTÓTIPO

Seguindo o modelo de VA proposto por Keim et al. [KKEM10], que combina análises automáticas e visualizações com interação do usuário, identificamos as etapas apresentadas na Figura 6.1. Esta figura ilustra como o protótipo que construímos e chamamos de VACS (*Visual Analytics para o Acompanhamento de Credit Scoring*) pode ser utilizado no gerenciamento de modelos. O protótipo está disponível para uso em [https://daianealdo.shinyapps.io/demo\\_vacs/](https://daianealdo.shinyapps.io/demo_vacs/). Após a entrada de dados e construção do modelo, ele passa a ser aplicado na análise dos dados utilizados em concessões de crédito. A construção do banco de dados diz respeito a etapa de agregação dos dados, ou seja, eles não são gerados analiticamente, mas sim utilizando contagens. Explicamos melhor esse tratamento na Seção 6.2. Utilizando esses dados, o VACS gera diversas análises e visualizações, que serão descritas nesse capítulo. Por fim, o analista utiliza o VACS para auxiliar na tomada de decisão sobre a performance do modelo, se ela segue satisfatória ou se é necessário voltar para a etapa de construção, na qual o modelo pode ser ajustado ou reconstruído. Portanto, o VACS permite acompanhar modelos construídos com qualquer tipo de técnica, pois o que será acompanhado é a saída do modelo (*Score*) e as variáveis que o compõe.

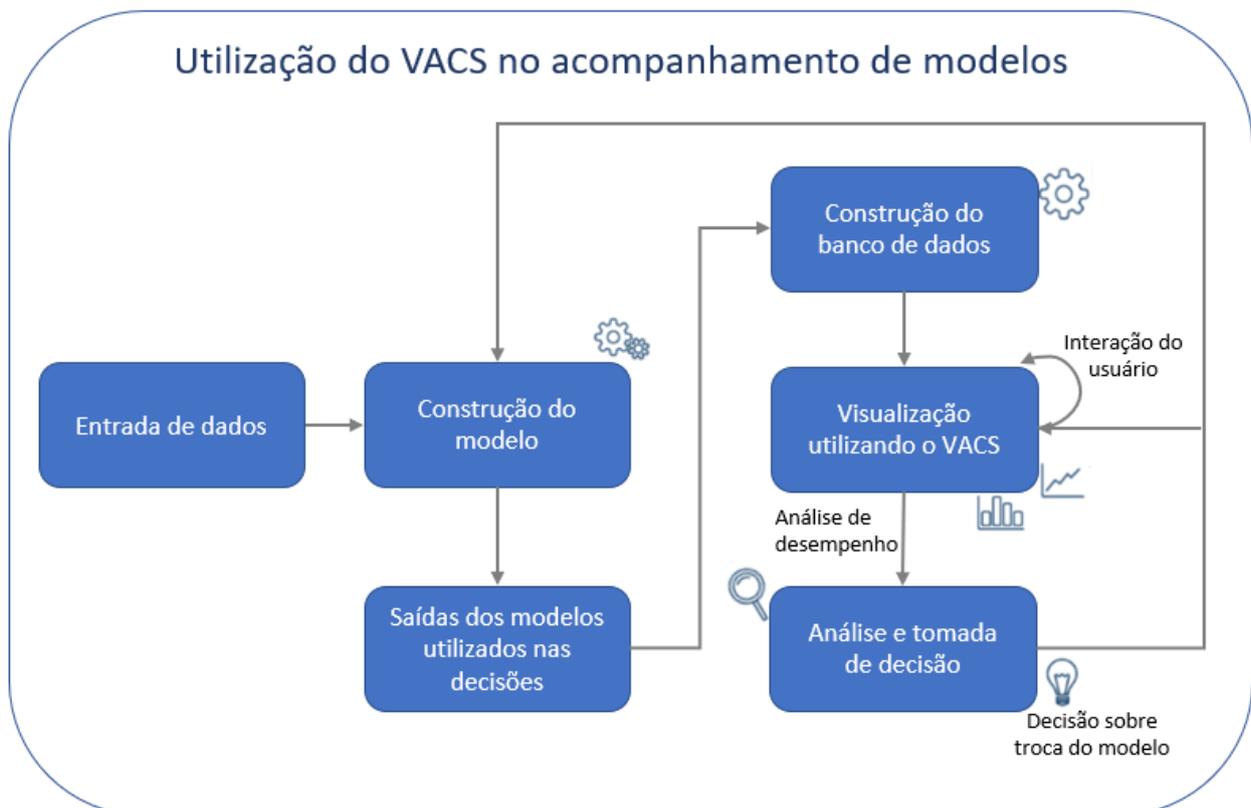


Figura 6.1 – Visão geral sobre o uso do VACS no gerenciamento de modelos.

Este capítulo contém uma descrição de como o VACS foi construído, incluindo suas funcionalidades. A Seção 6.1 aborda quais visualizações e indicadores optamos por incluir no protótipo. Na Seção 6.2 mostramos como funciona a entrada de dados do VACS e quais foram os bancos de dados utilizados nos modelos. A explicação de como construímos os dois modelos preditivos está na Seção 6.3 e na Seção 6.4 detalhamos as funcionalidades implementadas e suas respectivas finalidades.

## 6.1 Visualizações e Indicadores

Para construção do protótipo consideramos os resultados obtidos na RSL e nas entrevistas com especialistas de domínio. A Figura 6.2 resume os indicadores e técnicas de visualizações levantadas durante ambos estudos com uma sinalização em verde daqueles incluídos no VACS.

Todas as visualizações citadas nas entrevistas foram incluídas no protótipo. Acabamos não incluindo algumas visualizações identificadas na RSL porque seria necessário considerar os dados analíticos de *Score* em conjunto com as variáveis preditoras no banco de dados. A exceção é o gráfico de *radial chart* que acabou sendo incluído em uma recente atualização da RSL, após a construção do VACS. Com relação aos indicadores, o VDI é o único citado pelos especialistas que não foi incluído, pois não conseguimos encontrar referências na literatura referente a este indicador. Entre os enumerados através da RSL, ficaram de fora o *mean ration* e o *dimensionless confidence Interval*, pois seria necessário incluir outras informações no banco de dados, como o desvio padrão das variáveis preditoras.

## 6.2 Entrada de Dados

As IFs possuem um grande volume de informações sendo gerado a todo momento, pois são milhares de concessões de crédito realizadas que precisam ser monitoradas. Uma ferramenta de acompanhamento precisa considerar isso, para que não haja lentidão durante as interações com o usuário devido ao volume de dados. Os especialistas citaram que uma das formas de lidar com isso é a agregação prévia dos dados (ver Seção 5.7.2), ou seja, ao invés de carregar no banco de dados todos os casos para cada proposta de crédito, eles sumarizam essas informações em quantidades de acordo com informações de interesse. Por exemplo, ao invés de carregar a idade de todos os clientes no banco de dados, eles fazem uma contagem por faixa de idade. Assim, independentemente de serem milhares de informações sobre a variável, elas podem ser agregadas em poucas linhas.

	RSL	Especialistas
<b>Técnicas de Visualização</b>		
Gráfico de Barras ✓	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Gráfico de Linhas ✓	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Histograma ✓	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Gráfico de Dispersão ✓	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Parallel Error Plots	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Radial Chart	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Gráfico de Área	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Error Dependence Plots	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3-d firework plot	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<b>Indicadores</b>		
Curva ROC ✓	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Matriz de Confusão ✓	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Estatística Descritiva ✓	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Acurácia ✓	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Mean ration	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Dimensionless confidence interval	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Intervalo de Confiança ✓	<input type="checkbox"/>	<input checked="" type="checkbox"/>
KS1 ✓	<input type="checkbox"/>	<input checked="" type="checkbox"/>
KS2 ✓	<input type="checkbox"/>	<input checked="" type="checkbox"/>
GINI ✓	<input type="checkbox"/>	<input checked="" type="checkbox"/>
PSI ✓	<input type="checkbox"/>	<input checked="" type="checkbox"/>
IEP ✓	<input type="checkbox"/>	<input checked="" type="checkbox"/>
R2 ✓	<input type="checkbox"/>	<input checked="" type="checkbox"/>
WOE ✓	<input type="checkbox"/>	<input checked="" type="checkbox"/>
VDI	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Entropia ✓	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Figura 6.2 – Técnicas de visualização e indicadores levantados na RSL e nas entrevistas.

A fim de atender a necessidade dos participantes sobre facilidade de uso, conforme levantado nas entrevistas (ver Seção 5.7.2), e ao mesmo tempo conseguir calcular os indicadores e gerar as visualizações citadas na seção anterior, optamos por carregar dois conjuntos de dados distintos no protótipo. O primeiro deles com informações sumarizadas, no qual é feita uma contagem das observações nas categorias das variáveis do modelo e a contagem dos casos que estão naquela categoria que tiveram o evento de resposta do modelo (ver Figura 6.3). Neste banco também incluímos o número do modelo, indicativo se a informação é referente ao desenvolvimento ou a aplicação do modelo, nome das variáveis e suas respectivas categorias.

O segundo banco de dados contém as informações analíticas sem nenhuma agregação, ou seja, cada dado de interesse gerado pelo modelo. Isso foi necessário, pois o cálculo de alguns indicadores como KS2 e visualizações como gráfico de dispersão não poderiam ser construídos com os dados sumarizados. A Figura 6.4 mostra o layout do banco de dados analítico, no qual são incluídos os *Scores* gerados pelo modelo, informações exploratórias de interesse, variável resposta, data de referência e indicativo se o dado é referente ao desenvolvimento ou aplicação do modelo. Para fim de avaliação do protótipo, importamos os dados via código. Neste caso, é preciso carregar dois arquivos “csv” para cada modelo.

Variáveis	Tipo	Descrição
modelo	int	Indica número do modelo
des_apl	string	Indica se o dado é referente ao desenvolvimento
data	date	Data de referência
variavel	string	ID das variáveis preditoras que compõem o modelo
nome	string	Nome das variáveis preditoras que compõem o modelo
categoria	string	Categoria da variável preditora
n	int	Soma da quantidade de casos nas categorias
resposta	int	Soma de casos em que ocorreu o evento de resposta do modelo

Figura 6.3 – Layout de entrada de banco de dados sumarizados.

Como o VACS visa permitir a avaliação de diversos modelos em uma única interface, decidimos construir dois modelos de crédito para simular um cenário real de acompanhamento de modelos. No primeiro modelo, escolhemos um banco disponibilizado pela *U.S. Small Business Administration (SBA)*, devido ao grande volume de informações e por ele ter sido utilizado em decisões de crédito reais, o que é difícil já que as informações de IFs são extremamente sigilosas [LAF16]. No segundo modelo utilizamos o banco de dados *German*, pois é um conjunto amplamente utilizado em estudos de modelos de crédito [LAF16] e baseado em concessões reais. A seguir detalhamos os conjuntos de dados escolhidos.

Variáveis	Tipo	Descrição
período	date	Data de referência
modelo	int	Indica número do modelo
resposta	int	Flag indicando se ocorreu o evento de resposta do modelo
des_apl	string	Indica se o dado é referente ao desenvolvimento
Score	int	Probabilidade ou escore (saída do modelo)
preditora 1	string	Variável exploratória que deseja investigar
.	.	.
.	.	.
.	.	.
preditora n	string	Variável exploratória que deseja investigar

Figura 6.4 – Layout de entrada de banco de dados analíticos.

A SBA é uma empresa fundada em 1953, nos Estados Unidos, com o objetivo de auxiliar pequenas empresas a tomarem crédito. A SBA oferece garantias às IFs quando empresas atendidas pelo programa não conseguem honrar seus compromissos, atuando em parte do pagamento dessas dívidas. Empresas como FedEx e Apple já participaram do programa. O banco de dados disponibilizado contém informações de empresas que tomaram crédito utilizando como garantia a SBA, no período entre 1969 a 2014, contendo 899.164 observações, com 27 colunas. Utilizamos em nosso trabalho as concessões de crédito ocorridas no período entre janeiro de 2007 e abril de 2008 totalizando 81.786 observações. A escolha do período foi resultado da análise descritiva que mostra que em média os prazos dos contratos são maiores que cinco anos, ou seja, garantiremos na amostra apenas empréstimos que venceram até 2014 (último ano do banco de dados). Além disso, em média, há um maior número de contratos nesse período. Este banco de dados, cha-

mado aqui de BD1, contém informações como data de aprovação do crédito, data do início do contrato, tempo de duração, data em que foi identificada inadimplência, localização do banco, localização da empresa, etc.

O banco de dados *German*, que está disponível no UCI repository of Machine Learning Repository<sup>1</sup>, é amplamente utilizado em vários estudos de modelos de crédito. Ele contém 1.000 solicitações de empréstimos e 20 variáveis que descrevem as características do empréstimo. Neste banco, chamado aqui de BD2, também são utilizadas informações de concessões de crédito reais, que incluem idade do cliente, número de dependentes, número de contas existentes na instituição financeira, etc.

### 6.3 Modelos Preditivos

Com o objetivo de demonstrar o uso do protótipo, construímos um modelo para cada banco de dados. O modelo A utilizou o BD1 e o modelo B o BD2. Utilizamos regressão logística como técnica de modelagem para ambos os modelos. Apresentaremos de forma resumida nesta seção como os modelos foram construídos. Durante o processo de modelagem não focamos em otimizar os modelos comparando técnicas ou melhorando a performance, mas sim em gerar as saídas do modelo que serão entradas no protótipo.

Na construção dos modelos, utilizamos a metodologia proposta no trabalho de Selau e Ribeiro [SR09], que usaram uma sistemática para construção de modelos de previsão constituída de seis etapas: delimitação da população, seleção da amostra, análise preliminar, construção do modelo, escolha do modelo e passos para implantação. Os passos sobre escolha do modelo e implementação do mesmo não foram utilizados, pois não temos interesse em comparar técnicas. O agrupamento das categorias de variáveis foi realizado de acordo com o seu risco relativo, que mensura o risco associado de cada categoria das variáveis preditoras com a variável resposta. Considerando os cortes sugeridos no trabalho de [GGM12], agrupamos as categorias de acordo com o risco relativo, classificando nas seguintes classes: péssimo (0 a 0,5], muito ruim (0,5 a 0,67], ruim (0,67 a 0,9], neutro (0,9 a 1,1], bom (1,1 a 1,5], muito Bom (1,5 a 2,0], excelente (maior que 2). Todos os coeficientes do modelo A possuem significância estatística ( $p\text{-value} < 0,01$ ). Mas, no modelo B, como tínhamos menos informações no banco de dados para serem testadas e para maximizar o número de variáveis exibidas no protótipo, optamos por usar o máximo de variáveis possíveis, sem considerar, por exemplo, a significância estatística.

O modelo A possui oito variáveis, sete categóricas e uma contínua (“disbursementgross”) podemos observar quais são elas e qual sua relação com a variável resposta (ter pagado o empréstimo tomado) na Tabela 6.1. O sinal do coeficiente nos mostra se indivíduos que estão naquela categoria tem maior ou menor chance de terem pagado o

---

<sup>1</sup><https://archive.ics.uci.edu/ml/index.php>

empréstimo quando comparado com a categoria de referência (“ref”). Assim, por exemplo, é possível interpretar que empresas que possuem tipo de negócio rural são mais arriscadas, pois o coeficiente é negativo quando comparado a categoria de referência (negócio “urbano”). A variável com maior influência na predição da inadimplência foi referente ao prazo do empréstimo, pois possui coeficientes maiores em módulo. O modelo possui boa capacidade preditiva, com acurácia de 79% e KS2 de 58.

Tabela 6.1 – Variáveis do modelo preditivo A.

Categoria	Descrição	Coeficiente
bankstate_categrr2_1	CA, FL, SC	1,6
bankstate_categrr2_2	DE, UT, SD	-0,5
bankstate_categrr2_3	VA, NC, OR	0,8
bankstate_categrr2_4	ID, NH, GA, WA, CO	-1
bankstate_categrr2_5	OH, RI, MO, MS, PA, TX, WI, AL, VT, IA, MT, NJ, ME, TN, MI, KY, CT, MA, NM, NE, AR, IN, MN, MD, NV, KS, AK, LA, WY, OK, ND, AZ, HI, WV, DC	-0,9
bankstate_categrr2_6	outros	ref
naics_categrr3_1	81, 54, 23, 48, 49, 31, 33, 56, 72, 61, 71, 92, 22	-0,2
naics_categrr3_2	Código no Sistema de Classificação de Indústrias Norte Americanas 62, 11, 21	-0,9
naics_categrr3_3	outros	ref
term_catrr_1	de 85 a 135 meses	-3,6
term_catrr_2	de 66 a 84 meses	-2,4
term_catrr_3	de 49 a 65 meses	-1,1
term_catrr_4	de 1 a 48 meses	ref
noemp_cat3_1	Quantidade de funcionários de 1 a 4	0,1
noemp_cat3_2	5 ou mais	ref
urbanrural_1	Tipo de negócio rural	-0,6
urbanrural_2	urbano	ref
newexist_1	Tempo do negócio novo	-0,3
newexist_2	antigo	ref
revlinecr_cat3_1	Possui linha de crédito rotativa sim	-0,2
revlinecr_cat3_2	não	ref
disbursementgross_1	Valor desembolsado	$4 \times 10^{-6}$

Já o modelo B é constituído por seis variáveis, cinco categóricas e uma contínua (idade). Podemos observar quais são as variáveis e qual o seu relacionamento com a variável resposta (ser “good” no banco de dados) na Tabela 6.2. A interpretação dos coeficientes segue a mesma lógica do modelo A, o KS2 do modelo foi de 29 e a acurácia foi de 72%.

Para que a construção do protótipo ficasse mais próxima do cenário real citado nas entrevistas, onde o que é acompanhado é o desempenho do modelo na aplicação e não no desenvolvimento, foi necessário realizar algumas simulações nos dados após a construção dos modelos. Foram feitas simulações para quatro variáveis do modelo A, além da variável resposta e *Score* e uma para o modelo B. O principal ponto simulado é a aplicação do modelo, como apenas conseguimos treinar e testar o modelo no conjunto de dados apresentado, mas não o usar em situações reais, não poderíamos acompanhar sua performance ao longo dos meses. Optamos, então, por utilizar os resultados dos modelos

Tabela 6.2 – Variáveis do modelo preditivo B.

Categoria	Descrição	Coefficiente
duration_qtl_1	Duração	de 4 a 12 meses
duration_qtl_2		de 13 a 18 meses
duration_qtl_3		de 19 a 30 meses
duration_qtl_4		mais que 31 meses
housing_1	Tipo de residência	outros
housing_2		residência própria
housing_3		residência alugada
job_1	Emprego	highly skilled
job_2		unskilled and resident
job_3		skilled
purpose_1	Proposta	business
purpose_2		car
purpose_3		furniture/equipment
purpose_4		others
purpose_5		radio/tv
saving.accounts_1	Poupança	little
saving.accounts_2		missing
saving.accounts_3		moderate
saving.accounts_4		quite rich
saving.accounts_5		rich
age_1	Idade	0,01

desenvolvidos, simulando a data de desenvolvimento como “2020-01” e depois utilizamos os mesmos dados simulando a aplicação dos modelos nos seguintes meses: “2020-02”, “2020-03”, “2020-04”, “2020-05”, “2020-06” e “2020-07”. Logo, simulamos um cenário no qual desenvolvemos o modelo em janeiro de 2020 e ele é usado nas decisões de crédito entre fevereiro e julho de 2020.

No modelo A, simulamos o desenvolvimento do modelo em “2020-01” e demais meses como aplicação. Optamos por simular algumas situações adicionais citadas durante as entrevistas como gatilho para uma queda de desempenho no modelo (ver Capítulo 5). Simulamos também alterações na variável resposta e no *Score* nos meses entre fevereiro e julho de 2020, além de alterações nas variáveis preditoras. Já no modelo B, apenas replicamos os dados gerados em janeiro para os demais meses, ou seja, simulamos o mesmo desempenho após o modelo passar a ser utilizado nas decisões de crédito.

Na Tabela 6.3 apresentamos as situações simuladas em algumas variáveis do modelo A ao longo dos meses de aplicação (fevereiro a julho de 2020). Simulamos as seguintes situações relacionadas a performance do modelo: um deslocamento dos *Scores* para valores mais altos ao longo do tempo; e um aumento do percentual de resposta no geral, ou seja, veremos que em todos os decis de *Scores* há um aumento no % de resposta ao longo do tempo.

Já nas variáveis preditoras simulamos algumas situações que poderiam indicar que o modelo não está funcionando como deveria de acordo com o desenvolvimento: (a) Aumento no percentual de resposta; (b) Aproximação do percentual de resposta entre as categorias; (c) Oscilações nas quantidades das categorias; (d) Inversão no percentual de resposta; (e) Oscilações no percentual resposta.

Tabela 6.3 – Situações simuladas nos bancos de dados do modelo A.

Variável	Situação simulada
Score	Deslocamento dos Scores mais baixos para mais altos
resposta	Aumento do % de resposta
naics_categrr	(a) Aumento no % de resposta
noemp_cat	(a) Aumento no % resposta (b) Aproximação do % de resposta entre as categorias
bankstate_categrr	(c) Oscilações nas quantidades das categorias (d) Inversão no % de resposta (e) Oscilações no % resposta
new_exist_catrr	(a) Aumento no % resposta; (c) Oscilações nas quantidades das categorias; (d) Inversão no % de resposta

## 6.4 Funcionalidades do VACS

O VACS foi desenvolvido considerando a análise dos trabalhos relacionados (Capítulo 4) e as necessidades levantadas pelos especialistas de domínio (Capítulo 5). O seu principal objetivo é detectar perda de performance do modelo ou sinais que indiquem que o modelo não está aderente ao seu desenvolvimento. Para ilustrar as análises que podem ser feitas através das funcionalidades implementadas, utilizamos imagens referentes ao acompanhamento do modelo A.

O protótipo foi implementado usando Shiny<sup>2</sup>, pacote desenvolvido para usuários do Software R<sup>3</sup> que permite criar aplicações interativas em HTML. Ele apresenta quatro menus principais: o primeiro, “Resumo dos Modelos” (Figura 6.5-a), mostra os principais indicadores e visualizações de maneira resumida, com alertas sobre a performance dos modelos; o segundo e terceiro, “Modelo A” e “Modelo B” (Figura 6.5-b), exploram análises detalhadas dos modelos, tais como análise do *Score*, estabilidade das variáveis e indicadores; o quarto menu, “Catálogo” (Figura 6.5-c), traz informações adicionais dos modelos, tais como descrição dos indicadores apresentados, variáveis que o compõem e descrição de suas categorias. A seguir detalharemos cada uma destes menus e as visualizações disponibilizadas.

### 6.4.1 Resumo dos Modelos

O menu “Resumo dos Modelos” (Figura 6.5-a) possibilita ter acesso aos principais indicadores de performance e de estabilidade para os dois modelos analisados (Figura 6.5-

<sup>2</sup><https://shiny.rstudio.com/>

<sup>3</sup><https://www.r-project.org/>

e e f). As cores mudam conforme o estado do indicador: verde se está dentro do esperado, amarelo se está levemente fora do esperado e vermelho caso os indicadores estejam gravemente fora dos limites indicados na literatura. Assim, os usuários podem ter um panorama geral sobre a performance dos modelos. Além disso, incluímos uma seção de alerta para variáveis invertidas em cada modelo, indicando quais são elas em vermelho (Figura 6.5-g). Uma variável é considerada invertida quando há uma inversão nas suas categorias, por exemplo, quando passa de “excelente” para “mau”, ou vice-versa. O usuário pode selecionar qual mês de referência quer que seja exibido na tela (Figura 6.5-d). Por fim, adicionamos links rápidos aos nomes dos modelos “Modelo A” e “Modelo B” para possibilitar o acesso aos seus indicadores.

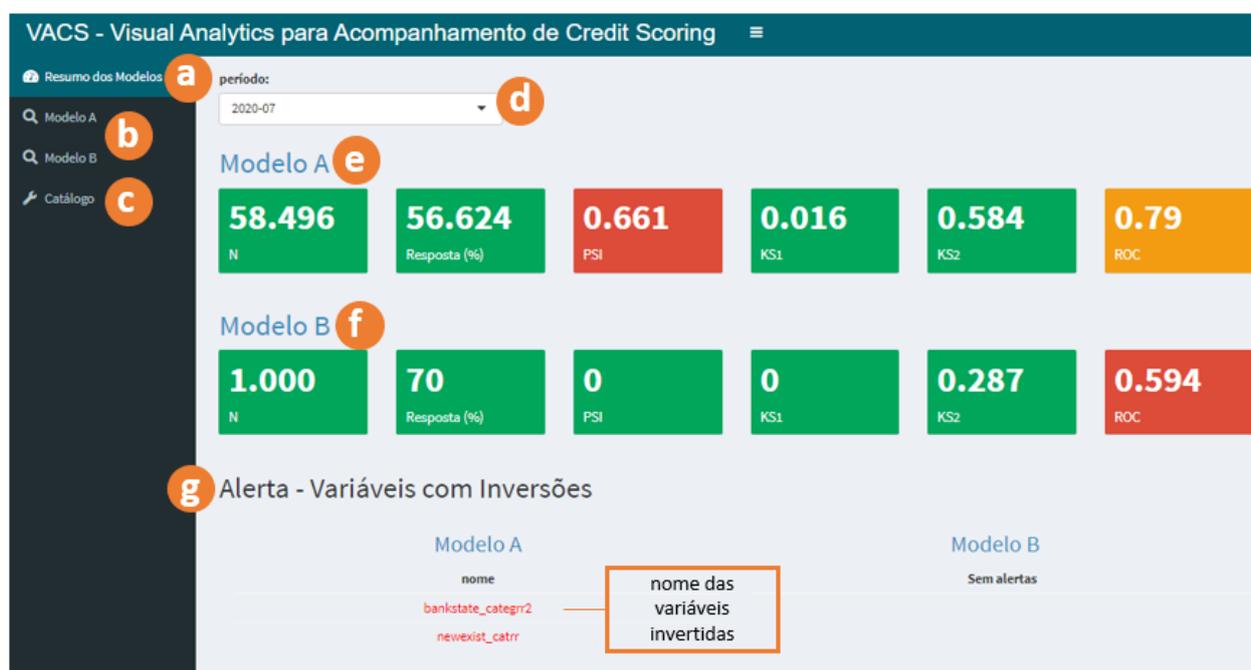


Figura 6.5 – Resumo dos modelos.

#### 6.4.2 Modelos

A seleção dos menus referentes aos modelos (A ou B) (Figura 6.5-b) abre um submenu com três opções para o modelo selecionado. A primeira opção, “Análise do *Score*” (Figura 6.6-a), dá acesso a seis abas com diversas visualizações que auxiliam nesta análise. A segunda opção, “Variáveis Predictoras” (Figura 6.9-a), apresenta duas abas com visualizações que auxiliam na análise de estabilidade das variáveis predictoras do modelo. A terceira opção, o submenu “Indicadores” (Figura 6.12-a), dá acesso a duas abas com tabelas dos indicadores referentes às variáveis predictoras do modelo e ao *Score* de saída do modelo.

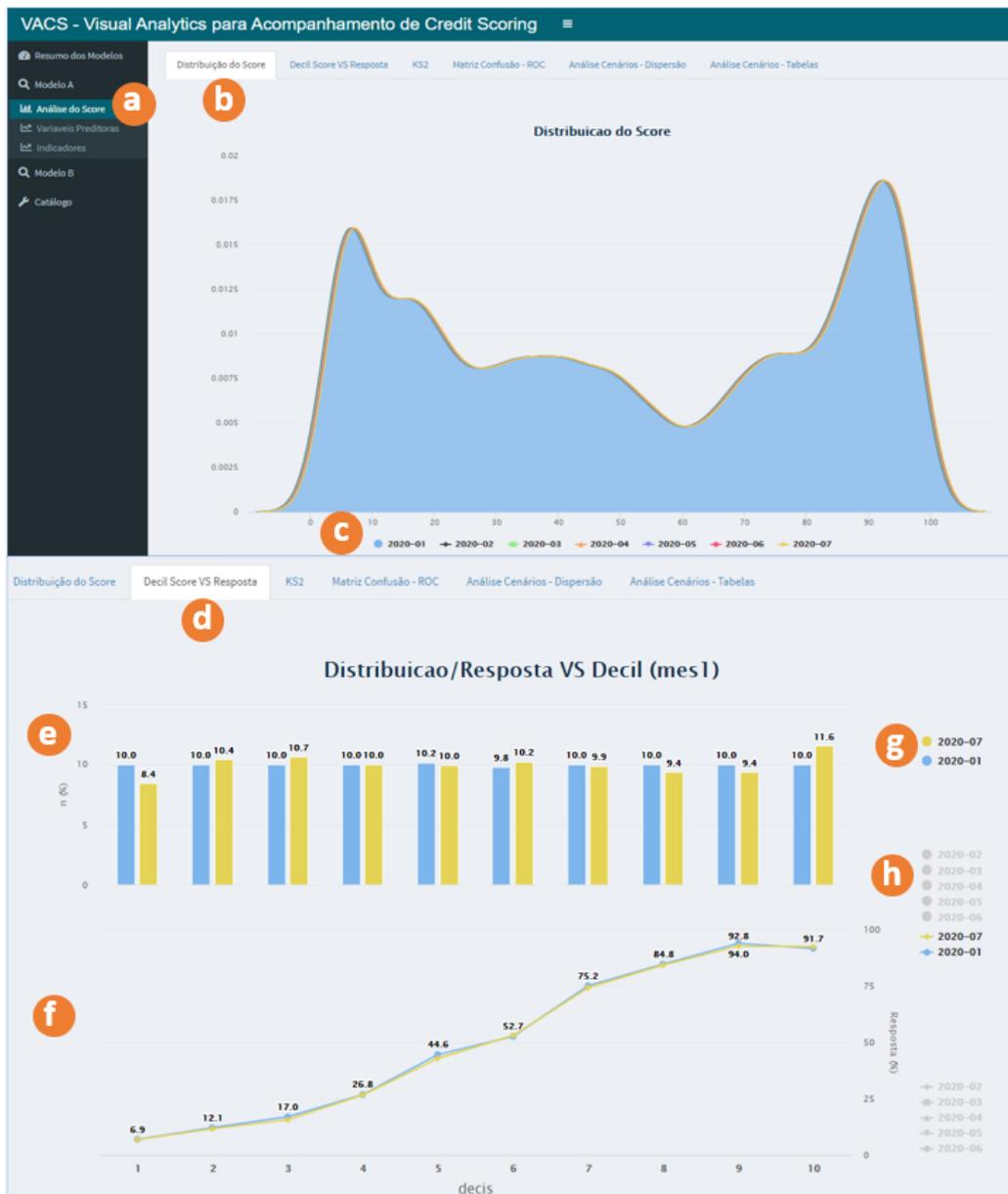


Figura 6.6 – Análise do Score - Distribuição do Score.

## Análise do Score

A aba “Distribuição do Score” (Figura 6.6-b) apresenta o gráfico da função densidade de probabilidade do *Score* para o período selecionado pelo usuário (Figura 6.6-c), com linhas de cores diferentes para cada mês (ver Seção 2.1.2). O usuário pode trocar a seleção a qualquer momento, sendo que o período de desenvolvimento do modelo é mostrado na cor azul para efeitos de comparação.

A seleção da aba “Decil Score VS Resposta” (Figura 6.6-d) dá acesso a um gráfico de barras que permite analisar a estabilidade do *Score*. Calculamos os cortes dos decis de *Score* marcados como desenvolvimento e aplicamos nos meses fora do desenvolvimento contidos na base de dados. Podemos verificar como as concentrações nos *Scores* do

modelo mudam (Figura 6.6-e) ao longo do período selecionado pelo usuário (Figura 6.6-g), comparando com o desenvolvimento. O percentual de resposta do modelo para cada decil de *Score* pode ser analisado no gráfico de linha (Figura 6.6-f), de acordo com os meses selecionados pelo usuário (Figura 6.6-h). Separamos as seleções dos meses em cada gráfico, pois acreditamos que as análises possam ser feitas independentemente uma da outra.

A aba “KS2” (Figura 6.7-a) mostra o gráfico do KS2 do modelo para o período selecionado (Figura 6.7-b), ilustrando a separação entre a distribuição acumulada dos marcados na base como “1” e a distribuição acumulada dos marcados como “0” (Figura 6.7-c). O gráfico ilustra em que *Score* ocorreu a maior separação entre as duas distribuições, o valor do indicador pode ser encontrado na aba “indicadores”. Importante destacar que o VACS pode ser utilizado no acompanhamento de qualquer modelo com resposta binária, uma vez que ele não distingue o que são clientes considerados como “bons” ou “ruins”. Nas entrevistas ficou claro que isso varia de acordo com a instituição financeira.



Figura 6.7 – Gráfico de KS2 e matriz de Confusão - Curva ROC.

A matriz de confusão do modelo é apresentada na aba “Matriz de Confusão - ROC” (Figura 6.7-d). Aqui o usuário pode selecionar em qual *Score* de predição ocorrerá o corte para marcação de resposta como “1”. Por exemplo, a partir do *Score* 75 gerado pelo modelo iremos marcar a observação como “bom”. Para isso, a barra de interação deve ser deslizada até o ponto de corte de interesse (Figura 6.7-e). Criamos essa funcionalidade para flexibilização da análise.

O usuário pode também selecionar o período que deseja avaliar (Figura 6.7-f), e para facilitar a seleção incluímos um botão que possibilita selecionar todos os meses ao mesmo tempo. A partir das seleções geramos a matriz (Figura 6.7-g) que irá comparar as predições com o que de fato ocorreu na resposta do modelo, ou seja, verificamos os verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos classificados pelo modelo. Além disso, apresenta também a curva ROC para o mesmo período de seleção da matriz (Figura 6.7-h), sendo que cada mês de referência possui uma cor distinta, para que as curvas possam ser comparadas.

A seleção da aba “Análise de Cenários - Dispersão” (Figura 6.8-a) dá acesso ao gráfico de dispersão (Figura 6.8-e) dos *Scores* de acordo com a variável selecionada (Figura 6.8-c), no período escolhido pelo usuário (Figura 6.8-b). Aqui também incluímos a opção de seleção de todas as datas ao mesmo tempo. Por uma questão de performance e para não “poluir” o gráfico, incluímos uma funcionalidade para seleção do tamanho da amostra (Figura 6.8-d), onde cada vez que o botão muda é gerada uma amostra aleatória sem reposição com o tamanho selecionado. Isso foi necessário, pois quando se tem muitos pontos neste gráfico ele pode demorar para carregar e dificultar a interpretação. Importante destacar que o eixo x do gráfico irá se adaptar ao tipo de variável selecionada na opção (Figura 6.8-c), sendo ela categórica ou contínua.

Como essa visualização busca informações do banco de dados analítico, as variáveis que se deseja analisar devem constar neste banco de dados. Essa foi uma visualização baseada na RSL, onde foi utilizada na análise de justiça dos *Scores* (ver Seção 4.2.2). O objetivo é analisar se os *Scores* preditos pelo modelo podem ter algum viés indesejado, por exemplo, se o modelo atribui valores de *Scores* menores para mulheres, o que no caso de modelos de crédito significaria que mulheres poderiam receber taxas de juros maiores que homens. Por isso o nome “Análise de justiça”, pretende-se avaliar se o modelo está sendo injusto com algum público específico.

Por fim, a seleção da aba “Análise de Cenários - Tabelas” (Figura 6.8-f) dá acesso a duas visualizações em formato de tabelas. A primeira contém o percentual de resposta do modelo (Figura 6.8-i) de acordo com as variáveis selecionadas nas linhas (Figura 6.8-g) e nas colunas (Figura 6.8-h). A tabela também é colorida por um mapa de calor, onde quanto maior os valores das colunas, mais forte a cor em verde, e quanto menor o valor, mais forte a cor em vermelho. O objetivo da introdução das cores foi facilitar a interpretação dos valores contidos na tabela, considerando o trabalho de [CHH<sup>+</sup>19] que utiliza o mesmo

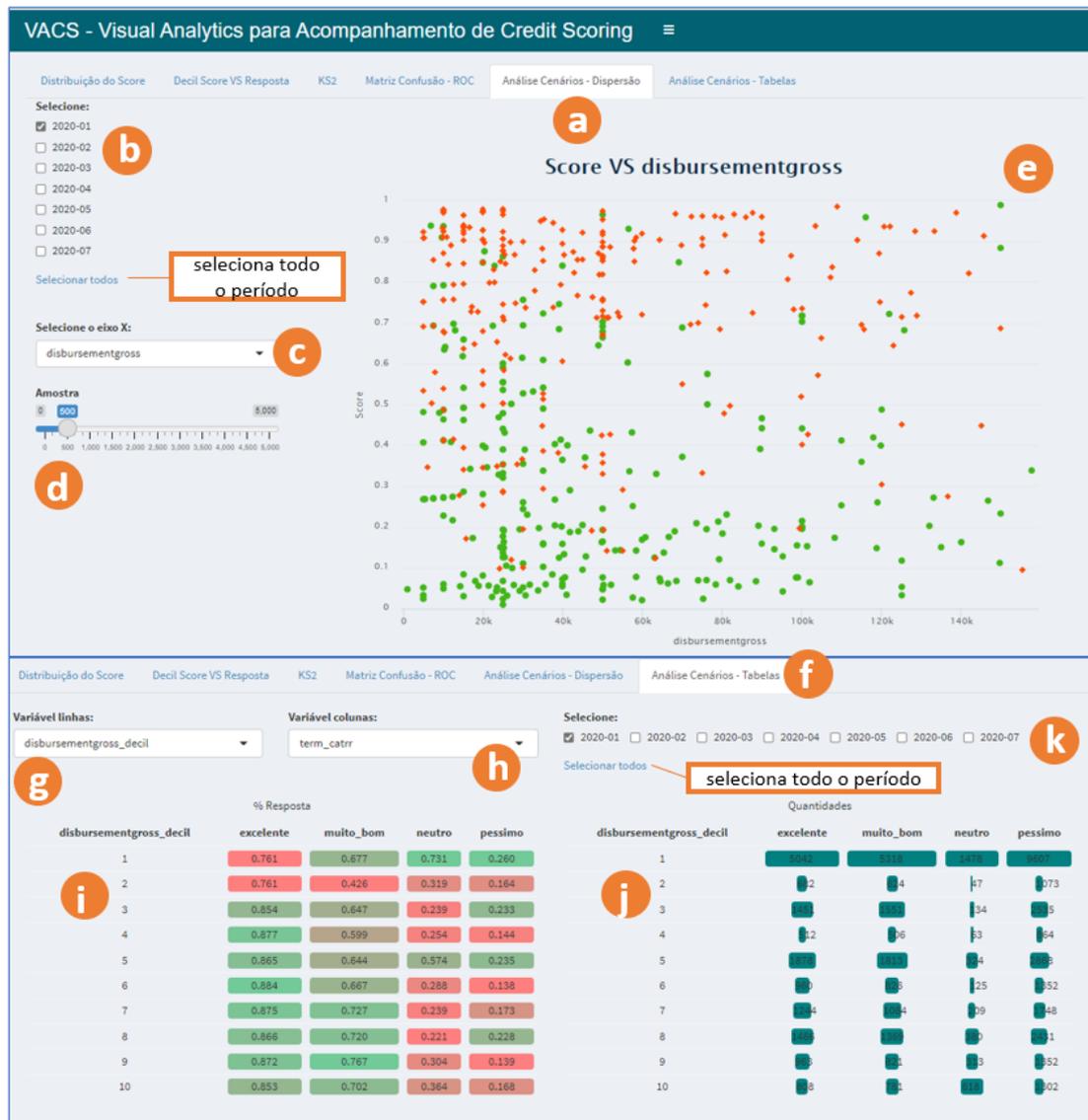


Figura 6.8 – Análise cenários dispersão e análise de tabelas.

conceito para deixar as matrizes de confusão mais intuitivas (ver Seção 4.2.1). A segunda tabela (Figura 6.8-j) contém as quantidades totais para as variáveis selecionadas nas linhas e nas colunas (Figura 6.8-g e h), permitindo verificar como os dados se distribuem entre as categorias de ambas as variáveis. Para facilitar a visualização das maiores concentrações dos dados, incluímos barras de cores, onde quanto maior o valor nas colunas maior o tamanho das barras. As interações do usuário neste menu mudam os dados para ambas as tabelas, além das colunas e linhas das tabelas o usuário pode também selecionar o mês de referência dos dados que será exibido (Figura 6.8-k).

## Variáveis Predictoras

O submenu “Variáveis Predictoras” disponível a partir dos menus de modelos do VACS (Figura 6.9-a), apresenta duas abas com visualizações que auxiliam na análise de

estabilidade das variáveis preditoras do modelo. Selecionando a aba “Variáveis Geral” (Figura 6.9-b) é possível visualizar todas as variáveis do modelo. Utilizamos o gráfico de barras empilhado para analisar a estabilidade da proporção do total nas categorias ao longo dos meses de referência (Figura 6.9-d). O usuário escolhe as categorias exibidas no gráfico, basta selecionar quais deseja retirar da visualização (Figura 6.9-c). A Figura 6.10-a apresenta um exemplo dessa interação, onde tiramos da seleção as categorias “muito\_mau”, “muito\_bom”, “bom” (Figura 6.10-a), logo são exibidas apenas as demais categorias.



Figura 6.9 – Variáveis preditoras - Análise geral.

O percentual de resposta para cada categoria apresentada no gráfico de barras empilhado é exibido no gráfico de linhas (Figura 6.9-f), que ajuda a verificar como foi a performance do modelo nas variáveis, comparando com o desenvolvimento (“2020-01”). É possível verificar, por exemplo, que a categoria “muito\_mau” além de aumentar ao longo dos meses ainda cruza com a categoria “mau”, ou seja, o percentual de resposta na categoria “mau” ficou maior que na categoria “muito\_mau”, algo que influencia o *Score* final calculado. Neste gráfico também é possível selecionar apenas as categorias de interesse (Figura 6.9-e). Na Figura 6.10-b podemos ver o gráfico apenas com as linhas das categorias selecionadas.

Para facilitar a análise e deixá-la mais interativa, incluímos *tooltips* em ambos os gráficos. Assim, ao passar o mouse na categoria ela aparece em destaque e com os valores contidos no ponto selecionado (Figura 6.10-c). Por fim, incluímos também a funcionalidade de *zoom*. Na Figura 6.10-d mostramos um exemplo de como ela funciona: ao arrastar o mouse nas categorias “bom” e “neutro” o gráfico foca apenas nelas. O usuário pode também desfazer o *zoom* no botão “Reset Zoom” que aparece automaticamente assim que o usuário aciona o *zoom* (Figura 6.10-e).

Através da aba “Variável Individual” (Figura 6.11-a), o usuário pode realizar a mesma análise de estabilidade e performance encontradas na aba “Variáveis Geral” (Figura 6.9-d e f). Todas as funcionalidades como *zoom*, *tooltips* e seleção das categorias exibidas também foram replicadas. Porém, o usuário realizará a análise de forma individual para cada variável de interesse selecionada (Figura 6.11-b)



Figura 6.10 – Variáveis preditoras - Análise geral - Exemplo de interação.



Figura 6.11 – Variáveis preditoras - Análise individual.



Figura 6.12 – Indicadores - Variáveis.

## Indicadores

O submenu “Indicadores” (Figura 6.12-a) disponível para o modelo selecionado através do menu principal do VACS, dá acesso a duas abas com tabelas dos indicadores referentes às variáveis predictoras do modelo e ao seu *Score* de saída.

Depois de selecionada esta opção, a aba “Indicadores - Variáveis” (Figura 6.12-b) apresenta uma tabela com os indicadores de entropia e *information value* para cada variável (Figura 6.12-d). As cores ajudam na interpretação dos valores, sendo que os limites utilizados para troca de cores foram atribuídos conforme a legenda (Figura 6.12-e) para o indicador *information value*. Além disso, utilizamos um degrade na cor azul para a entropia: quanto maior o valor, mais intensa é a cor. O usuário também pode selecionar quais datas de referência deseja analisar, sendo incluídas novas linhas para cada nova data selecionada (Figura 6.9-c). O botão “Selecionar todos” permite que o usuário analise todas as datas de referência ao mesmo tempo.

Na aba “Indicadores - *Score*” (Figura 6.13-a) trazemos os indicadores referentes ao *Score*, no âmbito de estabilidade e performance do modelo (Figura 6.13-c). Aqui os indicadores também apresentam cores de acordo com os limites da legenda (Figura 6.13-d). Os indicadores são mostrados nas colunas e, assim como nos indicadores das variáveis, as linhas aparecem de acordo com as datas de referências selecionadas (Figura 6.13-b).

Cada indicador possui seu método de cálculo e seus limites. Ambos foram definidos com base na literatura para os indicadores KS1 e KS2 [Sic10]. Utilizamos a *grey literature* para os seguintes indicadores: entropia<sup>4</sup>, *information value*<sup>5</sup>, *population stability*

<sup>4</sup><https://www.listendata.com/2015/04/decision-tree-in-r.html>

<sup>5</sup><https://www.listendata.com/2015/03/weight-of-evidence-woe-and-information.html>

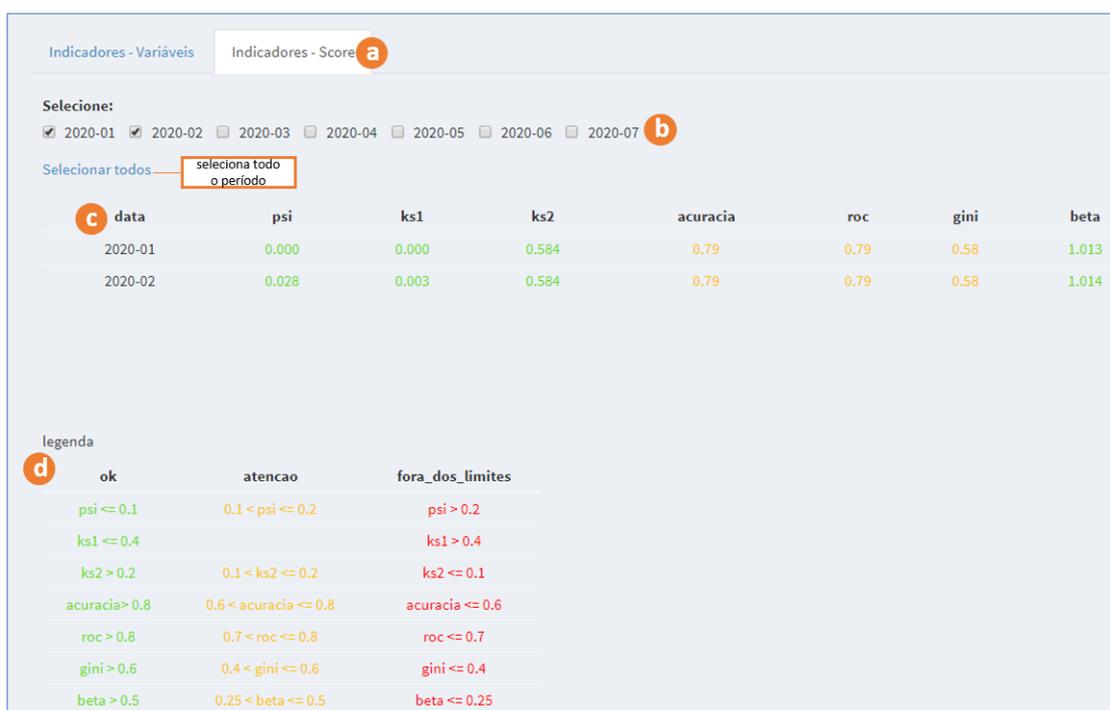


Figura 6.13 – Indicadores - Score.

*index*<sup>6</sup>, gini e acurácia<sup>7</sup>, beta<sup>8</sup> (coeficiente da regressão linear). A figura 6.16 apresenta a descrição de cada um dos indicadores.

### 6.4.3 Catálogo

O menu “Catálogo” (Figura 6.14-a) dá acesso a informações que auxiliam na sua interpretação. A aba “Variáveis - modelo 1” (Figura 6.14-b) possui uma tabela com a descrição de cada variável do modelo A e das categorias que o compõem (Figura 6.14-c). Já a aba “Variáveis - modelo 2” (Figura 6.15-a) traz as mesmas informações para o modelo B (Figura 6.15-b). Como o nome das categorias criadas durante o desenvolvimento podem não ser intuitivas, esse menu ajuda na interpretação do significado de cada uma delas.

Por fim, a aba “Descrição Indicadores” (Figura 6.16-a) apresenta uma tabela com a descrição dos indicadores utilizados (Figura 6.16-b). Isso irá ajudar principalmente a pessoas fora da área técnica na interpretação dos valores dos indicadores.

No próximo capítulo apresentamos os resultados da etapa de avaliação, através da qual buscamos entender se as necessidades levantadas nas entrevistas (ver Seção 5.7) foram atendidas. Os participantes puderam também dar suas opiniões e sugestões sobre as funcionalidades, visualizações, indicadores, alertas e interações construídas.

<sup>6</sup><https://www.listendata.com/2015/05/population-stability-index.html>

<sup>7</sup><https://www.listendata.com/2019/09/gini-cumulative-accuracy-profile-auc.html>

<sup>8</sup><https://researchwithfawad.com/index.php/lp-courses/basic-and-advance-data-analysis-using-smart-pls>

VACS - Visual Analytics para Acompanhamento de Credit Scoring

Resumo dos Modelos

Modelo A

Análise do Score

Variáveis Preditoras

Indicadores

Modelo B

Catálogo

Variáveis - modelo 1		Variáveis - modelo 2		Descrição Indicadores	
modelo	nome_variavel	descricao_variavel	categoria	descricao_categoria	
1	bankstate_catgrrr	UF de localização	excelente	CA, FL, SC	
1			muito_bom	VA, NC, OR	
1			bom	IL,NY	
1			mau	DE, UT, SD	
1			muito_mau	ID, NH, GA, WA, CO	
1			pessimo	OH, RI, MO, MS, PA, TX, WI, AL, VT, IA, MT, NJ, ME, TN, MI, KY, CT, MA, NM, NE, AR, IN, MN, MD, NV, KS, AK, LA, WY, OK, ND, AZ, HI, WV, DC	
1	naics_catgrrr	Código no Sistema de Classificação de Indústrias Norte Americanas	bom	outros	
1			neutro	81, 54, 23, 48, 49, 31, 33, 56, 72, 61, 71, 92, 22	
1			pessimo	62, 11, 21	
1	term_catrr	Prazo do empréstimo	excelente	de 1 a 48 meses	
1			muito_bom	de 49 a 65 meses	
1			neutro	de 66 a 84 meses	
1			pessimo	de 85 a 135 meses	
1	newexist_catrr	Tempo do negócio	mau	novo	
1			bom	antigo	
1	noemp_cat3	Quantidade de funcionários	[0,4]	de 0 a 4	
1			(4,11]	5 ou mais	
1	revlinecr_cat3	Possui linha de crédito rotativa	Y	sim	
1			N	não	
1	urbanruralrr	Tipo de negócio	muito_mau	rural	
1			bom	urbano	
1	disbursementgross	Valor de contrato	-	Variável Contínua	

Figura 6.14 – Catálogo - Variáveis modelo A.

Variáveis - modelo 1		Variáveis - modelo 2		Descrição Indicadores	
modelo	nome_variavel	descricao_variavel	categoria	descricao_categoria	
2	duration_qtl	Duração	1	de 4 a 12 meses	
2			2	de 13 a 18 meses	
2			3	de 19 a 30 meses	
2			4	mais que 31 meses	
2	housing	Tipo de residência	free	outros	
2			own	residência própria	
2			rent	residência alugada	
2	job	Emprego	0_1	unskilled and non-resident OU unskilled and resident	
2			2	skilled	
2			3	highly skilled	
2	purpose	Proposta	business	business	
2			car	car	
2			furniture/equipment	furniture/equipment	
2			others	others	
2			radio/tv	radio/tv	
2	saving.accounts	Poupança	little	little	
2			missing	missing	
2			moderate	moderate	
2			quite rich	quite rich	
2			rich	rich	
2	age	Idade	-	Variável Contínua	

Figura 6.15 – Catálogo - Variáveis modelo B.

Variáveis - modelo 1	Variáveis - modelo 2	Descrição Indicadores <b>a</b>
<b>Indicador <b>b</b></b>	<b>Descrição</b>	
Entropia	Determina a impureza da variável, ou seja, como as categorias são distintas entre si. Quanto mais próximo de 1 mais impuro (menos informação a variável agrega).	
IV - Information Value	Mensura o quanto uma variável é boa para distinguir uma resposta binária. Baixo valor de IV significa que ela pode não discriminar em um nível suficiente e pode ser removida como uma variável explicativa.	
PSI - Population Stability Index	Compara distribuições de Score de interesse com a distribuição de desenvolvimento do modelo.	
ks1 - Kolmogorov Smirnov	Avalia mudanças no público de desenvolvimento, comparando a distribuição de referência com a de desenvolvimento.	
ks2 - Kolmogorov Smirnov	Avalia a performance do modelo, comparando a distribuição acumulada dos clientes considerados como "bom" e a distribuição acumulada dos considerados como "mau".	
Acurácia	Avalia se o modelo classificou corretamente o cliente como "bom" ou "ruim".	
Gini	Mensura a performance do modelo, cálculo baseado no Area Under the Curve. Avalia as classificações geradas pelo modelo.	
Beta	Beta da regressão linear entre os Scores e o percentual de resposta médio encontrado para os Scores. Espera-se que esse valor fique próximo de 1.	

Figura 6.16 – Catálogo - Descrição dos indicadores.



## 7. VALIDAÇÃO DO PROTÓTIPO VACS

Neste capítulo, apresentamos os resultados obtidos na avaliação do protótipo construído. O objetivo da avaliação foi verificar se ele atende às necessidades levantadas durante as entrevistas com os especialistas e identificar novas sugestões de aprimoramento para o VACS. Além disso, na última seção apresentamos as contribuições e limitações do VACS e da pesquisa como um todo.

### 7.1 Metodologia e Perfil dos Participantes

Assim como na etapa de levantamento de requisitos (ver Capítulo 5), realizamos entrevistas semiestruturadas com questões abertas sobre o protótipo, para que os usuários pudessem dar suas opiniões de forma detalhada. O questionário aplicado nas entrevistas pode ser encontrado no Apêndice B.

Os participantes foram selecionados considerando a sua experiência com modelos de *Credit Scoring* (pelo menos dois anos). Foram entrevistados quatro participantes, sendo que dois deles participaram da primeira fase das entrevistas e dois deles não. Decidimos realizar as entrevistas com dois participantes que não haviam participado da primeira etapa, para avaliar a convergência das respostas sem o viés de participação no levantamento de requisitos.

Realizamos as entrevistas *online*, elas duraram em média 1 hora e 40 minutos. Como as entrevistas de *feedback* são mais longas que as de levantamento de requisitos, tivemos mais dificuldades para encontrar respondentes. Por exemplo, um dos participantes da primeira etapa, em virtude do tempo que precisaria reservar, desmarcou três vezes antes de cancelar a entrevista. Assim, optamos por um número menor de respondentes. Porém, como em pesquisas qualitativas, diferente de pesquisas quantitativas, o objetivo não é quantificar para generalizar os resultados, mas sim explorar a opinião dos entrevistados com uma maior profundidade, é comum um número reduzido de participantes [BR08].

As entrevistas foram divididas em quatro momentos. Primeiro, explicamos os objetivos do projeto e como a entrevista funcionaria. No segundo momento, apresentamos o VACS, explicando cada menu e como poderia ser feita a navegação e as análises. No terceiro momento, compartilhamos o link do protótipo<sup>1</sup> para que eles pudessem interagir e explorar o VACS, instigando-os a simularem a forma como monitoram modelos de crédito nas IFs. No quarto momento, aplicamos o questionário com questões sobre o perfil dos participantes, sobre o monitoramento de modelos e como seria a utilização do VACS nas IFs. Os participantes não carregaram nenhum conjunto de dados, eles exploraram os dois

---

<sup>1</sup>[https://daianealdo.shinyapps.io/demo\\_vacs/](https://daianealdo.shinyapps.io/demo_vacs/)

conjuntos de dados que já estavam disponíveis no protótipo. A Figura 7.1 apresenta o perfil dos participantes, referenciados aqui como P1, P2, P3 e P4.

	*P1	P2	P3	*P4
<b>UF de residência</b>				
RS	✓		✓	✓
SP		✓		
<b>Gênero</b>				
Feminino	✓			
Masculino		✓	✓	✓
<b>Idade</b>				
	32	26	35	40
<b>Graduação</b>				
Estatística	✓	✓		✓
Engenharia			✓	
<b>Pós-Graduação</b>				
Especialização	✓		✓	✓
Mestrado	✓			
<b>Tempo em IF's (anos)</b>				
	5,5	5	5	6,5
<b>Trabalha atualmente com modelagem de crédito?</b>				
sim		✓	✓	✓
não	✓			
<b>Tempo em modelagem (anos)</b>				
	4	2,3	2,4	6,5
<b>Quantas IFs trabalhou?</b>				
	1	2	1	1
<b>Cargo na empresa</b>				
Analista Pleno			✓	
Analista Sênior	✓	✓		
Especialista				✓

Figura 7.1 – Perfil dos respondentes, P1 e P4 participaram do levantamento de requisitos.

## 7.2 Resultados

Após interagirem com o VACS, questionamos os participantes sobre o seu uso no monitoramento e quais os ganhos ao comparar com a forma de acompanhamento que conhecem. Eles citaram que o protótipo ficou completo e permite realizar um monitoramento adequado, através da verificação da performance dos modelos e da análise das possíveis ações necessárias. Sobre os ganhos, eles destacaram a praticidade do acompanhamento já estar pronto, apenas sendo necessário a inserção dos dados. Ainda, comentaram sobre a padronização do acompanhamento, uma vez que as mesmas visualizações e indicadores são gerados para todos os modelos inseridos.

As funcionalidades disponíveis no menu “Resumo dos Modelos” foram citadas como ponto positivo por todos os participantes, pois assim conseguem obter alertas sobre os modelos utilizados pela IF. O participante P4 também citou como ganho alguns indicadores que ele não conhecia. Todos os participantes consideraram o VACS fácil de usar e que as análises podem ser realizadas de forma intuitiva, sem dificuldade na aprendizagem. O

participante P1 citou como dificuldade a funcionalidade de zoom utilizando o *touchpad* do *notebook* ao selecionar objetos do gráfico “arrastando” a seleção. Além disso, citou também que as datas utilizadas para a seleções de períodos de tempo poderiam estar ordenadas do mais novo para o mais antigo. O participante P3 citou como dificuldade os menus não estarem na ordem que ele utilizaria na análise.

Considerando como as visualizações e interações apresentadas no VACS ajudariam no monitoramento, P1 citou que os gráficos são bonitos e de fácil interpretação, o que favorece o uso em apresentações. P2 comentou sobre a facilidade na exploração de cenários, pois consegue avaliar hipóteses do motivo pelo qual o modelo não estaria performando adequadamente, avaliando efeitos de políticas de crédito e não apenas do modelo em si.

P3 citou como principal auxílio no monitoramento as visualizações permitirem comparar o desenvolvimento com a aplicação ao longo do tempo (ver Seção 6.3). P4 citou novamente a praticidade das visualizações já estarem construídas, sendo apenas necessária a inserção dos dados, e a primeira tela de resumo que já mostra um panorama geral, variáveis com categorias invertidas e principais indicadores. Perguntamos também sobre as funcionalidades, visualizações e indicadores presentes no VACS serem capazes de identificar de maneira eficiente a performance dos modelos. Todos os participantes responderam de forma positiva, justificando que a gama de indicadores, as análises de estabilidade e as visualizações utilizando cores permitiriam identificar de forma rápida qualquer desvio. Além disso, o participante P1 agregou que esse monitoramento poderia ser estendido para qualquer modelo com resposta binária.

Com relação a exploração de cenários e hipóteses em subpopulações, todos os participantes comentaram que as funcionalidades disponibilizadas na aba “Análise de Cenários - Dispersão” do submenu “Análise do Score” são muito úteis e que não haviam pensado nesse tipo de análise. Quando explicamos o conceito de análise de justiça, eles comentaram que fazia todo sentido, principalmente por que, segundo P1 e P2, pode haver vieses mesmo que as variáveis não estejam no modelo, por exemplo, idade e sexo. Além disso, poderiam utilizar para testar como os modelos estão distribuídos para diferentes públicos das IFs e avaliar pontos de corte nas políticas de crédito. P2 também comentou que isso gera ações propositivas na tomada de decisão, ao invés de apenas “olhar no retrovisor”. P2 e P4 citaram também a aba “Análise de Cenários - Tabelas” do submenu “Análise do Score” para testar hipóteses, sendo que P4 nunca tinha visto esse tipo de análise de segmentação e citou como ponto positivo. Já P3 disse que não utilizaria essa opção e sugeriu no lugar algum teste de calibragem dos modelos, avaliando os impactos nos *Scores* utilizando tabelas de comparação.

Dando sequência ao questionário, pedimos para que os participantes descrevessem como utilizariam o VACS para identificar quedas de performance do modelo ou falhas após sua implantação. P1 respondeu que caso o menu “Resumo dos modelos” não apresentasse nenhum alerta não precisaria verificar detalhadamente os demais menus. Caso

contrário, analisaria primeiro se foi algo pontual ou se há alguma tendência e depois faria as demais análises para identificar o que poderia estar ocorrendo. P2 citou que olharia a comparação entre o *Score* atual e o de desenvolvimento, verificando se há alguma mudança de comportamento, compararia a ordenação, verificando o KS2, além da análise exploratória. P3 olharia o menu “Resumo dos Modelos”, para verificar se algum indicador está com alerta, verificaria o KS2 e então se há algum alerta nas variáveis, analisando também o % de resposta por variável, para depois verificar a distribuição de *Score*, sempre comparando o atual com o desenvolvimento. Já P4 analisaria principalmente as variáveis, pois isso diz muito sobre o modelo, olharia também o menu “Resumo dos Modelos”, pois erros graves ou perda de performance significativa ficariam aparentes.

Também perguntamos sobre como as informações apresentadas no menu “Resumo dos Modelos” ajudariam no gerenciamento de modelos e se os participantes incluiriam mais indicadores. Todos responderam que ter uma página de resumo ajuda e que ela traz mais agilidade para o acompanhamento. P1 ainda ponderou que o que é apresentado no resumo guiaria como ele acessaria os demais menus do VACS e geraria *insights* para possível calibragem do modelo. P2 comentou que colocaria o menu projetado em uma TV para que toda a equipe pudesse estar ciente sobre a performance dos modelos, além disso, sugeriu a inclusão de uma comparação direta com o desenvolvimento. P3 gostou especialmente do alerta sobre variáveis com categorias invertidas, porém pontuou que se houvessem muitos modelos talvez o resumo não ficasse tão bom, mas que para poucos modelos atende. P4 ainda sugeriu a criação de um *Key Performance Indicator* (KPI) para estabilidade e para a discriminação que resumissem ainda mais a performance, pois caso a instituição financeira tenha muitos modelos ajudaria a resumir ainda mais.

Sobre o que mais gostaram no VACS, metade dos participantes citaram as informações apresentadas no menu “Resumo dos Modelos”. P1 citou a interatividade dos gráficos e das comparações, as visualizações serem “*Clean*”, não ter muitos botões em uma mesma página e de um mesmo botão já mudar a seleção para os diferentes gráficos na mesma página. P2 gostou da forma como o protótipo foi construído, da navegação dos menus, da interatividade, das opções apresentadas no menu “Resumo dos Modelos” e do monitoramento dos diferentes modelos estarem em uma única interface. Ele comentou que: “*É muito melhor que uma ferramenta estática, posso trocar os meses e investigar o que eu quero, é muito mais rápido que um Excel. O grande ganho é rapidez que eu consigo trocar de menu e acessar as informações, navegando pelo menu da esquerda*”. P3 também citou as funcionalidades apresentadas no menu “Resumo dos Modelos”, pois gera um conforto em não ter urgência de realizar o acompanhamento detalhadamente. P4 gostou das análises de cenários, principalmente da análise de justiça e segmentações que consegue fazer por meio das visualizações, pois foi uma novidade.

Perguntamos ainda se havia alguma análise apresentada no VACS que os participantes não conheciam e sobre como ela poderia contribuir no acompanhamento, e todos

os participantes citaram a análise de justiça como uma análise nova e importante. P1 citou a análise de ponto de corte por meio da matriz de confusão, disponível através da aba “Matriz de Confusão - ROC” no submenu “Análise do Score” e as informações disponibilizadas através da seleção do menu “Catálogo”, pois ajuda ter as descrições de forma rápida. Além disso, achou o VACS bem “democrático” por não ter relação com a técnica empregada, afirmando que é: *“Democrático, pois não está interessado na técnica, sendo resposta binário funciona, isso é um ponto importante”*.

P2 comentou que não conhecia o indicador *information value* e que ele pode ser bastante útil. P4 citou a análise de segmentações contida na aba “Análise de cenários - Tabelas” no submenu “Análise do Score”, pois é possível realizar cruzamentos específicos e investigações.

Quando questionados sobre como o uso de VA apresentado no VACS poderia contribuir no acompanhamento dos modelos das IFs, os participantes foram unânimes em afirmar que facilita e agiliza o trabalho. P1 e P4 citaram principalmente os alertas de cores utilizados, pois auxiliam a direcionar as ações a serem tomadas e na interpretação de pessoas não técnicas. P2 comentou que a padronização de cores ajuda nas apresentações dos resultados dentro da empresa, uma vez que permite uma familiarização com o significado das cores, facilitando as seleções nos gráficos. P3 gostou principalmente das visualizações gráficas e comentou que a visualização com cores ajuda muito, mas que os limites deveriam ser definidos pelas IFs. P4 acredita que as visualizações e as cores apresentadas no VACS são imprescindíveis na interpretação do acompanhamento.

Todos os participantes responderam que utilizariam o VACS no acompanhamento de modelos. As justificativas foram principalmente baseadas no protótipo ser bastante completo para as análises necessárias, a facilidade de navegação, a rapidez, a interatividade e o dinamismo, a Tabela 7.1 apresenta as respostas na íntegra. Sobre as dificuldades que o VACS encontraria para ser usado nas IFs, P1, P2, P3 citaram que devido a segurança da informação o VACS teria que ser homologado pelas IFs antes de ser usado e que isso poderia ser um obstáculo. P1 citou também que nem todas as áreas poderiam ter acesso a todos os menus, pois algumas informações como variáveis do modelo são sigilosas. P2 ainda comentou sobre preocupação com relação a base analítica, pois dependendo do tamanho poderia afetar na performance do VACS.

Questionamos os participantes sobre materiais e métodos relacionados ao acompanhamento dos modelos nas IFs e como os resultados da pesquisa podem ajudar nesse sentido. Todos responderam que o conhecimento sobre o monitoramento é interno das IFs e que não há materiais sobre o assunto. P1 citou que o material sobre os indicadores e sua descrição ajudaria a profissionais de áreas fora da modelagem de crédito entenderem sobre os assuntos que o time técnico das IFs apresentam, ajudando a difundir o conhecimento. P2 comentou que também pode auxiliar na padronização, pois cada instituição financeira foca em uma parte do monitoramento e traria *insights* de novas formas de acompanhamento.

Tabela 7.1 – Opiniões sobre utilização do VACS.

Partic.	Você gostaria de utilizar o VACS para o acompanhamento de modelos? Por quê?
P1	<i>“Com certeza. Acho que ele é fácil de usar, o que ajuda bastante, ele é de simples de visualização, não tem muita informação, muitas funcionalidades. Tem uma dimensão boa de qualidade, não ficou confuso empilhado ou com sobreposição, bem visual e fácil de navegar e traz praticidade para o dia a dia, de acompanhar no dia a dia testando com cenários diferentes, pensando em tamanho de amostra e simulação. Não tem dificuldade de uso em termos de clicks, arrastar, enfim usar a ferramenta é bem simples”.</i>
P2	<i>“Óbvio que sim, por que ele é fácil e rápido e indica de forma objetiva o que eu preciso melhorar e consigo de forma bem rápida responder questionamentos que podem me surgir, por exemplo, a distribuição de score está mudando? Consigo responder se estamos tendo algum preconceito no modelo. Conseguimos responder de forma rápida se tem alguma variável de políticas que está afetando o modelo. Além disso, a parte bem completa de variáveis e indicadores consigo de forma fácil”.</i>
P3	<i>“Sim, achei uma ferramenta bem interessante para acompanhar. Ele tem tudo que precisa para monitorar modelos, usaria”.</i>
P4	<i>“Sim, com certeza, é bem interessante abre um leque de possibilidades. Na minha experiência atual, já temos dashes de monitoramento, mas como no VACS tem funcionalidades diferentes que não costumo utilizar seria legal de testar e ver como funciona internamente”.</i>

Ele exemplificou que se sentiu motivado a começar a utilizar a análise de justiça que apresentamos no VACS. Como P3 trabalhou em apenas uma instituição financeira não se sentiu à vontade para opinar, mas comentou que encontrou dificuldades em encontrar materiais na literatura. P4 também comentou que pode trazer *insights* sobre como monitorar.

Por fim, perguntamos especificamente sobre sugestões para o aprimoramento do VACS, além de considerarmos as sugestões dadas ao longo do questionário ou quando os participantes estavam na etapa de interação com o VACS. Estas sugestões estão sintetizadas na Tabela 7.2.

Sobre a participação na pesquisa, os participantes elogiaram o projeto e o VACS, citando que a pesquisa poderia servir como guia no monitoramento de modelos. O P2 inclusive sugeriu o uso em disciplinas sobre modelos de *Credit Scoring*.

### 7.3 Contribuições e Limitações

A primeira contribuição deste trabalho foi a pesquisa com especialistas de domínio, pois pela falta de trabalhos relacionados ao tema de acompanhamento de modelos, acreditamos que a divulgação possa ser importante principalmente para IFs que estejam iniciando trabalhos com modelagem de crédito, ou até mesmo para auxiliar na padronização dos requisitos mínimos para o acompanhamento dos modelos de crédito nas instituições que já fazem uso desses modelos. A segunda contribuição foram as necessidades identificadas a partir das entrevistas com especialistas, que apoiaram a construção do VACS. As seguintes

Tabela 7.2 – Sugestões para aprimoramento do VACS.

Assunto / Menu	Sugestões
Botões	<ul style="list-style-type: none"> <li>- Incluir botão de voltar para a última tela. (P4, P3)</li> <li>- Datas na seleção (da mais recente para a mais antiga. (P1)</li> <li>- Quando selecionar as datas, elas mudarem de cor para cinza e não descer para parte inferior do gráfico. (P2)</li> <li>- Quando houver duas visualizações na mesma tela, ter botões de datas individuais. (P2)</li> <li>- Todos os botões pré-selecionados no último mês em todas as visualizações. (P3)</li> <li>- Pré-selecionar o desenvolvimento e último mês na aba "Decil Score VS Resposta". (P3)</li> </ul>
Indicadores	<ul style="list-style-type: none"> <li>- Análise gráfica ao longo do tempo. (P4)</li> <li>- Incluir indicadores sobre mudanças nas quantidades nas categorias de variáveis. (P1)</li> <li>- Mudaria os indicadores de <i>Score</i> para a primeira aba e de variáveis para a segunda. (P1, P3)</li> <li>- Link para análise de estabilidade das variáveis. (P2)</li> <li>- Mostrar o indicador <i>Information Value</i> ordenado do maior para o menor. (P2)</li> </ul>
Distribuição do <i>Score</i>	<ul style="list-style-type: none"> <li>- O gráfico de densidade de <i>Score</i> não ser suavizado. (P2)</li> <li>- Pintaria da mesma cor todos os meses do desenvolvimento. (P2)</li> </ul>
Análise cenários - Tabelas	<ul style="list-style-type: none"> <li>- Excluiria o menu. (P3)</li> <li>- Incluir possibilidade de visualizar as quantidades em percentual total de coluna e de linha. (P4)</li> </ul>
Análise cenários - Dispersão	<ul style="list-style-type: none"> <li>- Excluiria as cores de acordo com a resposta, para não ficar relacionado com performance. (P2)</li> </ul>
KS2	<ul style="list-style-type: none"> <li>- Incluir valor do indicador KS2. (P3)</li> </ul>
Matriz Confusão - ROC	<ul style="list-style-type: none"> <li>- Incluir valor do AUROC no gráfico da curva ROC. (P2, P4)</li> </ul>
Decil <i>Score</i> VS Resposta	<ul style="list-style-type: none"> <li>- Mudar formatação do rótulo de dados, pois dependendo da seleção não aparece os valores ou fica poluído. (P3)</li> </ul>
Variáveis Geral	<ul style="list-style-type: none"> <li>- Incluir alerta sobre inversões na análise de variáveis. (P3)</li> <li>- Alterar o atalho "Descrição Variáveis" para "Catálogo". (P4)</li> <li>- Gráfico de barras ir até 100%. (P3)</li> </ul>
Variáveis Individual	<ul style="list-style-type: none"> <li>- Incluir atalho para o menu com a descrição das variáveis. (P2)</li> </ul>
Resumo	<ul style="list-style-type: none"> <li>- Colocaria o símbolo de quando necessário. (P2)</li> </ul>
Catálogo	<ul style="list-style-type: none"> <li>- Trocar Modelo 1 e 2 para modelo A e B (padronização). (P3)</li> </ul>
Outros	<ul style="list-style-type: none"> <li>- Escala de cores para daltonismo. (P1)</li> <li>- VACS esperar vazio, para quando não tem a performance do modelo (análise real <i>timing</i>). (P1)</li> <li>- VACS pode ser para qualquer modelo de resposta binária. (P1)</li> <li>- Simulações com calibragem do modelo caso haja mudanças nas variáveis. (P2)</li> <li>- Colocaria o submenu de indicadores antes do submenu de variáveis preditoras. (P3)</li> </ul>

necessidades foram levantadas nas entrevistas: praticidade (N1), facilidade de uso (N2), explorar cenários e hipóteses (N3), alertas (N4), resumo dos modelos (N5) (ver Seção 5.7). Por fim, o protótipo VACS além de auxiliar no acompanhamento de modelos de *Credit Scoring* contribui com a análise de justiça neste acompanhamento, pois é um tema de extrema importância que encontramos na RSL e que não foi citado nas entrevistas. A seguir iremos discutir o que foi atendido pelo VACS em cada necessidade levantada pelos participantes.

Sobre a necessidade N1, o VACS não se limitou a nenhuma técnica de modelagem específica. Mesmo que tenhamos usado a regressão logística na construção dos modelos, qualquer outra técnica de modelagem pode ser aplicada, pois o que se acompanha são as entradas e saídas do modelo. Entretanto, modelos de *Deep Learning* não são recomendados para este contexto [GVBB<sup>+</sup>21b]. Os participantes da pesquisa de avaliação do protótipo comentaram sobre a praticidade na análise das variáveis de interesse, algo que buscamos na aba “Variável - Individual” do submenu “Variáveis Preditoras”. Outro ponto citado por eles foi a não necessidade de refazer todo o acompanhamento para cada novo modelo. Isso foi atendido parcialmente, pois a maior parte das funções estão genéricas para receber qualquer dado de entrada, porém foram necessários alguns ajustes personalizados para cada modelo. Por exemplo, o número de menus que aparece na interface deve se ajustar ao número de banco de dados inserido. Devido ao tempo não foi possível implementar essas funcionalidades, no entanto, a maior parte do VACS está genérica a qualquer banco de dados.

Com relação à necessidade N2, o VACS conseguiu atender todos os pontos levantados, com a utilização dos dados agregados na maior parte das visualizações geradas. O protótipo tem resposta rápida aos comandos. Durante o uso dos dados analíticos, na análise de justiça, incluímos ainda uma opção para os usuários trabalharem com amostras aleatórias. Limitamos o total dos dados a 5.000 observações para o gráfico não demorar para carregar a cada variável analisada. Utilizamos na maior parte das visualizações o pacote *highcharter*<sup>2</sup> do R, para que os gráficos fossem bonitos, fáceis e interativos. Além disso, sempre que fez sentido, incluímos nas visualizações a possibilidade de comparação com o desenvolvimento, sendo também possível fazer filtros para períodos específicos.

No que diz respeito à necessidade N3, três abas específicas do submenu “Análise do Score” atendem a essa necessidade: “Matriz de Confusão - ROC”; “Análise de Cenários - Dispersão” e “Análise de Cenários - Tabelas”. Aqui é possível testar alguns cenários, como o que acontece com o *Score* caso o ponto de corte seja alterado, ou como o *Score* está distribuído para variáveis específicas. Qualquer variável inserida no banco de dados analíticos aparecerá como opção de cruzamento, podendo ser utilizada para testes em subpopulações do modelo, por exemplo, públicos de rendas diferentes. Além disso, é possível fazer a análise de justiça dos modelos, basta o usuário colocar as variáveis de interesse, tais como sexo e raça, no banco analítico.

---

<sup>2</sup><https://cran.r-project.org/web/packages/highcharter/highcharter.pdf>

Sobre a necessidade N4, criamos alguns alertas de cores sobre a performance dos indicadores e sobre o cruzamento das variáveis nos menus do protótipo. No entanto, aqui há espaço para melhoria no VACS, pois os limites dos indicadores poderiam ser passados por parâmetro pelos usuários. Além disso, poderia ser incluída alguma funcionalidade que mandasse e-mails de alertas caso os indicadores estivessem fora dos limites de controle atribuídos.

Sobre a necessidade N5, dois menus foram criados para atender essa necessidade: “Resumo dos Modelos” e “Catálogo”. Aqui acreditamos que o VACS tenha atendido totalmente, pois foi construída uma visão gerencial com os principais alertas de ambos os modelos e uma seção com informações gerais que ajudam a entender os modelos e a interpretar os indicadores.

Percebemos ainda duas limitações no VACS: a conexão com o banco de dados e a homologação da ferramenta pela instituição financeira. Os usuários que utilizam conexão direta com o banco de dados teriam uma etapa adicional de exportação dos dados. Por isso acreditamos que a abordagem atenderá principalmente a modelos que já façam essa manipulação. Isso seria um ponto de atenção principalmente no caso de modelos que precisem ser acompanhados em tempo real, onde a conexão direta com o banco de dados é obrigatória. Como em vários momentos os participantes comentaram que os dados das IFs são extremamente sigilosos, acreditamos que antes de usar o VACS os analistas terão uma etapa de homologação com a equipe de segurança da informação da empresa. Portanto, a principal contribuição do VACS é a divulgação dos resultados e *insights* de análises que os analistas podem replicar no acompanhamento dos modelos dentro das IFs em que atuam. Isso também contornaria as limitações, já que eles poderiam fazer todas as conexões necessárias e não teriam uma etapa de homologação.



## 8. CONCLUSÃO

Inicialmente, apresentamos os resultados obtidos com a RSL sobre o uso de técnicas de visualização no acompanhamento de modelos de *Credit Scoring* com o objetivo de analisar trabalhos relacionados ao tema. Assim, foi possível perceber que a dificuldade da explicabilidade de modelos de aprendizado de máquina propiciou diversas pesquisas que mostraram valor no uso de técnicas de visualização para auxiliar tanto na modelagem, como na interpretação de resultados. Além disso, também foi possível perceber que essas técnicas de visualização ainda não foram aplicadas a modelos de *Credit Scoring*. Apresentamos ainda um estudo com 13 profissionais da área para auxiliar no levantamento das necessidades relacionadas a uma abordagem de VA que os ajudasse no acompanhamento de modelos de *Credit Scoring*. Descrevemos detalhadamente a metodologia utilizada e sintetizamos os resultados em cinco necessidades principais do usuário.

Com base nos resultados da RSL e das entrevistas com os especialistas de domínio propusemos o protótipo de uma abordagem de VA para o acompanhamento de modelos de *Credit Scoring*, chamado de VACS. A fim de avaliá-lo, construímos ainda dois modelos preditivos de crédito, para que fosse possível criar cenários de acompanhamento dos modelos que foram utilizados como entrada para o VACS. Na etapa de avaliação, entrevistamos quatro especialistas de domínio para que eles avaliassem o VACS no acompanhamento dos dois modelos criados. Foi possível verificar que, de forma geral, as necessidades levantadas durante a primeira parte das entrevistas foram atendidas com as funcionalidades construídas e que os usuários gostariam de utilizar o VACS no acompanhamento de seus modelos. Portanto, as principais contribuições deste trabalho são:

- Os resultados obtidos com a RSL que possibilitaram perceber que há uma lacuna sobre o tema na literatura, confirmada na pesquisa com especialistas. Com os trabalhos relacionados, avaliamos como as técnicas de visualização utilizadas na explicabilidade de modelos de aprendizado de máquina podem contribuir em modelos de *Credit Scoring*.
- As entrevistas com especialistas que permitiram fazer um levantamento das necessidades dos usuários e compreender que, apesar da falta de materiais na literatura, as IFs acompanham seus modelos de crédito e a maneira como esse acompanhamento deve ser feito é algo ensinado dentro das próprias instituições. Além disso, identificamos também as principais dificuldades na construção desses acompanhamentos, as necessidades dos usuários para uma ferramenta de acompanhamento e o perfil dos profissionais que trabalham na área.
- A criação do protótipo VACS que possibilitou colocar em prática o uso de VA no acompanhamento dos modelos de *Credit Scoring*. As interações, as técnicas de visualiza-

ções e as análises foram baseadas na RSL e nas entrevistas com especialistas. Assim, conseguimos exemplificar as funcionalidades necessárias para uma ferramenta de acompanhamento criada para esse fim. Introduzimos no protótipo o conceito de análise de justiça, algo que vem sendo bastante discutido no contexto de modelos de aprendizado de máquina e que acreditamos ter grande importância nas avaliações de modelos de crédito, independente da técnica de modelagem empregada. Além disso, conforme destacado durante as entrevistas, o VACS também poderia ser utilizado para o acompanhamento de modelos que não sejam de crédito, desde que eles tenham sua variável resposta binária.

- A avaliação do VACS através de novas entrevistas com especialistas de domínio que permitiu avaliar o protótipo, suas visualizações e funcionalidades, e como o que foi criado atende as necessidades levantadas na primeira etapa das entrevistas. Todas as melhorias comentadas ao longo da entrevista foram sumarizadas em uma lista de possibilidades para futuras abordagens que visem permitir o acompanhamento de modelos de *Credit Scoring*.

Sendo assim, podemos responder a nossa principal questão de pesquisa, pois concluímos que VA pode contribuir no acompanhamento de modelos de *Credit Scoring* por meio de um conjunto de visualizações interativas que: sejam fáceis de usar e rápidas de gerar; permitam testar hipóteses e explorar as variáveis de diferentes maneiras; e que consigam resumir um grande volume de dados e informações.

Além disso, acreditamos que esse é o primeiro trabalho na literatura que consegue apresentar como os acompanhamentos de modelos *Credit Scoring* são feitos, respondendo a QS1 que não havia sido respondida com a RSL. Os resultados obtidos com essa pesquisa podem ser utilizados tanto por instituições que estejam iniciando seus modelos de crédito, como para gerar *insights* de novas análises para as instituições que já trabalhem com esses modelos. Por fim, consideramos que também há potencial para uso acadêmico, pois, como citado nas entrevistas, um modelo de VA para este acompanhamento poderia ser introduzido em disciplinas sobre ensino de modelagem.

Na Seção 7.3 apresentamos as contribuições e limitações do VACS. Como trabalho futuro, pretendemos fazer a sua generalização para que seja possível adaptar-se aos diferentes modelos que as IFs possuem e que podem ser inseridos na ferramenta. Ainda, iremos corrigir e aplicar as sugestões feitas na etapa de avaliação, conforme descrito na Seção 7.2. Além disso, o VACS poderia ser testado em outros contextos fora da área crédito, como, por exemplo, modelos de propensão a compras de produto ou no ensino de disciplinas de modelos preditivos. Assim, poderíamos avaliar se os indicadores disponíveis são suficientes e se alguma outra visualização poderia ser incluída.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [AH14] Admati, A.; Hellwig, M. “The bankers’ new clothes: What’s wrong with banking and what to do about it—updated edition”. Princeton University Press, 2014, 424p.
- [AT19] Areosa, I.; Torgo, L. “Explaining the performance of black box regression models”. In: Proceedings of IEEE International Conference on Data Science and Advanced Analytics, 2019, pp. 110–118.
- [B+10] Bolton, C.; et al.. “Logistic regression and its application in credit scoring”, Ph.D. Thesis, University of Pretoria, 2010, 240p.
- [BCGL17] Bequé, A.; Coussement, K.; Gayler, R.; Lessmann, S. “Approaches for credit scorecard calibration: An empirical analysis”, *Knowledge-Based Systems*, vol. 134, Oct 2017, pp. 213–227.
- [BM15] Bravo, C.; Maldonado, S. “Fieller stability measure: a novel model-dependent backtesting approach”, *Journal of the Operational Research Society*, vol. 66–11, Nov 2015, pp. 1895–1905.
- [BR08] Bickman, L.; Rog, D. J. “The sage handbook of applied social research methods”. Sage Publications, 2008, 680p.
- [BWTK19] Bhombe, A.; Walukar, K.; Thakare, Y.; Kamble, S. “Comparative analysis of two BI tools: Micro strategy and tableau”. In: Proceedings of International Conference on Advancements in Computing & Management, 2019, pp. 559–565.
- [CB02] Casella, G.; Berger, R. L. “Statistical inference”. Duxbury, 2002, 688p.
- [CC17] Creswell, J. W.; Creswell, J. D. “Research design: Qualitative, quantitative, and mixed methods approaches”. Sage Publications, 2017, 404p.
- [CCDM19] Cardoso, V. R. d. S.; Campos, L. A.; Dantas, J. A.; Medeiros, O. R. d. “Fatores relacionados à liquidez estrutural dos bancos no Brasil”, *Revista Contabilidade & Finanças*, vol. 30–80, Apr 2019, pp. 252–267.
- [CG16] Chen, T.; Guestrin, C. “Xgboost: A scalable tree boosting system”. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
- [CHH+19] Cashman, D.; Humayoun, S. R.; Heimerl, F.; Park, K.; Das, S.; Thompson, J.; Saket, B.; Mosca, A.; Stasko, J.; Endert, A.; et al.. “A user-based

visual analytics workflow for exploratory model analysis”. In: Proceedings of Computer Graphics Forum, 2019, pp. 185–199.

- [DBH20] De Baets, S.; Harvey, N. “Using judgment to select and adjust forecasts from statistical models”, *European Journal of Operational Research*, vol. 284, Aug 2020, pp. 882–895.
- [DCP20] Dastile, X.; Celik, T.; Potsane, M. “Statistical and machine learning models in credit scoring: A systematic literature survey”, *Applied Soft Computing*, vol. 91, Jun 2020, pp. 106263.
- [DGR+21] Dunkerley, F.; Galai, K.; Rohr, C.; Gloinson, E. R.; Favaro, M.; Phillips, W. D. “Future of the credit information market”. Rand Corporation Publications, 2021, 70p.
- [DKKSA18] Demirguc-Kunt, A.; Klapper, L.; Singer, D.; Ansar, S. “The global index database 2017: Measuring financial inclusion and the fintech revolution”. World Bank Publications, 2018, 148p.
- [DLZ+19] Dodge, J.; Liao, Q. V.; Zhang, Y.; Bellamy, R. K.; Dugan, C. “Explaining models: an empirical study of how explanations impact fairness judgment”. In: Proceedings of the International Conference on Intelligent User Interfaces, 2019, pp. 275–285.
- [Eva11] Evans, T. “Cinco explicações para a crise financeira internacional”, *Revista Tempo do Mundo*, vol. 3–1, Apr 2011, pp. 9–30.
- [FHDT+12] Fabbri, S.; Hernandez, E.; Di Thommazo, A.; Belgamo, A.; Zamboni, A.; Silva, C. “Using information visualization and text mining to facilitate the conduction of systematic literature reviews”. In: Proceedings of International Conference on Enterprise Information Systems, 2012, pp. 243–256.
- [GC15] Gingerich, M. J.; Conati, C. “Constructing models of user and task characteristics from eye gaze data for user-adaptive information highlighting”. In: Proceedings of the Conference on Artificial Intelligence, 2015, pp. 1728–1734.
- [GGM12] Gouvêa, M. A.; Gonçalves, E. B.; Mantovani, D. M. N. “Aplicação de regressão logística e algoritmos genéticos na análise de risco de crédito”, *Revista Universo Contábil*, vol. 8–2, Apr 2012, pp. 84–102.
- [Gon05] Gonçalves, E. B. “Análise de risco de crédito com o uso de modelos de regressão logística, redes neurais e algoritmos genéticos”, Ph.D. Thesis, Universidade de São Paulo, 2005, 105p.

- [GVBB+21a] Gunnarsson, B. R.; Vanden Broucke, S.; Baesens, B.; Oskarsdottir, M.; Lemahieu, W. “Deep learning for credit scoring: Do or don’t?”, *European Journal of Operational Research*, vol. 295–1, Mar 2021, pp. 292–305.
- [GVBB+21b] Gunnarsson, B. R.; Vanden Broucke, S.; Baesens, B.; Óskarsdóttir, M.; Lemahieu, W. “Deep learning for credit scoring: Do or don’t?”, *European Journal of Operational Research*, vol. 295–1, Nov 2021, pp. 292–305.
- [GWGv19] Garcia Caballero, H.; Westenberg, M.; Gebre, B.; van Wijk, J. “V-awake: A visual analytics approach for correcting sleep predictions from deep learning models”, *Computer Graphics Forum*, vol. 38–3, Mar 2019, pp. 1–12.
- [Hai11] Hair, J. F. “Multivariate Data Analysis: An Overview”. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, chap. Multicollinearity, pp. 904–907.
- [Hay07] Haykin, S. “Redes neurais: princípios e prática”. Bookman Editora, 2007, 898p.
- [HL00] Hosmer, D. W.; Lemeshow, S. “Applied logistic regression”. Wiley, 2000, 518p.
- [JAC15] JANg, D.-H.; Anderson-Cook, C. M. “Firework plots for evaluating the impact of outliers and influential observations in generalized linear models”, *Quality Technology & Quantitative Management*, vol. 12–4, Apr 2015, pp. 423–436.
- [KKEM10] Keim, D.; Kohlhammer, J.; Ellis, G.; Mansmann, F. “Mastering the information age: Solving problems with visual analytics”. Eurographics Association, 2010, 518p.
- [KMF+17] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T. “Lightgbm: A highly efficient gradient boosting decision tree”, *Advances in Neural Information Processing Systems*, vol. 30, Dec 2017, pp. 4–9.
- [KS98] Kennedy, A. B. W.; Sankey, H. R. “The thermal efficiency of steam engines. Report of the committee appointed to the council upon the subject of the definition of a standard or standards of thermal efficiency for steam engines: With an introductory note. (Including appendixes and plate at back of volume).” Source: <https://doi.org/10.1680/imotp.1898.19100>, Jan. 1898.
- [LAF16] Louzada, F.; Ara, A.; Fernandes, G. B. “Classification methods applied to credit scoring: Systematic review and overall comparison”, *Surveys in Operations Research and Management Science*, vol. 21–2, Feb 2016, pp. 117–134.
- [LGL+19] Leonel, R.-A.; Gabriel, S.-P.; Linda, K. T.-M.; Hector, M. P.-M.; Jose, P.-P. “A graphical user interface for fast evaluation and testing of machine

learning models performance”. In: Proceedings of International Workshop on Biometrics and Forensics, 2019, pp. 1–5.

- [Lon03] Longhurst, R. “Semi-structured interviews and focus groups”, *Key Methods in Geography*, vol. 3–2, Jan 01, 2003, pp. 143–156.
- [MESB18] Mecca, M. S.; Eckert, A.; Santos, L. C.; Biasio, R. “Análise do impacto das taxas de cartões de crédito no resultado líquido de uma rede comercial varejista sob a perspectiva de diferentes regimes tributários”, *Revista Estudo & Debate*, vol. 25–3, Dec 2018, pp. 244–263.
- [MT12] Miranda, C.; Tiburcio, B. “A nova cara da pobreza rural: desafios para as políticas públicas”. Instituto Interamericano de Cooperação para a Agricultura, 2012, 200p.
- [NCC18] Neagoe, V.-E.; Ciotec, A.-D.; Cucu, G.-S. “Deep convolutional neural networks versus multilayer perceptron for financial prediction”. In: Proceedings of the International Conference on Communications, 2018, pp. 201–206.
- [RR11] Rezac, M.; Rezavc, F. “How to measure the quality of credit scoring models”, *Czech Journal of Economics and Finance*, vol. 61–5, Jan 2011, pp. 486–507.
- [SBMK18] Sopan, A.; Berninger, M.; Mulakaluri, M.; Katakam, R. “Building a machine learning model for the soc, by the input from the soc, and analyzing it for the soc”. In: Proceedings of IEEE Symposium on Visualization for Cyber Security, 2018, pp. 1–8.
- [SC10] Sousa, A. F. d.; Chaia, A. J. “Política de crédito: uma análise qualitativa dos processos em empresas”, *Revista de Gestão*, vol. 7–3, Jun 2010, pp. 197–209.
- [Sic10] Sicsu, A. L. “Credit Scoring: desenvolvimento, implantação, acompanhamento”. Blucher, 2010, 200p.
- [Sid12] Siddiqi, N. “Credit risk scorecards: developing and implementing intelligent credit scoring”. John Wiley & Sons, 2012, 196p.
- [SL21] Szepannek, G.; Lubke, K. “Facing the challenges of developing fair risk scoring models”, *Frontiers in Artificial Intelligence*, vol. 4, Oct 2021, pp. 1–9.
- [SR09] Selau, L. P. R.; Ribeiro, J. L. D. “Uma sistemática para construção e escolha de modelos de previsão de risco de crédito”, *Gestão e Produção*, vol. 16, Set 2009, pp. 398–413.

- [SS14] Schlechtingen, M.; Santos, I. F. “Wind turbine condition monitoring based on scada data using normal behavior models. part 2: Application examples”, *Applied Soft Computing*, vol. 14, Jun 2014, pp. 447–460.
- [TCE17] Thomas, L.; Crook, J.; Edelman, D. “Credit scoring and its applications”. Society for Industrial and Applied Mathematics, 2017, 264p.
- [TH19] Taplin, R.; Hunt, C. “The population accuracy index: A new measure of population stability for model monitoring”, *Risks*, vol. 7–2, May 2019, pp. 53.
- [Tho00] Thomas, L. C. “A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers”, *International Journal of Forecasting*, vol. 16–2, Apr 2000, pp. 149–172.
- [TMGL20] Troisi, O.; Maione, G.; Grimaldi, M.; Loia, F. “Growth hacking: Insights on data-driven decision-making from three firms”, *Industrial Marketing Management*, vol. 90, Oct 2020, pp. 538–557.
- [TSR+21] Tripathi, D.; Shukla, A. K.; Reddy, B. R.; Bopche, G. S.; Chandramohan, D. “Credit scoring models using ensemble learning and classification approaches: A comprehensive survey”, *Wireless Personal Communications*, vol. 123, Oct 2021, pp. 1–28.
- [VGB20] Veiga, F. D. S.; Gibran, S. M.; Bonsere, S. F. M. “Open banking: Expectativas e desafios para o mercado financeiro no brasil”, *Administração de Empresas em Revista*, vol. 1–15, Apr 2020, pp. 203–226.
- [WPB+19] Wexler, J.; Pushkarna, M.; Bolukbasi, T.; Wattenberg, M.; Viégas, F.; Wilson, J. “The what-if tool: Interactive probing of machine learning models”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 26–1, Oct 2019, pp. 56–65.
- [WZWY21] Wang, J.; Zhang, W.; Wang, L.; Yang, H. “Investigating the evolution of tree boosting models with visual analytics”. In: Proceedings of IEEE Pacific Visualization Symposium, 2021, pp. 186–195.
- [ZHC21] Zhou, J.; Huang, W.; Chen, F. “Facilitating machine learning model comparison and explanation through a radial visualisation”, *Energies*, vol. 14–21, Apr 2021, pp. 1–20.
- [ZWM+18] Zhang, J.; Wang, Y.; Molino, P.; Li, L.; Ebert, D. S. “Manifold: A model-agnostic framework for interpretation and diagnosis of machine learning models”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 25–1, Aug 2018, pp. 364–373.



## APÊNDICE A – LEVANTAMENTO DE REQUISITOS: QUESTIONÁRIO APLICADO

### Parte 1: Levantamento do Perfil.

1. Cidade e estado de residência?
2. Qual seu gênero?
3. Qual a sua idade?
4. Qual a sua formação (graduação e pós-graduação)?
5. Há quanto tempo você atua em instituições financeiras?
6. Você trabalha atualmente com modelagem de crédito?
7. Quanto tempo você trabalha ou trabalhou na área?
8. Em quantas instituições financeiras já trabalhou? Qual a UF de localização?
9. Qual o seu cargo dentro da empresa atual?

### Parte 2: Ferramentas, Dados, Técnicas de Visualização.

10. Você já trabalhou com ferramentas que fossem específicas para monitoramento de modelos de *Credit Scoring*?  
[SIM] Quais?  
[NÃO] Você tem conhecimentos de ferramentas com essa finalidade?  
[SIM] Quais?
11. Você já utilizou ferramentas para o monitoramento de modelos que não fossem específicas para esta finalidade?  
[SIM] Quais?  
[SIM] Essas ferramentas atenderam as necessidades esperadas para um monitoramento apropriado? Por quê?
12. Quais os formatos de dados geralmente são utilizados como entrada nas ferramentas que você utilizava para monitoramento?
13. Como é feito o pré-processamento dos dados antes de serem usados na ferramenta?
14. No monitoramento de modelos, são utilizadas visualizações que auxiliam na interpretação dos dados?  
[SIM] Quais?

15. Na sua opinião, quais tipos de análises e visualizações uma ferramenta de monitoramento de modelos deveria conter?
16. Na sua opinião, um Abordagem de *Visual Analytics* para o monitoramento de modelos contribuiria para o gerenciamento dos modelos de crédito da instituição financeira? Por quê?

**Parte 3: Atividades do Monitoramento de Modelos de *Credit Scoring*.**

17. De forma geral nas instituições que você trabalha ou já trabalhou, os modelos construídos são monitorados?
18. Você acha importante que os modelos sejam monitorados? Por quê?
19. Quais são as dificuldades encontradas na construção do monitoramento de modelos?
20. Quais são as etapas (atividades principais) de um monitoramento de modelos? Como garantir que o modelo atende o esperado?
21. Qual a periodicidade em que o monitoramento ocorre?
22. O monitoramento dos diferentes modelos acontece em uma mesma interface? Você acha importante que isso ocorra? Por quê?
23. Qual a relevância em acompanhar a distribuição dos dados antes que ocorra o desfecho do modelo?

**Parte 4: Indicadores.**

24. São utilizados indicadores de crédito de curto prazo no monitoramento dos modelos?  
[SIM] Quais?  
[NÃO] Você acha relevante que eles sejam acompanhados?
25. Como são identificados possíveis problemas após a implantação de um modelo?
26. Quais métricas estatísticas deveriam constar em um monitoramento?
27. Se você pudesse escolher qualquer funcionalidade em uma Abordagem de monitoramento de modelos, quais seriam elas? (Lista de desejos)
28. Gostaria de deixar algum comentário geral ou alguma sugestão que não estava no questionário?

## APÊNDICE B – *FEEDBACK VACS*: QUESTIONÁRIO APLICADO

### Parte1: Levantamento do Perfil

1. Cidade e estado de residência?
2. Qual seu gênero?
3. Qual a sua idade?
4. Qual a sua formação (graduação e pós-graduação)?
5. Há quanto tempo você atua em instituições financeiras?
6. Você trabalha atualmente com modelagem de crédito?
7. Quanto tempo você trabalha ou trabalhou na área?
8. Em quantas instituições financeiras já trabalhou? Qual a UF de localização?
9. Qual o seu cargo dentro da empresa atual?

### Parte 2: Questões relacionadas as necessidades dos usuários.

10. Como você utilizaria o VACS no acompanhamento de modelos de *Credit Scoring*? Como ele traria ganhos comparado aos métodos de acompanhamento que você conhece?
11. As visualizações/funcionalidades/indicadores presentes no VACS permitiriam identificar de maneira eficiente se os dados monitorados estão performando conforme o esperado? Justifique sua resposta.
12. Você achou o VACS fácil de usar? Quais funcionalidades apresentam maior e menor dificuldade? Justifique sua resposta.
13. De que maneira as visualizações e interações disponibilizadas no VACS ajudariam no monitoramento dos modelos?
14. Como você conseguiria explorar cenários e hipóteses em subpopulações de interesse?
15. Descreva como você utilizaria o VACS para identificar a queda de performance do modelo ou falhas após a implantação. Quais visualizações/funcionalidades/indicadores você utilizaria para isso?

16. Como o menu 'resumo dos modelos' ajudaria no gerenciamento dos modelos utilizados pelas IFs? Quais indicadores ou funcionalidades você incluiria?

**Parte 3: Perguntas Gerais.**

17. Quais as funcionalidades que você mais gostou? Por quê?
18. Havia alguma análise ou funcionalidade que você não conhecia? Como ela poderia contribuir no acompanhamento de modelos de *Credit Scoring*?
19. Você gostaria de utilizar o VACS para o acompanhamento de modelos? Por quê?
20. Haveria dificuldades para uso do VACS dentro das instituições financeiras? Quais e por quê?
21. O uso de *Visual Analytics* apresentado no VACS pode contribuir no acompanhamento de modelos utilizados nas IF's? Por quê?
22. Como os resultados da pesquisa contribuiriam para a forma como os modelos são monitorados? Na sua opinião, a forma como os modelos deveriam ser monitorados é amplamente difundido nas IFs? Por quê?
23. Quais sugestões você teria para aprimorar o VACS? Por exemplo, outros indicadores, visualizações ou funcionalidades.
24. Quais são os comentários que você gostaria de adicionar sobre o VACS e em relação a sua participação na pesquisa?

## APÊNDICE C – TERMO DE CONSENTIMENTO ENTREVISTAS

### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Nós, Dalane Rodrigues Baldo (aluna de mestrado) e Isabel Harb Manssour (professora orientadora), responsáveis pela pesquisa intitulada "Visual Analytics para o Acompanhamento de Modelos de Credit Scoring", estamos fazendo um convite para você participar como voluntário nesse estudo.

Esta pesquisa tem como objetivo criar uma abordagem de análise visual que auxilie no acompanhamento de modelos de Credit Scoring. Acreditamos que ela irá contribuir na gestão dos modelos, auxiliando as instituições financeiras na gestão de risco de crédito e tomada de decisão sobre a manutenção dos modelos.

Para sua realização será preciso fazer entrevistas seguindo um roteiro semiestruturado, pois mesmo nos baseando em estudos científicos sobre o tema, entendemos que a participação de profissionais com experiência nesta área pode enriquecer nosso trabalho, principalmente auxiliando no levantamento de atividades e necessidades. Por este motivo, realizamos o convite para sua participação, através de uma entrevista seguindo um roteiro semiestruturado.

Sua participação constará de forma voluntária e você poderá pedir para retirar-se deste consentimento ou solicitar o encerramento da gravação de áudio a qualquer momento sem penalidades ou perda de despesas decorrentes de sua participação. Importante ressaltar que o objetivo deste estudo não é avaliar o participante, mas sim, entender os processos de monitoramento de Credit Scoring e as ferramentas usadas para isso. O uso que faremos dos registros efetuados durante a entrevista é estritamente limitado a atividades acadêmicas e será garantido o seu anonimato e confidencialidade.

Entendemos que há riscos mínimos durante essa atividade como: Desconfortos ou aborrecimento ao responder o questionário durante entrevista; Divulgação de dados sigilosos sobre a instituição em que o respondente trabalha; Desconforto, constrangimento ou alterações de comportamento durante gravações de áudio. Com o objetivo de mitigar esses riscos, elaboramos um roteiro semiestruturado de questões, para otimizar o tempo com o entrevistado, não coletaremos dados associados a empresa de trabalho e informamos que a gravação de áudio será utilizada apenas para auxiliar na coleta de informações fornecidas durante a entrevista.

Você tem o direito de pedir uma indenização por qualquer dano que resulte da sua participação no estudo. Não há benefícios diretos e a curto prazo para os participantes. Porém, como disponibilizaremos online a Abordagem que será construída, os participantes poderão utilizá-la no futuro para o acompanhamento de seus modelos ou para insights sobre o uso de Visual Analytics nos acompanhamentos dos modelos já construídos.

Durante todo o período da pesquisa você tem o direito de esclarecer qualquer dúvida ou pedir qualquer outro esclarecimento, bastando para isso entrar em contato, com Dalane Rodrigues Baldo, no telefone celular (51) 99723-0137 ou Isabel H. Manssour no telefone celular (51) 99955-4948 a qualquer hora.

Outras informações importantes:

- Você tem garantido o seu direito de não aceitar participar ou de retirar sua permissão, a qualquer momento, sem nenhum tipo de prejuízo ou retaliação, pela sua decisão.
- As informações desta pesquisa serão confidenciais, e serão divulgadas apenas em eventos ou publicações científicas, não havendo identificação dos participantes, a não ser entre os responsáveis pelo estudo, sendo assegurado o sigilo sobre sua participação.
- Se por algum motivo você tiver despesas decorrentes da sua participação neste estudo com transporte e/ou alimentação, você será reembolsado adequadamente pelos pesquisadores.
- Ao assinar este termo de consentimento, você autoriza a gravação de áudio e não abre mão de nenhum direito legal que teria de outra forma.
- A entrevista tem um tempo estimado de 45 até 75 minutos para sua conclusão.

Caso você tenha qualquer dúvida quanto aos seus direitos como participante de pesquisa, entre em contato com Comitê de Ética em Pesquisa da Pontifícia Universidade Católica do Rio Grande do Sul (CEP-PUCRS) em (51) 33203345, Av. Ipiranga, 6681/prédio 50 sala 703, CEP: 90619-900, Bairro Partenon, Porto

Rubrica do participante

Rubrica do pesquisador resp.

Figura C.1 – Termo de consentimento utilizado nas entrevistas - Página 1/2.

Alegre – RS, e-mail: cep@pucrs.br, de segunda a sexta-feira das 8h às 12h e das 13h30 às 17h. O Comitê de Ética é um órgão independente constituído de profissionais das diferentes áreas do conhecimento e membros da comunidade. Sua responsabilidade é garantir a proteção dos direitos, a segurança e o bem-estar dos participantes por meio da revisão e da aprovação do estudo, entre outras ações.

Não assine este termo de consentimento a menos que tenha tido a oportunidade de fazer perguntas e tenha recebido respostas satisfatórias para todas as suas dúvidas. Se você concordar em participar deste estudo, você rubricará todas as páginas e assinará e datará duas vias originais deste termo de consentimento. Você receberá uma das vias para seus registros e a outra será arquivada pelo responsável pelo estudo.

Eu, \_\_\_\_\_, após a leitura deste documento e de ter tido a oportunidade de conversar com o pesquisador responsável, para esclarecer todas as minhas dúvidas, acredito estar suficientemente informado, ficando claro para mim que minha participação é voluntária e que posso retirar este consentimento a qualquer momento sem penalidades ou perda de qualquer benefício. Estou ciente também dos objetivos da pesquisa, dos procedimentos aos quais serei submetido, dos possíveis danos ou riscos deles provenientes e da garantia de confidencialidade e esclarecimentos sempre que desejar.

Diante do exposto expresso minha concordância de espontânea vontade em participar deste estudo.

\_\_\_\_\_  
Assinatura do participante da pesquisa ou de seu representante legal

\_\_\_\_\_  
Assinatura de uma testemunha

### **DECLARAÇÃO DO PROFISSIONAL QUE OBTVE O CONSENTIMENTO**

Expliquei integralmente este estudo ao participante ou ao seu cuidador. Na minha opinião e na opinião do participante e do cuidador, houve acesso suficiente às informações, incluindo riscos e benefícios, para que uma decisão consciente seja tomada.

Data: \_\_\_\_\_

\_\_\_\_\_  
Daiane Rodrigues Baldo  
Aluna de Mestrado em Ciência da Computação  
PPGCC – Escola Politécnica

\_\_\_\_\_  
Isabel Harb Manssour  
Professora Orientadora  
PPGCC – Escola Politécnica

Rubrica do participante

Rubrica do pesquisador resp.

Figura C.2 – Termo de consentimento utilizado nas entrevistas - Página 2/2.