

ANÁLISE EXPLORATÓRIA E PREDITIVA DE PADRÕES DE USO EM VIDEOCONFERÊNCIAS

¹MARCELO DE SOUZA

¹Universidade do Estado de Santa Catarina (UDESC)

<marcelo.desouza@udesc.br>

DOI: 10.21439/conexoes.v19.3671

Resumo. Ferramentas de videoconferência facilitam a comunicação, oferecendo recursos úteis como o compartilhamento de imagem, voz e arquivos. Sua popularidade intensificou-se nos últimos anos, uma vez que muitas atividades passaram a ser realizadas de forma remota com o suporte desse tipo de tecnologia. Na educação, ferramentas de videoconferência são usadas para aulas a distância, reuniões, palestras e eventos científicos. O ConferênciaWeb é uma ferramenta de videoconferência desenvolvida e mantida pela Rede Nacional de Ensino e Pesquisa (RNP). Ela foi projetada para atender às demandas dos setores de ensino e pesquisa, sendo usada por muitas instituições de ensino superior, professores, pesquisadores e estudantes de todo o país. Este trabalho tem por objetivo explorar os dados registrados durante o uso do ConferênciaWeb, analisando a atividade dos usuários e a qualidade da sua experiência durante o uso da ferramenta. Os dados foram obtidos junto à RNP, contendo registros de mais de 18 milhões de usuários dos últimos quatro anos. Após o pré-processamento e preparação dos dados, foram aplicadas técnicas de visualização para a análise exploratória dos dados. Com isso, foi possível extrair informações relevantes, como o comportamento dos usuários em relação ao uso dos recursos oferecidos pela ferramenta e padrões nas avaliações da experiência de uso. Além disso, foram explorados algoritmos de classificação para a predição da qualidade da experiência do usuário. Em particular, foram aplicados modelos de *random forest* com acurácia médias de até 86%, além de analisadas as variáveis mais importantes para a tarefa preditiva. Os resultados alcançados a partir desta pesquisa permitem extrair informações relevantes para a tomada de decisão tanto por parte dos desenvolvedores da ferramenta, quanto dos usuários responsáveis por sua aplicação em atividades de ensino e pesquisa.

Palavras-chave: análise exploratória de dados; visualização de dados; aprendizagem de máquina.

EXPLORATORY AND PREDICTIVE ANALYSIS OF USAGE PATTERNS IN VIDEOCONFERENCING

Abstract. Videoconferencing tools ease the communication, offering useful features like image, voice and file sharing. Its popularity has increased in the last years, since several activities began to be carried out remotely with the support of this kind of technology. In education, video conferencing tools are used for remote classes, meetings, lectures and scientific events. ConferênciaWeb is a videoconferencing tool developed and maintained by the National Network of Education and Research (RNP). It was designed to meet the demands of the teaching and research sectors, and has been used by several institutions, professors, researchers and students across the country. This work aims to explore the data recorded during the use of ConferênciaWeb, analyzing user activity and the quality of their experience while using the tool. The data was obtained from the RNP, containing records of more than 18 million users over the last four years. After pre-processing and preparing the data, we applied visualization techniques for exploratory data analysis. We have extracted relevant information, such as the user behavior in exploring the features offered by the tool, and patterns in user experience evaluations. Furthermore, we explored classification algorithms to predict the quality of user experience. In particular, we applied *random forest* models with average accuracies of up to 86%. We also analyzed the most important variables for the predictive task. The results achieved from this research allow extracting relevant information for decision-making both by the tool developers and the users responsible for its application in teaching and research activities.

Keywords: exploratory data analysis; data visualization; machine learning.



1 INTRODUÇÃO

Com o avanço da tecnologia, várias atividades acontecem de forma remota atualmente, muitas das quais realizadas presencialmente no passado recente. Esse cenário se intensificou com a pandemia de COVID-19, em que diversos setores tiveram que se adaptar a um novo contexto em que o contato social foi restringido. Como resultado, a popularidade de reuniões virtuais e plataformas de videoconferência aumentou de forma expressiva, permitindo que várias atividades pudessem ser realizadas sem contato presencial (Silence *et al.*, 2021; Hurst; Withington; Kolivand, 2022). Tais tecnologias são adotadas nas mais diversas áreas, como saúde (Hassan *et al.*, 2020; Haleem *et al.*, 2021; Barbosa *et al.*, 2021; Lieux *et al.*, 2021), justiça (Sourdin; Li; McNamara, 2020), eventos científicos (Sarabipour, 2020; Falk; Hagsten, 2021; Raby; Madden, 2021), pesquisa (Archibald *et al.*, 2019; Gray *et al.*, 2020) e educação (Quezada; Talbot; Quezada-Parker, 2020; Oliveira *et al.*, 2021; Bond *et al.*, 2021; Camilleri; Camilleri, 2022). Essa nova realidade aumenta o interesse no desenvolvimento e melhoria de soluções de videoconferência. Além disso, pesquisas em torno dessas ferramentas e suas aplicações têm ganhado destaque. Exemplos incluem o mapeamento das ferramentas existentes (Correia; Liu; Xu, 2020), discussões sobre a privacidade dos usuários (Sandhu *et al.*, 2023) e o desenvolvimento de diretrizes para a adoção dessas tecnologias (Rubinger *et al.*, 2020).

Um exemplo de ferramenta de videoconferência é o ConferênciaWeb¹, mantido pela Rede Nacional de Ensino e Pesquisa (RNP)². Trata-se de um serviço em nuvem cujo objetivo é apoiar a realização de aulas, reuniões e palestras de forma virtual. A RNP trabalha para fornecer recursos atrativos na ferramenta, visando oferecer uma boa experiência aos usuários. Exemplos incluem funcionalidades de interação instantânea e colaborativa, compartilhamento de arquivos e de tela. Além disso, a ferramenta possui integração com sistemas educacionais, como Moodle e EduPlay. O ConferênciaWeb foi concebido para atender às necessidades dos setores de ensino e pesquisa, tendo como público-alvo os usuários do sistema RNP, que inclui professores, pesquisadores e estudantes.

O ConferênciaWeb coleta uma série de dados durante seu uso e os armazena na forma de *logs*. São registradas as características dos usuários e suas interações com os recursos do sistema. Por exemplo, é possível saber se o usuário ligou sua câmera, interagiu através do bate papo ou compartilhou sua tela. Além disso, o sistema registra a nota que o usuário atribuiu para sua experiência com a ferramenta. Apesar de possuir um conjunto de dados vasto e com potencial de geração de informações de valor, a RNP não possui uma metodologia definida para sua exploração. Diferentes abordagens podem ser adotadas para a exploração desses dados, incluindo técnicas estatísticas, de visualização e de aprendizagem de máquina. A aplicação dessas técnicas tem o potencial de extrair informações relevantes, como o comportamento dos usuários, além de permitir a predição da sua experiência durante o uso da ferramenta.

Dante desse cenário, este trabalho tem por objetivo estudar padrões de uso dos clientes do ConferênciaWeb e predizer a experiência dos usuários. O estudo é conduzido em três etapas. Na primeira, são feitos a coleta e o pré-processamento dos dados. Em seguida, é realizada uma análise exploratória dos dados, analisando o comportamento dos usuários e extraíndo diversas informações de interesse. Finalmente, são aplicados modelos de classificação baseados em *random forest* (Breiman, 2001) para a predição da experiência dos usuários.

As informações extraídas da base de dados e os resultados dos modelos de predição são importantes para apoiar a tomada de decisões em relação à ferramenta. Entender o comportamento dos usuários é fundamental para identificar falhas, planejar melhorias e desenvolver novas funcionalidades. Com um modelo eficaz de predição da experiência dos usuários, é possível identificar automaticamente experiências insatisfatórias durante o uso da ferramenta. Tal capacidade é essencial para a criação de uma ferramenta adaptativa, que responde às necessidades dos usuários e fornece funcionalidades conforme sua experiência de uso. Além disso, esses modelos de predição permitem identificar as variáveis que determinam uma boa (ou má) experiência de uso, fornecendo um possível direcionamento dos esforços para a melhoria da ferramenta.

2 TRABALHOS RELACIONADOS

Esta seção discute os trabalhos relacionados, os quais aplicam técnicas de análise exploratória e visualização de dados, bem como modelos preditivos no contexto de ferramentas de videoconferência. Tal como esta pesquisa,

¹O ConferênciaWeb pode ser acessado em: <https://conferenciaweb.rnp.br>.

²O website da RNP pode ser acessado em: <https://www.rnp.br>.

ANÁLISE EXPLORATÓRIA E PREDITIVA DE PADRÕES DE USO EM VIDEOCONFERÊNCIAS

esses trabalhos analisam padrões de uso e o comportamento dos usuários, com foco em aplicações na educação e na melhoria dessas ferramentas.

He (2013) analisa a participação e o comportamento de estudantes em ambientes educacionais de transmissão de vídeos ao vivo. O autor analisa padrões de interação dos estudantes com o professor e outros participantes através dos recursos de perguntas e de bate papo, explorando a frequência de uso desses recursos e identificando os tópicos de discussão mais frequentes. He (2013) também analisa a correlação entre a atividade dos estudantes e sua nota final, mostrando que uma maior interação com a ferramenta de perguntas está relacionada a um melhor desempenho. Além disso, ele leva em consideração os diferentes cursos analisados e mostra que o comportamento dos estudantes varia conforme a área do conhecimento, tal como o impacto da sua atividade no desempenho final.

Ez-zaouia, Tabard e Lavoué (2020) propõem um *dashboard* para auxílio a tutores em ambientes educacionais baseados em videoconferência. As visualizações fornecidas pelo *dashboard* permitem que tutores identifiquem aspectos emocionais dos estudantes, tarefa desafiadora em ambientes de educação remota devido à dificuldade em desenvolver relações socioafetivas nessas condições. Além de usar as interações do tutor e do estudante com os recursos da videoconferência (e.g. compartilhamento de arquivos, mensagens de bate papo e uso de marcadores), a ferramenta analisa a imagem do estudante e reconhece emoções expressas por ele. Essas informações são usadas para construção do *dashboard*, fornecendo *feedback* ao tutor quanto às emoções identificadas no estudante durante a reunião. Nesse âmbito, Ez-zaouia, Tabard e Lavoué (2020) ressaltam que o uso da ferramenta contribuiu para que tutores tivessem maior consciência das emoções dos alunos, usando essa informação para planejamento de suas aulas.

Com uma proposta similar, Samrose *et al.* (2021) apresentam uma ferramenta de *dashboard* que apresenta informações sobre o comportamento dos usuários de videoconferências. Para isso, a ferramenta avalia as imagens e o áudio da reunião, identificando diferentes eventos de interesse, como os sentimentos e falas dos participantes. Com base nessas informações, a ferramenta cria *dashboards* personalizados para cada usuário, fornecendo uma visão geral e diferentes informações ao longo da reunião, incluindo índices de participação, sentimentos pessoais e métricas de concordância entre os participantes. Os autores discutem uma série de melhorias em ferramentas de videoconferência que podem ser alcançadas pelo uso do *dashboard* proposto. Os autores ainda reportam que o uso dos *dashboards* aumentou a efetividade de reuniões e a inclusão dos seus participantes, melhorando o entendimento das dinâmicas das reuniões.

Sandhu *et al.* (2023) exploram modelos de aprendizagem profunda para a classificação de vídeos extraídos de ferramentas de videoconferência. O objetivo é identificar indicadores que favoreçam o engajamento de estudantes durante o ensino remoto. Tais indicadores variam desde o uso de recursos tecnológicos (e.g. compartilhamento de tela e uso de softwares interativos) até o comportamento do professor (e.g. estímulo para que os estudantes interajam) e seus movimentos (e.g. gestos e expressões faciais). O modelo proposto atingiu desempenho satisfatório na classificação de vídeos segundo esses indicadores (acurácia superior a 70% na maioria dos cenários analisados), o que permite sua aplicação na análise de reuniões em tempo real, identificando a presença de comportamentos que contribuem para o engajamento dos estudantes.

Os trabalhos supracitados possuem algumas limitações em relação à abordagem proposta nesta pesquisa. Apesar de analisar o comportamento dos usuários em um ambiente de transmissão de vídeos, He (2013) explora um conjunto limitado de dados de interação dos usuários, e as técnicas propostas não são aplicadas em ambientes de videoconferência, os quais apresentam mais recursos de interação e, consequentemente, um conjunto de dados mais rico. Embora extraiam informações a partir da imagem e do áudio das reuniões, os trabalhos de Ez-zaouia, Tabard e Lavoué (2020) e Samrose *et al.* (2021) também se limitam a um conjunto restrito de dados de *log*. Além disso, a exemplo de He (2013), esses trabalhos não apresentam modelos de aprendizagem de máquina para predição de variáveis de interesse, como o engajamento ou desempenho dos estudantes. Sandhu *et al.* (2023) preenchem essa lacuna, porém o modelo de predição adotado pelos autores é específico para processamento de imagens, não sendo adequado para lidar com dados tabulares. O trabalho não explora, portanto, os dados de *log* da ferramenta de videoconferência.

Em contraste com as abordagens apresentadas, este trabalho propõe a exploração de dados de *log* de uma ferramenta de videoconferência, analisando padrões de uso através da análise exploratória e visualização desses dados. É usado um vasto conjunto de dados contendo informações detalhadas sobre a atividade dos usuários (e.g. uso dos recursos da ferramenta, interação com outros participantes e papel na reunião) e características das reuniões (e.g. duração e o *status* de gravação). Além disso, esses dados são explorados por modelos de aprendizagem de

máquina para a predição da experiência do usuário, com base na nota atribuída por ele em relação à sua experiência durante a reunião. Tal como os trabalhos apresentados, esses resultados podem ser explorados para implementar melhorias na ferramenta de videoconferência, bem como aumentar a qualidade da sua aplicação em contextos diversos. Finalmente, as avaliações experimentais descritas em todos os trabalhos apresentados foram realizadas com um conjunto restrito de reuniões e usuários (algumas dezenas). Por outro lado, este trabalho explora uma base de dados de uma ferramenta de videoconferência consolidada, contendo reuniões realizadas ao longo vários anos e mais de 18 milhões de registros de usuários.

3 METODOLOGIA

O trabalho está dividido em três etapas principais. A primeira consiste na coleta e pré-processamento dos dados (Seção 3.1), seguida da análise exploratória e visualização dos dados (Seção 3.2), e finalizando com a predição da experiência de uso (Seção 3.3). A coleta, o pré-processamento e a análise descritiva dos dados foram realizadas usando a biblioteca *pandas*. A visualização dos dados foi realizada usando a biblioteca *matplotlib*. Os modelos de predição foram desenvolvidos e avaliados usando a biblioteca *Scikit-learn*.

3.1 Coleta e pré-processamento dos dados

Os dados foram anonimizados e disponibilizados pela RNP através de arquivos estruturados. No processo de coleta, os dados de diferentes arquivos foram mesclados, produzindo duas bases de dados distintas. A primeira delas contém os dados dos usuários de cada reunião, enquanto a segunda armazena as informações das reuniões. Ainda durante a coleta, dados que não são de interesse para o estudo foram descartados, como o navegador ou idioma do sistema usados pelo participante. Também foram excluídos os dados de participantes do tipo robô. Após a coleta e organização das bases de dados, foi realizado um pré-processamento dos dados, a fim de prepará-los para as etapas seguintes.

O primeiro passo do pré-processamento dos dados consiste na sua limpeza. Para isso, inicialmente foram identificados e tratados dados faltantes. Algumas reuniões não possuem informação (i.e. elas não foram registradas). Neste caso, dados de usuários associados a essas reuniões foram excluídos da base de dados. Também foram excluídos dados de usuários sem seus horários de entrada ou saída, e dados a respeito do uso de recursos (e.g. microfone, câmera) sem os horários de início ou fim do uso. Para finalizar, foram excluídos dados de reuniões com duração inferior a cinco minutos, e de usuários com tempo de permanência na reunião inferior a um minuto. Com essa limpeza, todos os registros possuem dados completos e relevantes para a pesquisa, incluindo as informações da reunião e os tempos de permanência dos usuários e de uso dos recursos.

Uma vez realizada a limpeza, foram criadas novas variáveis a partir dos dados coletados. Foi computada a duração de cada reunião e o tempo de permanência de cada usuário, bem como seu tempo percentual de permanência em relação à duração da reunião. Também foram computados o número de usuários, o número e percentual de moderadores e visualizadores, a nota média atribuída pelos usuários de cada reunião e o percentual de usuários que avaliaram ou não sua experiência. Uma vez que a atividade dos usuários é um dos principais interesses deste estudo, foram criadas diversas variáveis a partir dos registros de interação com os recursos da ferramenta. Para cada recurso (câmera, microfone, compartilhamento de tela, apresentação e levantar a mão) foram computados a quantidade de vezes que o usuário usou o recurso, o tempo total em que ficou ativado, e o tempo percentual em que ficou ativado em relação ao tempo de permanência do usuário na reunião. Além disso, para cada reunião foram computadas as distribuições do tempo relativo de permanência dos usuários e do tempo relativo de uso de cada recurso.

A Tabela 1 lista as variáveis das bases de dados de usuários e reuniões. As variáveis criadas a partir dos dados coletados são destacadas com asterisco. Note que vários dados são apresentados de forma resumida pela tabela. Por exemplo, a variável “uso de cada recurso” se refere à quantidade e aos tempos (absoluto e percentual) de uso dos diferentes recursos que a ferramenta oferece. Logo, esse item corresponde a diversas variáveis da base de dados de usuários. O mesmo se aplica a outras variáveis, como as distribuições de tempo de permanência e uso de cada recurso na base de dados de reuniões.

Para finalizar, a variável que define o papel do usuário foi transformada de categórica para numérica. Apesar de não ser necessário, tal procedimento é recomendado para a aplicação de modelos de *random forest*. Para isso, foi adotada a técnica de *one-hot encoding*, que substitui a variável categórica em questão (papel) por duas variáveis

Tabela 1: Variáveis das bases de dados de usuários e reuniões.

Usuários	Reuniões
Identificador	Identificador
Horários de ingresso e saída	Horários de criação e término
Papel (moderador ou visualizador)	Reunião gravada ou não
Nota atribuída pela experiência	Uso de recursos adicionais
* Tempo de permanência	* Duração
* Tempo percentual de permanência	* Número de moderadores e visualizadores
* Uso de cada recurso	* Percentual de moderadores e visualizadores
	* Número de usuários
	* Tempo percentual de permanência
	* Uso de cada recurso
	* Nota atribuída pela experiência
	* Percentual de usuários com e sem nota
	* Distribuição do tempo de permanência
	* Distribuição de uso de cada recurso

numéricas, uma para cada possível valor da variável original (moderador ou visualizador). Caso o usuário tenha o papel de moderador, a variável correspondente recebe o valor 1, enquanto a outra variável recebe o valor 0. O contrário acontece quando o usuário tem o papel de visualizador.

3.2 Análise exploratória e visualização dos dados

A etapa de análise exploratória e visualização dos dados tem três objetivos principais: (i) conhecer e entender os dados, obtendo informações gerais sobre as variáveis e a tendência dos dados; (ii) extrair informações de interesse para o suporte à tomada de decisões, como o comportamento dos usuários; e (iii) fornecer *insights* para a aplicação de modelos para a predição da experiência dos usuários. Esta etapa está dividida em dois passos: análise descritiva e análise gráfica.

Na análise descritiva (Igual; Seguí, 2017), foram calculadas estatísticas descritivas para as variáveis, como medidas de tendência central (média, mediana e moda), dispersão (amplitude, variância e desvio padrão) e percentis. Também foram analisadas as distribuições de variáveis de interesse, como aquelas relacionadas ao uso dos recursos e às avaliações dos usuários. Com isso, foram identificados comportamentos interessantes nos dados. Como exemplo, a maioria dos usuários não interage com os recursos da ferramenta. Pouco mais de 20% dos usuários ativa seu microfone alguma vez durante a reunião. Esse percentual diminui para cerca de 10% quando analisadas as ativações de câmera, e cai para menos de 8% para os recursos de apresentação, compartilhamento de tela e levantar a mão.

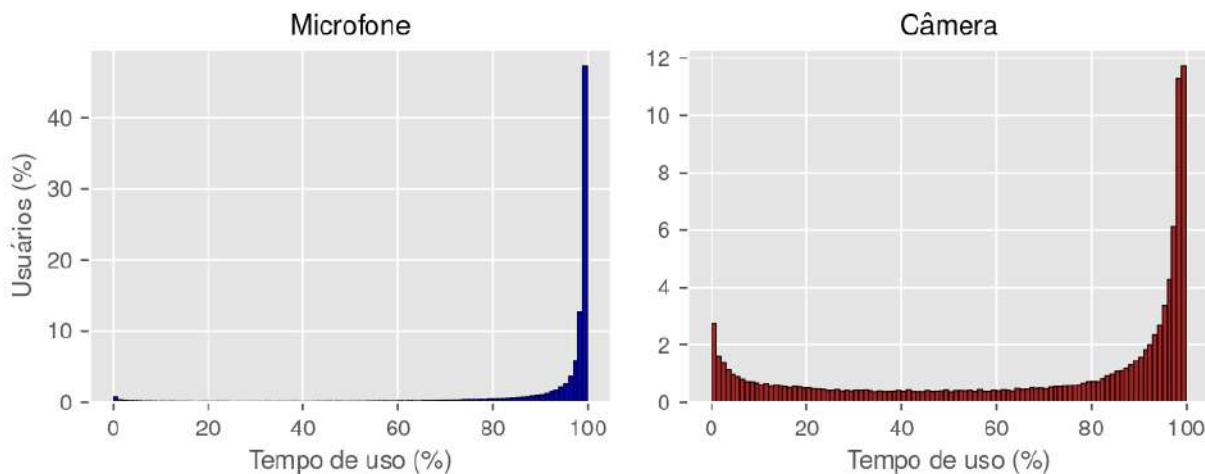
Na análise gráfica (Chen; Härdle; Unwin, 2007), os dados são explorados em maior profundidade, a fim de extraír informações relevantes sobre a ferramenta, a atividade dos usuários e sua experiência de uso. Neste passo, as distribuições de variáveis foram analisadas visualmente, facilitando a identificação de comportamentos interessantes nos dados, os quais foram investigados em maiores detalhes. Além disso, as visualizações (e.g. histogramas e diagramas de dispersão) foram exploradas para analisar correlações entre variáveis, permitindo estudar o comportamento dos usuários em função de diferentes características das reuniões.

3.3 Predição da experiência do usuário

Nesta etapa, foram construídos modelos de *random forest* (Breiman, 2001) para a predição da experiência do usuário. Em vez de considerar somente a nota atribuída ao final do uso da ferramenta, este trabalho se baseia nos conceitos do *Net Promoter Score* (NPS), uma abordagem bastante difundida para avaliar a experiência e a satisfação de clientes³. O NPS foi proposto por Reichheld (2003) e busca medir a lealdade dos clientes avaliando o resultados de pesquisas de satisfação onde eles atribuem um valor de 0 a 10 para a chance de indicar a empresa, produto ou

³O NPS é usado pela RNP para avaliação contínua da qualidade dos seus produtos.

Figura 1: Distribuição dos tempos percentuais de uso do microfone e da câmera.



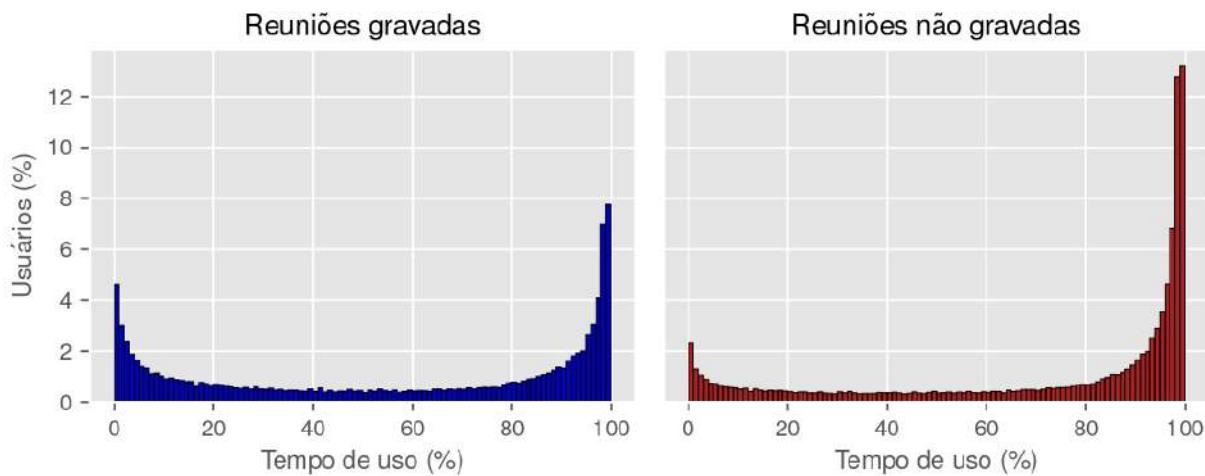
serviço a um amigo ou familiar. Se sua avaliação estiver entre 0 e 6, ele é classificado como um *detrator*. Caso seja 7 ou 8, ele é classificado como *neutro*. Caso seja 9 ou 10, ele é um *promotor*. O NPS é calculado subtraindo o percentual de *detratores* do percentual de *promotores*. Este trabalho foca em predizer a classe de um usuário (i.e. *detrator*, *neutro* ou *promotor*). De maneira análoga, pode-se computar a média das avaliações dos usuários de uma reunião, e definir sua classe usando a mesma abordagem. Logo, os modelos preditivos propostos podem ser aplicados tanto na classificação de usuários como na classificação de reuniões.

Novas variáveis foram criadas para armazenar a classe dos usuários e das reuniões segundo os conceitos do NPS. Como já identificado na etapa de análise exploratória de dados, a maioria dos usuários que avaliam sua experiência na ferramenta o fazem de maneira positiva. A nota 10 é a mais frequente, seguida da nota 9. Com isso, as bases de dados apresentam desbalanceamento de classes. No caso dos usuários, 88,2% são *promotores*, enquanto somente 6,2% são *neutros* e 5,6% são *detratores*. No caso das reuniões, os percentuais são de 89,8%, 6,0% e 4,2% para as classes *promotor*, *neutro* e *detrator*, respectivamente. Tal situação exige a aplicação de estratégias para tratar esse desbalanceamento. Para isso, foram testadas abordagens baseadas em *undersampling* e *oversampling* (Batista; Prati; Monard, 2004; Mohammed; Rawashdeh; Abdullah, 2020). No *undersampling*, amostras da classe majoritária são removidas até que a base de dados esteja balanceada, enquanto que no *oversampling* novas amostras das classes minoritárias são criadas até que a base de dados esteja balanceada.

Para o *undersampling*, foi adotado o método *Near Miss* (Mani; Zhang, 2003), o qual se baseia no algoritmo de *k-Nearest Neighbors* (kNN). Esse método seleciona exemplos a serem removidos da classe majoritária conforme sua distância com os elementos das demais classes. Na versão usada, são selecionados os elementos com menor distância média dos *k* elementos mais distantes das classes minoritárias, sendo *k* um parâmetro do método. Para o *oversampling*, foi usado o método SMOTE (Chawla *et al.*, 2002), do inglês *Synthetic Minority Oversampling Technique*. Esse método gera novos dados para as classes minoritárias via interpolação de dados similares.

Uma vez tratado o desbalanceamento da base de dados, os modelos de *random forest* foram construídos considerando uma divisão de 60% dos dados para treinamento, 20% para validação e 20% para teste. Ou seja, o conjunto de treinamento é usado para a construção dos modelos, enquanto o conjunto de validação é usado para o ajuste dos seus hiperparâmetros. Esse processo é chamado de configuração do modelo. O modelo configurado, i.e. com seus hiperparâmetros ajustados, é então avaliado no conjunto de teste. Os hiperparâmetros de modelos de *random forest* incluem, por exemplo, o número de árvores de decisão (estimadores) e sua profundidade máxima. A configuração desses modelos é uma etapa importante para se obter o máximo desempenho na tarefa preditiva (Feurer; Hutter, 2019). Para essa tarefa, foi usado o método de busca aleatória (Bergstra; Bengio, 2012) com 200 iterações e a acurácia como medida de qualidade a ser otimizada. Para a avaliação dos modelos configurados foram usadas métricas de acurácia, precisão, *recall* e medida F.

Além de serem facilmente interpretáveis, modelos de *random forest* permitem a aplicação de técnicas para

Figura 2: Tempo percentual de uso da câmera em reuniões gravadas e não gravadas.

avaliar a importância das variáveis do conjunto de dados (Breiman, 2001; Strobl *et al.*, 2008). Essas técnicas analisam o papel das variáveis nas diferentes árvores construídas, identificando aquelas que melhor dividem os dados e, por consequência, mais contribuem para a tarefa preditiva. A informação sobre a importância das variáveis é de grande relevância para a tomada de decisão, pois indica pontos prioritários para intervenções e melhorias no domínio de aplicação. Para concluir o estudo apresentado neste trabalho, foi conduzida uma análise das variáveis mais importantes para a classificação da experiência dos usuários do ConferênciaWeb.

4 RESULTADOS

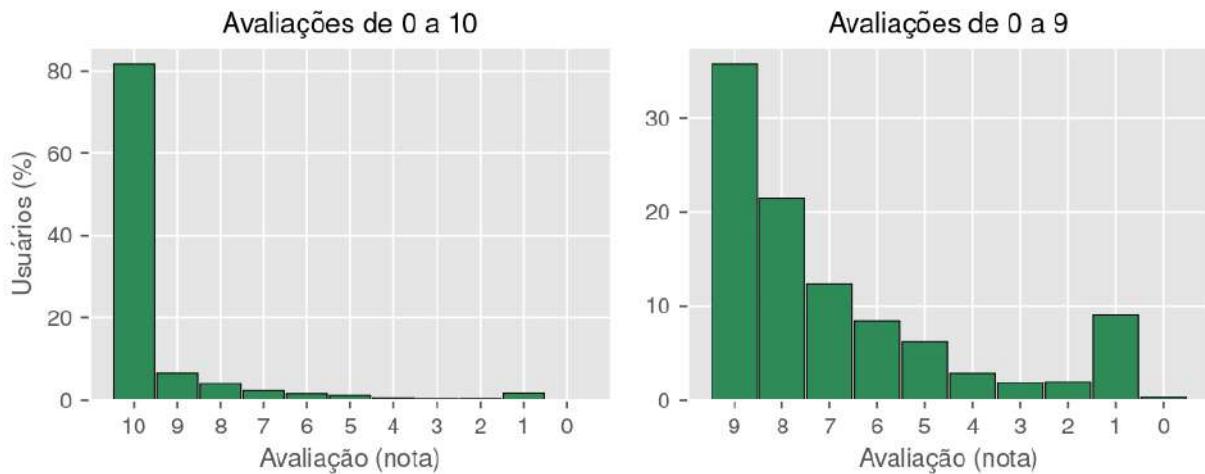
Esta seção apresenta os resultados desta pesquisa. A Seção 4.1 detalha os resultados da análise exploratória e visualização de dados. A Seção 4.2 detalha os resultados da construção e avaliação de modelos de classificação da experiência do usuário.

4.1 Análise exploratória e visualização de dados

A análise exploratória dos dados foca em dois aspectos: o comportamento dos usuários quanto ao uso dos recursos oferecidos pela ferramenta; e as avaliações feitas pelos usuários quanto à sua experiência de uso. Quanto ao uso dos recursos, foram analisadas as distribuições do tempo de uso. A Figura 1 mostra os resultados relacionados ao uso do microfone e da câmera, que são os recursos mais populares da ferramenta. São filtrados somente os usuários que interagiram com esses recursos, e apresentado o percentual de usuários conforme o tempo percentual de uso. Pode-se perceber que entre os usuários que ativaram seu microfone, a maioria deles permaneceu com o recurso ativado por mais de 80% do seu tempo de permanência na reunião. O comportamento é similar em relação ao uso da câmera. Neste caso, há um percentual expressivo de usuários que mantiveram a câmera ativada até 10% do seu tempo de permanência, seguido de uma quantidade constante de usuários com tempo de ativação entre 10% e 80% do tempo de permanência. Similar ao uso do microfone, a maioria dos usuários que ativaram a câmera mantiveram o recurso ativado por mais de 80% do seu tempo de permanência. Em resumo, a maioria dos usuários que ativam o microfone ou a câmera permanecem com o recurso ativado por praticamente todo o tempo de reunião. Esse comportamento é comum para usuários que conduzem a reunião, como moderadores, palestrantes e professores.

A Figura 2 explora mais a fundo o comportamento de uso da câmera, apresentando o percentual de usuários em função do tempo percentual de uso desse recurso para reuniões gravadas e não gravadas. Pode-se perceber que o tempo de ativação da câmera diminui de forma expressiva em reuniões gravadas. O percentual de usuários aumenta para menores tempos de uso (0% a 20%), ao passo que diminui para maiores tempos de uso (80% a 100%) em comparação com reuniões não gravadas. O percentual de usuários que permanece com a câmera ativada durante toda sua permanência na reunião cai de mais de 12% para cerca de 8%, por exemplo. Esses resultados sugerem que

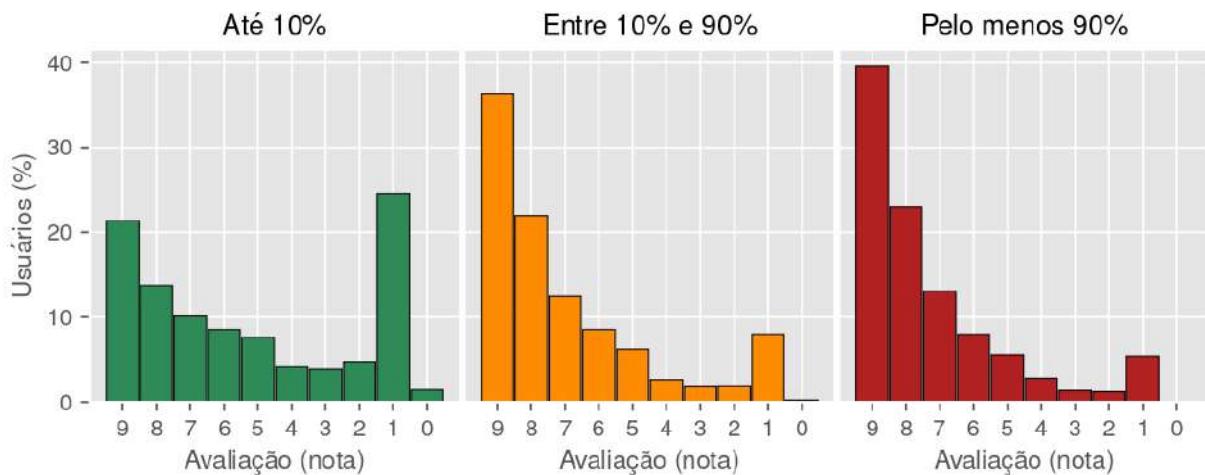
ANÁLISE EXPLORATÓRIA E PREDITIVA DE PADRÕES DE USO EM VIDEOCONFERÊNCIAS

Figura 3: Distribuição das notas atribuídas pelos usuários.


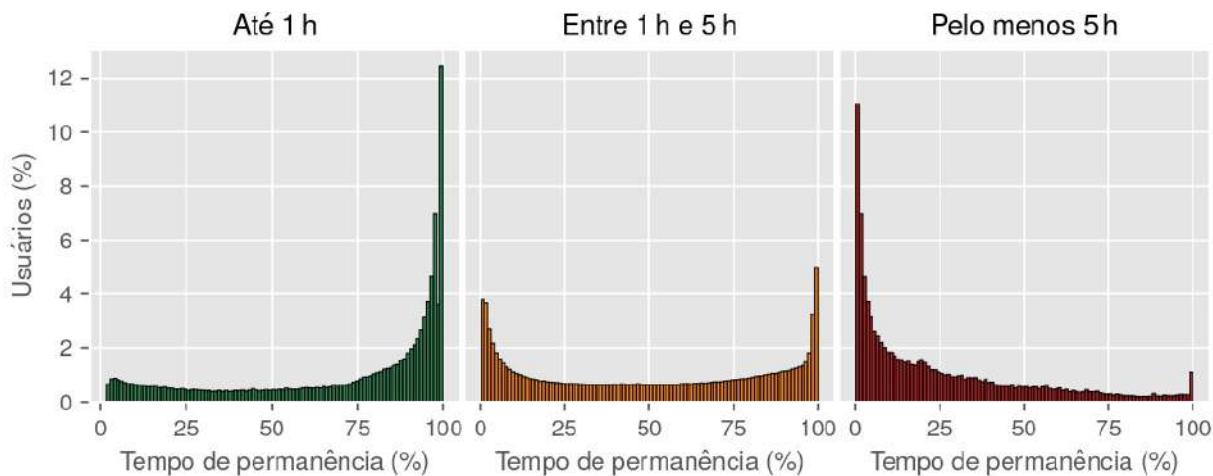
a gravação da reunião inibe muitos usuários a manterem sua câmera ativada, evitando o registro da sua imagem.

Com relação às avaliações dos usuários, foi analisada a distribuição das notas atribuídas para a experiência de uso. A Figura 3 mostra o percentual de usuários que atribuiu cada valor possível de nota, considerando somente usuários que avaliaram sua experiência. Pode-se notar que mais de 80% dos usuários atribuem nota máxima à sua experiência. Ao filtrar as avaliações de 0 a 9 (lado direito da Figura 3), percebe-se que o percentual de usuários decresce à medida que o valor de nota diminui, com exceção da nota 1. Neste caso, há um aumento expressivo, chegando a quase 10% dos usuários. Ou seja, usuários insatisfeitos com sua experiência de uso costumam atribuir essa nota.

Ao explorar esse comportamento em maior profundidade, percebe-se uma relação entre o tempo de permanência do usuário na reunião e a frequência da atribuição da nota 1 nas avaliações. A Figura 4 mostra a distribuição de notas de 0 a 9 entre os usuários com tempo de permanência de até 10% (lado esquerdo), entre 10% e 90% (centro) e pelo menos 90% (lado direito) do tempo de duração da reunião. Como pode-se notar, a frequência de atribuição da nota 1 é maior entre os usuários que deixaram a reunião no início (até 10%). Cerca de 25% desses usuários atribuíram nota 1, o que representa uma quantidade maior de usuários do que qualquer outra nota entre 0 e 9. Para

Figura 4: Avaliações dos usuários segundo o tempo de permanência na reunião.


ANÁLISE EXPLORATÓRIA E PREDITIVA DE PADRÕES DE USO EM VIDEOCONFERÊNCIAS

Figura 5: Distribuição dos tempos de permanência conforme a duração da reunião.


as outras faixas de tempo de permanência, esse percentual cai para menos de 10%. Essa insatisfação do usuário pode explicar o abandono precoce da reunião, o que motiva a investigação em relação à experiência de uso.

Também foi analisado o tempo de permanência dos usuários em relação à duração das reuniões. A Figura 5 mostra a distribuição do tempo percentual de permanência dos usuários para reuniões com duração de até 1 h (lado esquerdo), entre 1 h e 5 h (centro) e pelo menos 5 h (lado direito). Para reuniões curtas (de até 1 h) a maioria dos usuários permanece durante praticamente toda a reunião (75% ou mais do tempo de duração). Esse comportamento muda à medida que o tempo de reunião aumenta. Para reuniões longas (com 5 h ou mais), a maioria dos usuários permanece somente entre 0% e 25% do tempo de duração. A Figura 6 apresenta o tempo percentual de permanência dos usuários moderadores (lado esquerdo) e visualizadores (lado direito) em função da duração da reunião. O comportamento é similar entre os dois grupos. À medida que o tempo de duração aumenta, em especial a partir de 5 h, o tempo percentual de permanência dos usuários diminui gradativamente. Em reuniões muito longas (entre 10 h e 15 h, por exemplo) poucos usuários permanecem mais de 40% do tempo.

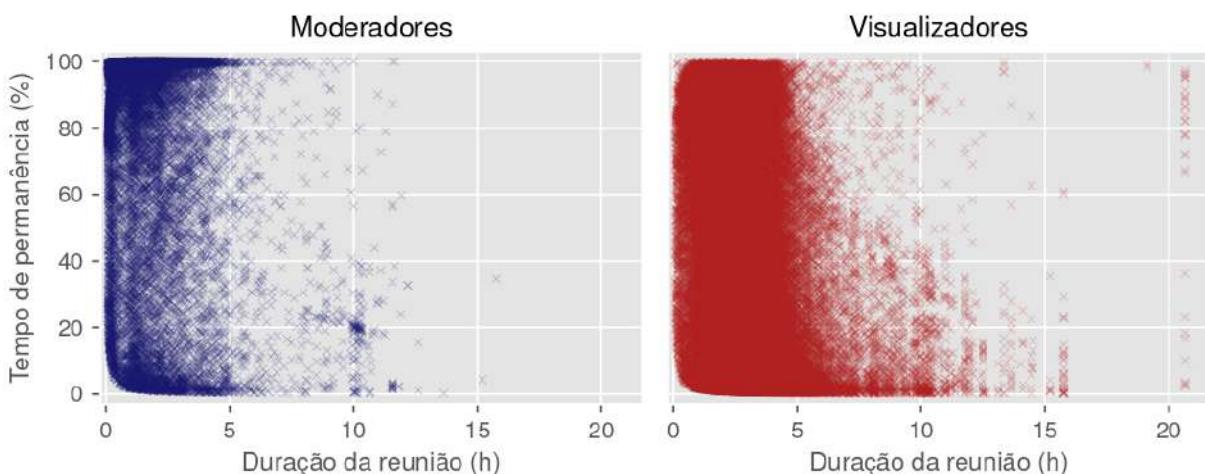
Figura 6: Tempo de permanência dos usuários conforme a duração da reunião.


Tabela 2: Hiperparâmetros dos modelos de *random forest* e melhores valores encontrados.

Cenário	Estimadores inteiro: [64, 256]	Profundidade inteiro: [16, 128]	Critério de divisão categórico: { <i>gini</i> , <i>entropy</i> , <i>log_loss</i> }
Usuários (RF + US)	186	17	<i>log_loss</i>
Usuários (RF + OS)	230	45	<i>gini</i>
Reuniões (RF + US)	120	16	<i>entropy</i>
Reuniões (RF + OS)	251	83	<i>log_loss</i>

Tabela 3: Desempenho dos modelos de *random forest* (valores percentuais).

Cenário	Modelo	Acurácia	Precisão	Recall	Medida F
Usuários	RF + US	72,3	(69,5; 59,2; 87,8)	(44,6; 72,3; 99,9)	(54,3; 65,1; 93,5)
	RF + OS	79,6	(79,9 ; 79,7 ; 79,0)	(84,8 ; 83,2 ; 70,6)	(82,3 ; 81,4 ; 74,6)
Reuniões	RF + US	77,2	(73,0; 66,3; 91,8)	(60,3; 72,3; 99,6)	(66,0; 69,1; 95,5)
	RF + OS	86,1	(84,9 ; 85,8 ; 87,4)	(83,4 ; 79,1 ; 95,8)	(84,2 ; 82,3 ; 91,4)

4.2 Predição da experiência dos usuários

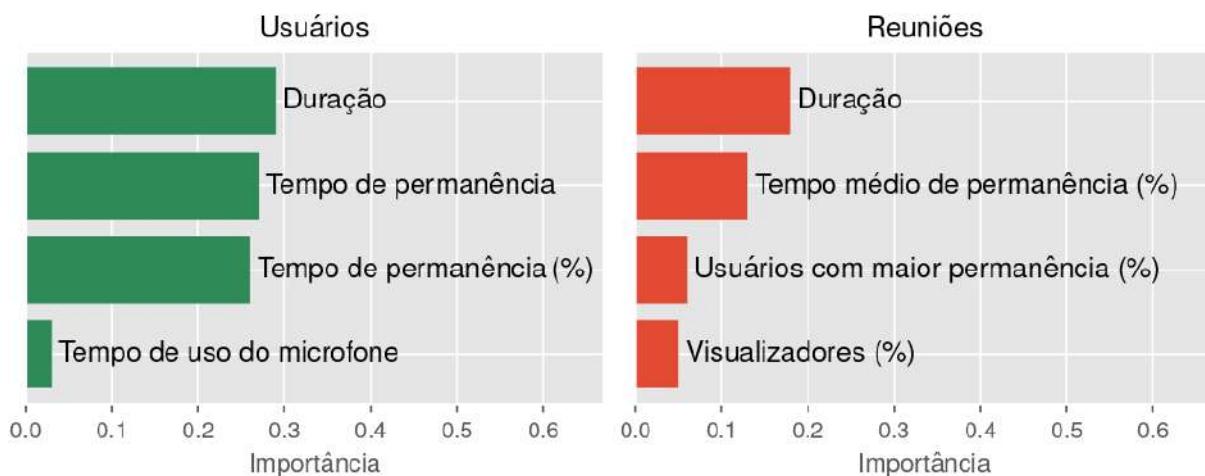
Para a predição da experiência de uso, são explorados modelos de *random forest* (RF) para a classificação tanto de usuários quanto de reuniões em *detrator*, *neutro* ou *promotor* (Reichheld, 2003). Para isso, os modelos são construídos para os dois cenários usando o conjunto de dados de treinamento. Os hiperparâmetros são configurados conforme o desempenho dos modelos no conjunto de validação. Finalmente, os modelos com os hiperparâmetros ajustados são avaliados no conjunto de teste. A Tabela 2 apresenta cada hiperparâmetro (número de estimadores, profundidade máxima e critério de divisão) com o respectivo tipo e intervalo ou conjunto de possíveis valores. São apresentados os melhores valores encontrados no processo de configuração considerando os dois cenários de classificação e as estratégias de *undersampling* (US) e *oversampling* (OS) para balanceamento de classes da base de dados. Como pode-se perceber, os melhores valores dos modelos que usam *undersampling* possuem um número menor de estimadores e uma menor profundidade máxima, em comparação com os modelos que usam *oversampling*. Em relação ao critério de divisão, as três abordagens foram selecionadas nos diferentes casos.

A Tabela 3 resume os resultados da aplicação dos modelos de *random forest* nos cenários estudados. São apresentados os valores de acurácia, precisão, *recall* e medida F para a classificação de usuários e reuniões usando *undersampling* e *oversampling*. Os valores dessas últimas três métricas são apresentados separadamente para as classes *detrator*, *neutro* e *promotor*, respectivamente. Os melhores resultados para cada cenário estão destacados em negrito. Observa-se que os modelos apresentam bom desempenho em todos os cenários, com valores de acurácia superiores a 70% em uma tarefa preditiva com três classes. Além disso, os valores de precisão e *recall* também são satisfatórios, especialmente para os modelos que usam *oversampling*. Os bons valores de medida F confirmam a relevância da acurácia obtida.

Pode-se perceber que os modelos que usam a estratégia de *oversampling* possuem melhor desempenho que os demais, i.e. maiores valores de acurácia. O mesmo se observa para as métricas de precisão, *recall* e medida F para as classes *detrator* e *neutro*. Já para a classe *promotor*, os modelos que usam *undersampling* apresentam melhor desempenho. *Promotor* é a classe majoritária da base de dados original, o que indica que a abordagem de *oversampling* favorece a predição das classes minoritárias ao gerar novas amostras dessas classes similares às existentes. Por outro lado, a abordagem de *undersampling*, que remove amostras da classe majoritária, beneficia a predição dessa classe.

Com o objetivo de obter informações mais aprofundadas sobre os dados e sobre a tarefa preditiva, foi medida a importância de cada variável dos modelos com melhor desempenho de ambos os cenários (i.e. os que usam *oversampling* – RF + OS). O resultado é apresentado pela Figura 7, que mostra as quatro variáveis mais importantes e sua contribuição para a classificação de usuários e reuniões. As variáveis relacionadas ao tempo, tanto de duração da reunião quanto de permanência dos usuários, são as mais importantes para a predição. No caso da classificação de usuários, a duração da reunião tem uma contribuição de quase 30% na tarefa preditiva, enquanto os tempos

Figura 7: Importância das variáveis para a tarefa preditiva.



de permanência absoluta e percentual do usuário na reunião contribuem com pouco mais de 25% cada. Na classificação de reuniões, a duração também é a variável mais importante, com contribuição de quase 20% na tarefa preditiva, seguida do tempo percentual médio de permanência e do percentual de usuários que permaneceu na reunião por pelo menos 80% da sua duração. As variáveis de tempo de uso do microfone e percentual de visualizadores também figuram entre as variáveis mais importantes na classificação de usuários e reuniões, respectivamente.

5 DISCUSSÃO

Além de comporem um estudo de caso interessante da aplicação de técnicas de ciência de dados e inteligência artificial, a pesquisa apresentada e os resultados obtidos apresentam uma contribuição relevante para o domínio explorado. As informações obtidas pela análise exploratória e visualização dos dados permitem entender o comportamento dos usuários do ConferênciaWeb. Além disso, as visualizações permitem acompanhar a atividade dos usuários e o uso dos recursos oferecidos pela ferramenta. Com isso, é possível analisar reuniões em função das suas características e da atividade dos usuários (e.g. definir os recursos mais importantes conforme o tipo de reunião). As técnicas exploradas neste trabalho permitem visualizar a distribuição das notas e monitorar as avaliações dos usuários, facilitando a extração de padrões. Por exemplo, é possível identificar o aumento de avaliações ruins e verificar sua relação com outras variáveis, o que permite investigar possíveis causas para experiências de uso negativas.

Os modelos de aprendizagem de máquina construídos neste trabalho se mostraram eficazes para a classificação de usuários e reuniões segundo os conceitos do NPS, apresentando excelente desempenho para essa tarefa. Esses resultados foram possíveis devido à metodologia adotada. Inicialmente, a base de dados foi preparada usando técnicas para o balanceamento das classes. Então, os modelos de *random forest* com melhor desempenho nos cenários estudados foram obtidos mediante a configuração dos seus hiperparâmetros. Como resultado, os modelos obtidos constituem uma importante ferramenta de análise de dados, pois permitem classificar a experiência de usuários e reuniões em tempo real. Além disso, a interpretabilidade dos modelos de *random forest* e a análise da importância das variáveis fornecem informações relevantes para o domínio de aplicação, indicando os fatores que mais impactam na qualidade da experiência de uso, o que permite concentrar esforços nas características mais importantes para a melhoria da ferramenta.

O estudo apresentado contribui para o fomento a decisões baseadas em dados. Por exemplo, é possível identificar recursos da ferramenta pouco explorados pelos usuários, e conduzir uma pesquisa para identificar possíveis razões (e.g. o recurso ter pouca relevância ou não funcionar corretamente). Em função do resultado, é possível aprimorar ou corrigir o recurso, ou ainda incentivar o uso através da própria ferramenta. Ao identificar potenciais usuários *detratores* durante o uso da ferramenta, é possível aplicar estratégias de intervenção a esses usuários, como a aplicação de instrumentos para obter maiores informações sobre a sua percepção da ferramenta e sugestões

ANÁLISE EXPLORATÓRIA E PREDITIVA DE PADRÕES DE USO EM VIDEOCONFERÊNCIAS

de melhorias. Tais informações são úteis para aprimorar a ferramenta, melhorando a experiência de uso e convertendo usuários *detratores* em *promotores*. Além disso, com a identificação de uma potencial reunião *detradora*, é possível fornecer essa informação ao responsável (e.g. moderador, professor ou palestrante), juntamente com dicas de melhoria.

Além das contribuições apresentadas, este trabalho demonstra o potencial da aplicação de técnicas de análise e visualização de dados, bem como de modelos de classificação para a exploração de dados de *log*, não somente de ferramentas de videoconferência, mas de soluções de tecnologia para outros domínios. A aplicação da ciência de dados na exploração de *logs* tem o potencial de contribuir diretamente para a melhoria das ferramentas associadas, uma vez que permite a extração de informações relevantes para a tomada de decisões. Neste sentido, este trabalho contribui para a ciência de dados ao fornecer diretrizes para a aplicação de diferentes técnicas na análise de dados de *log*. Além disso, as técnicas apresentadas podem ser aplicadas a dados oriundos de outros cenários. Por exemplo, uma prestadora de serviços que armazena dados sobre seus clientes (e.g. frequência de uso dos serviços, pagamentos realizados) e possui resultados de pesquisas de satisfação pode explorar os mesmos modelos de predição para classificar seus clientes e identificar potenciais *detratores*.

6 CONSIDERAÇÕES FINAIS

Este trabalho apresenta um estudo da ferramenta ConferênciaWeb a partir dos seus dados de *log*. São aplicadas técnicas de análise exploratória e visualização de dados, identificando o comportamento dos usuários e sua atividade em relação ao uso dos recursos oferecidos pela ferramenta. Além disso, são analisadas as avaliações fornecidas pelos usuários em relação à sua experiência de uso. São construídos modelos de *random forest* para a classificação da experiência de usuários e reuniões. Após o ajuste dos seus hiperparâmetros, os modelos foram avaliados em conjunto com técnicas de balanceamento de classes baseadas em *undersampling* e *oversampling*. Ao final, foram identificadas as variáveis mais importantes dos modelos de *random forest*.

A etapa de análise exploratória e visualização dos dados permitiu extrair informações relevantes. Por exemplo, foi possível identificar que a maior parte dos usuários que ativam seu microfone ou câmera permanece com o recurso ativado por praticamente toda a reunião. Esse comportamento muda quando a reunião está sendo gravada. Além disso, foi possível identificar um padrão nas avaliações negativas da ferramenta e sua relação com o tempo percentual de permanência dos usuários, bem como a redução desse tempo à medida que aumenta a duração da reunião. Os modelos de *random forest* se apresentaram eficazes na classificação de usuários e reuniões, obtendo acurárias médias entre 72% e 86% e bons valores de precisão, *recall* e medida F.

Esses resultados ressaltam as contribuições do trabalho no suporte à tomada de decisões. As informações extraídas a partir da base de dados permitem entender o comportamento dos usuários e fornecem um direcionamento para a implementação de melhorias na ferramenta. Além disso, os modelos de *random forest* permitem a classificação de usuários e reuniões em tempo real, permitindo a identificação de experiências ruins de forma precoce e a atuação para entender suas causas e mitigar potenciais efeitos negativos. Além das contribuições para o domínio explorado, este trabalho fornece diretrizes para aplicação das técnicas apresentadas, tanto na análise de *logs* de ferramentas computacionais como para outros cenários de exploração de dados.

Este trabalho pode ser estendido em diferentes direções. Pode ser realizado um estudo mais abrangente, explorando dados de outras ferramentas de videoconferência, bem como modelos adicionais de predição (e.g. kNN, regressão logística ou redes neurais). Nessa mesma direção, pode-se explorar funcionalidades de interação e outros recursos sofisticados de visualização de dados, produzindo *dashboards* para a análise e monitoramento em tempo real dos dados de *log* do ConferênciaWeb. Essas visualizações podem incluir as previsões dos modelos de *random forest*, mostrando usuários e reuniões de cada classe e permitindo a filtragem e detalhamento dos dados em função da classificação. Finalmente, os modelos de *random forest* podem ser aplicados para a predição de outras variáveis de interesse, como o abandono precoce dos usuários, contribuindo para a tomada de decisões em relação a outros aspectos da ferramenta.

REFERÊNCIAS

- ARCHIBALD, M. M.; AMBAGTSHEER, R. C.; CASEY, M. G.; LAWLESS, M. Using zoom videoconferencing for qualitative data collection: Perceptions and experiences of researchers and participants. **International Journal of Qualitative Methods**, v. 18, p. 1–8, 2019. Disponível em: <https://doi.org/10.1177/1609406919874596>.
- BARBOSA, W.; ZHOU, K.; WADDELL, E.; MYERS, T.; DORSEY, E. R. Improving access to care: Telemedicine across medical domains. **Annual Review of Public Health**, v. 42, n. 1, p. 463–481, 2021. Disponível em: <https://doi.org/10.1146/annurev-publhealth-090519-093711>.
- BATISTA, G. E. A. P. A.; PRATI, R. C.; MONARD, M. C. A study of the behavior of several methods for balancing machine learning training data. **SIGKDD Explor. Newsl.**, Association for Computing Machinery, New York, NY, USA, v. 6, n. 1, p. 20–29, 2004. ISSN 1931-0145. Disponível em: <<https://doi.org/10.1145/1007730.1007735>>.
- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. **Journal of Machine Learning Research**, v. 13, n. 2, 2012.
- BOND, M.; BEDENLIER, S.; MARÍN, V. I.; HÄNDEL, M. Emergency remote teaching in higher education: Mapping the first global online semester. **International Journal of Educational Technology in Higher Education**, SpringerOpen, v. 18, n. 1, p. 1–24, 2021.
- BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001.
- CAMILLERI, M. A.; CAMILLERI, A. C. Remote learning via video conferencing technologies: Implications for research and practice. **Technology in Society**, v. 68, p. 1–10, 2022. ISSN 0160-791X. Disponível em: <https://www.sciencedirect.com/science/>.
- CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. SMOTE: Synthetic minority over-sampling technique. **Journal of Artificial Intelligence Research**, v. 16, p. 321–357, 2002.
- CHEN, C.-h.; HÄRDLE, W. K.; UNWIN, A. **Handbook of Data Visualization**. 1. ed. Heidelberg: Springer Berlin, 2007.
- CORREIA, A. P.; LIU, C.; XU, F. Evaluating videoconferencing systems for the quality of the educational experience. **Distance Education**, Routledge, v. 41, n. 4, p. 429–452, 2020. Disponível em: 10.1080/01587919.2020.1821607.
- EZ-ZAOUIA, M.; TABARD, A.; LAVOUÉ, E. EMODASH: A dashboard supporting retrospective awareness of emotions in online learning. **International Journal of Human-Computer Studies**, v. 139, p. 1–15, 2020. ISSN 1071-5819. Disponível em: <https://www.sciencedirect.com/science/article/>.
- FALK, M. T.; HAGSTEN, E. When international academic conferences go virtual. **Scientometrics**, Springer, v. 126, n. 1, p. 707–724, 2021.
- FEURER, M.; HUTTER, F. Hyperparameter optimization. In: Frank Hutter, Lars Kotthoff, Joaquin Vanschoren. **Automated Machine Learning: Methods, Systems, Challenges**. Cham, Switzerland: Springer International Publishing, 2019. (The Springer Series on Challenges in Machine Learning), p. 3–33.
- GRAY, L. M.; WONG-WYLIE, G.; REMPEL, G. R.; COOK, K. Expanding qualitative research interviewing strategies: Zoom video communications. **The Qualitative Report**, Fort Lauderdale, v. 25, n. 5, p. 1292–1301, 2020.
- HALEEM, A.; JAVAID, M.; SINGH, R. P.; SUMAN, R. Telemedicine for healthcare: Capabilities, features, barriers, and applications. **Sensors International**, v. 2, p. 1–12, 2021. ISSN 2666-3511. Disponível em: <https://www.sciencedirect.com/S2666351121000383>.
- HASSAN, A.; MARI, Z.; GATTO, E. M.; CARDOZO, A.; YOUN, J.; OKUBADEJO, N.; BAJWA, J. A.; SHALASH, A.; FUJIOKA, S.; ALDAAJANI, Z.; CUBO, E.; GROUP the I. T. S. Global survey on telemedicine utilization for movement disorders during the COVID-19 pandemic. **Movement Disorders**, v. 35, n. 10, p. 1701–1711, 2020. Disponível em: <https://movementdisorders.onlinelibrary.wiley.com/>.
- HE, W. Examining students' online interaction in a live video streaming environment using data mining and text mining. **Computers in Human Behavior**, v. 29, n. 1, p. 90–102, 2013. ISSN 0747-5632. Including Special Section Youth, Internet, and Wellbeing.
- HURST, W.; WITTINGTON, A.; KOLIVAND, H. Virtual conference design: Features and obstacles. **Multimedia Tools and Applications**, Springer, v. 81, n. 12, p. 16901–16919, 2022.

ANÁLISE EXPLORATÓRIA E PREDITIVA DE PADRÕES DE USO EM VIDEOCONFERÊNCIAS

- IGUAL, L.; SEGUÍ, S. **Introduction to Data Science:** A python approach to concepts, techniques and applications. Cham: Springer, 2017.
- LIEUX, M.; SABOTTKE, C.; SCHACHNER, E. R.; PIRTLE, C.; DANRAD, R.; SPIELER, B. Online conferencing software in radiology: Recent trends and utility. **Clinical Imaging**, v. 76, p. 116–122, 2021. ISSN 0899-7071. Disponível em: <https://www.sciencedirect.com/science/>.
- MANI, I.; ZHANG, J. kNN approach to unbalanced data distributions: A case study involving information extraction. In: **ICML. Proceedings of the Workshop on Learning from Imbalanced Datasets**. Washington, D.C., 2003. v. 126, n. 1, p. 1–7.
- MOHAMMED, R.; RAWASHDEH, J.; ABDULLAH, M. Machine learning with oversampling and undersampling techniques: Overview study and experimental results. In: **11th International Conference on Information and Communication Systems (ICICS)**. New York: IEEE Press, 2020. p. 243–248.
- OLIVEIRA, G.; TEIXEIRA, J. G.; TORRES, A.; MORAIS, C. An exploratory study on the emergency remote education experience of higher education students and teachers during the COVID-19 pandemic. **British Journal of Educational Technology**, v. 52, n. 4, p. 1357–1376, 2021. Disponível em: <https://bera-journals.onlinelibrary.wiley.com/>.
- QUEZADA, R. L.; TALBOT, C.; QUEZADA-PARKER, K. B. From bricks and mortar to remote teaching: A teacher education program's response to COVID-19. **Journal of Education for Teaching**, Routledge, v. 46, n. 4, p. 472–483, 2020. Disponível em: <https://doi.org/10.1080/02607476.2020.1801330>.
- RABY, C. L.; MADDEN, J. R. Moving academic conferences online: Understanding patterns of delegate engagement. **Ecology and Evolution**, Wiley Online Library, v. 11, n. 8, p. 3607–3615, 2021.
- REICHHELD, F. F. The one number you need to grow. **Harvard Business Review**, v. 81, n. 12, p. 46–55, 2003.
- RUBINGER, L.; GAZENDAM, A.; EKHTIARI, S.; NUCCI, N.; PAYNE, A.; JOHAL, H.; KHANDUJA, V.; BHANDARI, M. Maximizing virtual meetings and conferences: A review of best practices. **International Orthopaedics**, Springer, v. 44, p. 1461–1466, 2020.
- SAMROSE, S.; MCDUFF, D.; SIM, R.; SUH, J.; ROWAN, K.; HERNANDEZ, J.; RINTEL, S.; MOYNIHAN, K.; CZERWINSKI, M. MeetingCoach: An intelligent dashboard for supporting effective & inclusive meetings. In: **Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems**. New York, NY, USA: Association for Computing Machinery, 2021. (CHI '21). Disponível em: <https://doi.org/10.1145/3411764.3445615>.
- SANDHU, R. K.; VASCONCELOS-GOMES, J.; THOMAS, M. A.; OLIVEIRA, T. Unfolding the popularity of video conferencing apps: A privacy calculus perspective. **International Journal of Information Management**, v. 68, p. 1–17, 2023. ISSN 0268-4012. Disponível em: <https://www.sciencedirect.com/science/article>.
- SARABIPOUR, S. Virtual conferences raise standards for accessibility and interactions. **eLife**, eLife Sciences Publications, Ltd, v. 9, p. 1–9, 2020.
- SILENCE, C.; RICE, S. M.; POLLOCK, S.; LUBOV, J. E.; OYESIKU, L. O.; GANESHRAM, S.; MENDEZ, A.; FEENEY, F.; KOUROSH, A. S. Life after lockdown: Zooming out on perceptions in the post-videoconferencing era. **International Journal of Women's Dermatology**, v. 7, n. 5, Part B, p. 774–779, 2021. ISSN 2352-6475. Disponível em: <https://www.sciencedirect.com/science/>.
- SOURDIN, T.; LI, B.; MCNAMARA, D. M. Court innovations and access to justice in times of crisis. **Health policy and technology**, Elsevier, v. 9, n. 4, p. 447–453, 2020.
- STROBL, C.; BOULESTEIX, A.-L.; KNEIB, T.; AUGUSTIN, T.; ZEILEIS, A. Conditional variable importance for random forests. **BMC Bioinformatics**, Springer, v. 9, p. 1–11, 2008.