DOI: 10.5748/20CONTECSI/PSE/AIT/7236

eLocator: e207236

MODELO DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE MENSAGENS DE USUÁRIOS DO SISTEMA DE SAÚDE QUANTO À PRESENÇA DE NECESSIDADES DE SAÚDE

Adriana Camargo De Brito – https://orcid.org/0000-0003-4019-6063
Instituto De Pesquisas Tecnológicas Do Estado De São Paulo

Gustavo Torres Custodio – https://orcid.org/0000-0002-3215-1693
Instituto De Pesquisas Tecnológicas Do Estado De São Paulo

Rogerio Silicani Ribeiro – https://orcid.org/0000-0002-7957-7405
Agile Healthtech

Rubens Carvalho Silveira – https://orcid.org/0000-0003-1862-1772
Agile Healthtech

Renata Luciria Monteiro – https://orcid.org/0000-0002-8436-6303
Agile Healthtech

Rodrigo Cabrera Castaldoni – https://orcid.org/0009-0009-1008-6681
Agile Healthtech

MACHINE LEARNING MODEL FOR CLASSIFICATION OF HEALTH SYSTEM USER MESSAGES REGARDING THE PRESENCE OF HEALTH NEEDS

ABSTRACT

The electronic transmission of messages represents a mode of communication employed across various services, including those in the healthcare sector. This study introduces a system for the active search of patients, wherein messages are received and categorized either for automatic response or for redirection to a human responder, depending on health-related necessities. The database employed in this study is one with two imbalanced classes. To address this issue, ChatGPT 3.5 was utilized to generate additional examples for the minority class in the database. Three scenarios were compared: the first using the original imbalanced database, the second with a database of synthetic messages generated by ChatGPT, and the third with ChatGPT-generated messages, albeit containing colloquial language and errors in Portuguese. In all models, the XGBoost algorithm was employed. The scenario in which XGBoost demonstrated the best discrimination capacity between messages was the one trained on the database of synthetic messages without Portuguese errors.

Keywords: Natural Language Processing; text classification; GPT

MODELO DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE MENSAGENS DE USUÁRIOS DO SISTEMA DE SAÚDE QUANTO À PRESENÇA DE NECESSIDADES DE SAÚDE

RESUMO

O envio de mensagens eletrônicas é uma forma de comunicação usada em diversos serviços, incluindo os da área da saúde. Neste trabalho, é proposto um sistema de busca ativa de pacientes, em que mensagens são recebidas e classificadas em mensagens que são respondidas automaticamente ou mensagens que devem ser direcionadas para serem respondidas por um humano, em função de necessidades de saúde. A base de dados utilizada neste trabalho é uma base com as duas classes desbalanceadas. Para lidar com esse problema, foi utilizado o ChatGPT3.5 para gerar mais exemplos da classe minoritária na base de dados. Foram comparados 3 cenários: o primeiro comdados na base original desbalanceada, o segundo com a base de dados de mensagens sintéticas geradas peloChatGPTe o terceirocom mensagens do ChatGPT, mas contendo linguagem coloquial e erros de português. Em todos os modelos foi utilizado o algoritmo XGBoost. O cenário onde o XGBoost apresentou melhor capacidade de discriminação entre as mensagens foi aquele treinado na base de dados commensagens sintéticas sem erros de português.

Palavras-chave:Processamento de Linguagem Natural; classificação de texto; GPT

1. INTRODUÇÃO

Atualmente 93% dos brasileiros comunicam-se por mensagens eletrônicas (CETIC, 2023), especialmente em aplicativos de smartphones. Esse padrão de comunicação tem se estendido para vários setores, incluindo os da área da saúde. Nesse contexto, as empresas prestadoras de serviços de saúde têm o desafio de gerenciar um volume crescente de mensagens de pacientes. Isso gera a necessidade de se ter grandes equipes para atender tais mensagens, o que nem sempre se tem disponibilidade. Diante disso, há possibilidade de criação de sistemas automatizados que possam ampliar a capacidade de interação entre pacientes e as equipes de saúde.

Modelos de aprendizado de máquina para classificar mensagens de texto podem ser criados com uso de métodos de Processamento de Linguagem Natural – PLN (Aubaid e Mishra, 2020; Hu et al. 2022; Lu et al. 2023; Robert M. Cronin et al., 2015; Robert M. Cronin et al.,2017; Sapozhnikova e Gordeeva 2019; Sulieman et al.,2017). Tais sistemas podem organizar e distribuir mensagens de pacientes, por exemplo, conforme a complexidade da demanda ou de acordo com a presença ou não de uma necessidade de saúde. Uma mensagem com indicação de necessidades de saúde poderia ser enviada para um atendente humano, enquanto uma mensagem sem tais necessidades poderia ser atendida por um chatbot com respostas pré-programadas. Isso permite a otimização das equipes de trabalhos das empresas sem causar prejuízos aos pacientes.

Nesse sentido, foi feita uma busca de artigos internacionais por palavras-chave como “machinelearning”, “natural languageprocessing”, “textmessages”, “patients”. Foram encontrados 50 artigos publicados desde 2003, porém, a maioria destes abordou contextos específicos relacionados à saúde mental, câncer, diabetes, doença cardiovascular, abuso de drogas ou doenças infecciosas. Poucos estudos avaliaram as mensagens enviadas para serviços de menor complexidade com objetivo de encontrar pessoas com necessidades médicas e prover informação, educação e direcionamento de serviços de forma personalizada. Neste trabalho são apresentados os estudos que abordam contextos gerais de classificação de mensagens, que foram considerados mais aderentes ao escopo deste artigo.

O objetivo do trabalho é efetuar uma classificação binária de mensagens de pacientes de uma empresa startup da área médica para determinar se a mensagem deve ser respondida por um atendente humano ou de forma automática, respectivamente, em função da presença ou ausência de necessidades de saúdemanifestadas no texto.

2. TRABALHOS RELACIONADOS

Huhet al. (2013) analisaram 8239 mensagens enviadas para um portal eletrônico de discussões referentes à doença diabetes, com objetivo de classificar a necessidade ou não de interação de um moderador especialista no assunto. Foram usados três tipos de características. Primeiro o texto foi vetorizado com a técnica Bag ofWords–BoW para Processamento de Linguagem Natural – PLN.O BoWcria um vetor considerando o número de vezes que uma palavra aparece no texto, sem levar em conta sequência ou posição da palavra no texto, sendo uma das formas mais simples de vetorizar textos.Os autores eliminaram do texto aspalavras não essenciais (stop words), com exceção de pronomes, que foram considerados relevantes para a pesquisa. Segundo, foram incluídas como características a frequência de ocorrência de palavras associadas a emoções positivas e negativas, derivadas de um estudo anterior (LIWC). Terceiro, foi registrado como característica o comprimento do tópico. O modelo foi treinado com o algoritmo de classificaçãoNaiveBayes, que assume a independência entre os preditores. Foi realizada uma seleção das características mais significativas utilizando o método dosqui-quadrado. O modelo foi treinado em uma base de dados com quantidade equivalente de classes e testado em dois tipos de bases de dados: com classes equilibradas e com classes desequilibradas. Os melhores resultados foram obtidos com a combinação das três seleções de

características(BoW, LIWC e comprimento do tópico) e base de teste com dados equilibrados. Comparado aos dados desequilibrados, o uso de dados equilibrados mostrou uma pontuação F1 mais alta (0,48 versus 0,54 F1-score). Com o modelo de classificação dos textos, seria viável reduzir a quantidade de mensagens que necessitariam da interação do moderador.

Cronin et al (2015) analisaram 1000 mensagens selecionadas randomicamente de uma base de 2,5 milhões de usuários de um portal de um hospital norte americano. As mensagens foram classificadas emcinco classes:informações clínicas, médicas, logísticas, sociais e outras. Os textos foram vetorizados com a técnica BoW e usados como características, em conjunto com os identificadores de conceito único (CUIs), que representam doenças e procedimentos, por exemplo, e tipos semânticos da linguagem médica (STYs), para facilitar a identificação de contextos. Foram utilizadosquatro classificadores:com um identificador de expressões regulares, com Regressão Logística, com NaiveBayes e com Random Forest. Foi utilizada técnica de validação cruzada com 5 folds. Os resultados mostraram que com a Regressão Logística e as Random Forest os modelos têm desempenhos similares e superaram o NaiveBayes na identificação de necessidades de saúde em mensagens de pacientes. A regressão logística foi mais eficiente em categorizar informações clínicas e médicas, enquantoRandom Forest se destacou em outras categorias. O uso de diferentes combinações de características, como BoW, CUIs e STYs, influenciou os resultados, com cada classificador apresentando melhores desempenhos em categorias específicas. Os ganhos em precisão ao adicionar CUIs e STYs foram modestos, indicando que a combinação de palavras é mais preditiva do que ferramentas de Processamento de Linguagem Natural (PLN) mais avançadas. Textos de pacientes tendem a ter menos conteúdo biomédico formal, o que pode limitar a eficácia de métodos de PLN de ordem superior. A pesquisa sugere que um sistema híbrido que combine diferentes classificadores pode ser mais eficaz para determinar as categorias de necessidades de saúde em mensagens individuais.

Suliemanet al. (2021) utilizaram 3000 mensagens randomicamente extraídas de uma base de 20 milhões de mensagens de um hospital norte americano. Os autores classificaram as mensagens em cinco categorias, assim como Cronin et al (2015). As mensagens foram convertidas em características utilizando os seguintes métodos de PLN: BoW, Bag ofPhrases (BoP) e representaçãobaseada em grafos. Além disso, foi usado o Word2Vec e Paragraph2Vec para representar vetores de palavras e parágrafos.O Word2Vec produz representações vetoriais de palavras, de modo que palavras com significados semelhantes tenham representações vetoriais similares. Ele captura relações semânticas e sintáticas complexas entre palavras, diferente do BoW, que não possui essa capacidade. O Paragraph2Vec é uma extensão do Word2Vec desenvolvida para capturar a representação vetorial de parágrafos ou documentos inteiros. Para classificar as mensagens de acordo com diferentes necessidades de comunicação, foram desenvolvidos quatro classificadores binários. Foram utilizadosRegressão Logística eRandom Forest, com BoW, BoP e vetores de grafos. Além disso, foi empregada uma Rede Neural Profunda (DNN), onde os vetores de todas as palavras em uma mensagem foram somados e alimentados na rede, com uma camada softmax para classificar as mensagens em rótulos positivos ou negativos. Os parâmetros da DNN foram ajustados para otimização. Também foi implementado um modelo de Rede Neural Convolucional (CNN) usando a incorporação Word2Vec, onde cada mensagem foi convertida em uma matriz de vetores de palavras. A CNN consistiu em várias camadas,com ajustes finos nos parâmetros do modelo para se adequar à análise. Foram testados vários parâmetros do modelo, sendo utilizado o Word2Vec treinado tanto em documentos clínicos quanto no Google News. Os resultados indicaram que abordagens que consideram a semântica das mensagens é mais eficaz, principalmente com CNN, obtendo área abaixo da curva ROC acima de 0,9. Embora as técnicas convencionais como BoW sejam aceitáveis, vetores de grafos e vetores de parágrafo mostraram benefícios adicionais. Os modelos com CNN foram particularmente eficazes na classificação de comunicações sociais e informativas, mas enfrentaram desafios com mensagens de conteúdo misto.

Para classificar mensagens de pacientes, nos trabalhos analisados, os autores efetuaram tanto classificações binárias quanto multinomiais, utilizaram, desde técnicas simples de Processamento de Linguagem Natural para vetorizar palavras como BoW, até as mais sofisticadas com Word2Vec ou Paragraph2Vec. Para criar os modelos de classificação, foram utilizados algoritmos como NaiveBayes, Regressão Logística, Random Forest, CNNs ou DNNs. Observou-se que é possível obter resultados satisfatórios com modelos simples, obtendo-se acréscimos no desempenho com uso de técnicas ou ferramentas mais complexas, o que varia em função das características do banco de dados, dentre outros aspectos.

3. METODOLOGIA

Contexto

Uma empresa startup utiliza uma plataforma própria de mensagens integrada ao WhatsApp para buscar ativamente pessoas que tem dúvidas de saúde ou necessitam algum cuidado para a prevenção e o tratamento de doenças e para indicar soluções personalizadas. A startup fornece orientações, educação em saúde e direcionamento de profissionaispara beneficiários de planos de saúde.

Metade das interações que ocorrem são com pessoas que recebem o contato de um atendente da empresa, trocam mensagens de texto com a equipe enão apresentam dúvidas ou necessidades de cuidado preventivo ou terapêutico. A outra metade é atendida por pessoas com perfil especializado na área de saúde.

Mensalmente a empresa recebe da ordem de mil mensagens. Como a empresa tem um grupo reduzido de profissionais para lidar com essas mensagens, foi construído um modelo de aprendizado de máquina paraclassificá-lasem categorias de tal modo que parte delas é respondida de modo automatizado. Isso representa um ganho de produtividade significativo nos serviços da empresa, aumentando sua capacidade operacional e reduzindo custos.

As mensagens que são atendidas por atendentes humanos são intituladas “manuais - com necessidades”, são aquelas que contém assuntos relacionados a dores, incômodos, dúvidas em tratamentos, agendamento de exames e consultas, que requerem a presença de um profissional, em virtude de sua complexidade. As mensagens que são atendidas por mecanismos automatizados são denominadas “automáticas - sem necessidades”, são aquelas que contém solicitações administrativas, opiniões ou agradecimentos e todas as que não se inserem na primeira categoria.

As mensagens utilizadas no trabalho foram anonimizadas, usadas somente para treinamento, validação e teste dos modelos, com conteúdo e dados sensíveis de pacientes sendo mantidosem sigilo de acordo com princípios da Lei Geral de Proteção de Dados (Brasil, 2018). Vale ressaltar que o mecanismo desenvolvido neste trabalho se encontra em produção atualmente e aqui são apresentados os estudos elaborados anteriormente à sua implantação.

Base de dados

Foram utilizadas mensagens da plataforma da empresa, selecionando-se as primeiras respostas dos pacientes, logo após o envio da mensagem de busca ativa (mensagens disparadas em lote para iniciar o contato com o paciente e identificar se há necessidade de saúde). A base de dados da plataforma apresenta 1.279 mensagens de texto iniciais, trocadas entre os clientes e a empresa, no período de 2020 a 2023 nas cidades de São Paulo, Rio de Janeiro, Belo Horizonte e Vitória.

Foram selecionadas todas as mensagens enviadas pelos usuários que interagiram com a mensagem de busca ativa até a primeira resposta de atendimento do promotor de saúde. Estas respostas foram rotuladas como: “manuais - com necessidade médica” e “automáticas - sem necessidade médica”, por um profissional de saúde que conhece a operação e o modelo de atendimento e possui conhecimento técnico adequado.

A base de dados possui maior quantidade de mensagens “manuais - com necessidade” (%) e somente 11% de mensagens da classe “automáticas - sem necessidade”. Esse tipo de situação pode levar ao modelo a prever somente uma classe, prejudicando o seu desempenho (Kotsiantis et al., 2006).Como há um desbalanceamento significativo entre a quantidade de mensagens com os dois rótulos, foram criadas mensagens sintéticas utilizando o ChatGPT 3.5 para aumentar a quantidade de mensagens da classe minoritária (data augmentation) e reduzir o desequilíbrio entre as classes, a partir de exemplos de mensagens contidas na base de dados original.

Para testar a eficiência dessa estratégia de data augmentation, foram criados três modelos, com e sem tal artifício, como descrito a seguir:

- **Modelo 1:** base de dados com as mensagens originaisda plataforma da empresa com 1.279 mensagens (896 para treino e 383 para teste), sendo 11% dessas mensagens classificadas como “automáticas - sem necessidade” e as demais (89%) “manuais - com necessidade”;
- **Modelo 2:**base de dados 2.379 mensagens, sendo 50% as mensagens originais da plataforma da empresa e as demais geradas peloChatGPTconsiderando português sem erros. Foram utilizadas 1.666 para treino e 713 para teste, sendo 29,2% das mensagens classificadas como “automáticas sem necessidade” e 70,7% “manuais - com necessidade”;
- **Modelo 3:** base com dados com 2.509 mensagens, das quais 50% correspondem a mensagens originais da plataforma da empresa, 25% são geradas pelo ChatGPT sem erros de português e 25% geradas pelo ChatGPT com erros simulando a linguagem coloquial (ortográficos e gramaticais comuns). Foram utilizadas 1.796 mensagens para treino e 713 para teste, sendo 30,8% dessas mensagens classificadas como “automáticas - sem necessidade” e as demais 69,2% “manuais - com necessidade”.

Pré-processamento

Foi feito um pré-processamento dos dados com limpeza das mensagens, removendo emojis, mídias, acentos, números, caracteres especiais e efetuando-se a transformação de letras maiúsculas em minúsculas. Em seguida foi realizada a segmentação de palavras (tokenização), os tokens gerados passaram pelo processo de lematização assim diminuindo a variabilidade de tokens que podem trazer o mesmo significado (ex: conjugação verbal). O peso de cada token foi calculado pela técnica da frequência do termo – inverso da frequência nos documentos (tf-idf). Essa é uma medida estatística comumente usada para avaliar a importância de uma palavra em um documento dentro de um conjunto ou corpus de documentos.

Detalhamento dos modelos

Os modelos de classificação binária de mensagens foram desenvolvidos na linguagem em R (software R versão 4.2.1), e seus pacotes (*tidyverse, tidytext, readxl, abjutils, Information, cutpointr, tidymodels, funModeling, vip, rsample, forcats, pROC, tm*).

O conjunto de dados tratados e rotulados foram divididos em 70% para treino e 30% para validação. Foi feito teste em uma base de dados inédita para todos os modelos com 500 mensagens, oriundas da base de dados da plataforma da empresa.

Foi utilizado o XGboost (Chen et al., 2015), quese enquadra na categoria de ensemble, um método que utiliza modelos de árvores que minimizem a função de perda. Essa ferramenta também é adequada para o contexto do trabalho e teve um desempenho bomem trabalhos com bases de dados com quantidades desbalanceadas de amostras entre classes (Wang et al. 2020; ZhangandChen 2021). Para otimização do modelo foi realizado ajuste de hiperparâmetros com técnica de validação cruzada com 7 folds. Os hiperparâmetros utilizados nos modelos são indicados na Tabela 1.

Tabela 1 – Hiperparâmetros utilizados nos modelos

Descrição do hiperparâmetro	**Nome**	**Modelo 1**	**Modelo 2**	**Modelo 3**
Quantidade de variáveis/ palavras sorteadas por árvore.	mtry	690	913	939
Quantidade mínima de observações dentro de um nó para se considerar dividir em duas folhas novas.	min_n	3	3	3
Quantidade máxima de nós que cada arvore terá (profundidade máxima)	tree_depth	7	7	7
Taxa de aprendizado	learn_rate	0,02	0,02	0,02
Parâmetro regularizador	loss_reduction	$1{,}1\times 10^{-10}$	$1{,}1 \times 10^{-10}$	$1{,}1 \times 10^{-10}$
Proporção de amostras para sortear por árvore	sample_size	0,994	0,994	0,994
Nº de arvores	tree	1000	1000	1000

Métricas de avaliação

Para avaliar o desempenho dos modelos e comparar sua eficiênciaforam utilizadascomo métricas a acurácia, a sensibilidade (recall) e a especificidade para os seguintes pontos de corte:0,39 para os modelos 1 e 3 e0,45para o modelo 2. Também foi considerada a área sob a curva ROC (ROC-AUC). Um ponto de corte igual a 0,45, por exemplo, indica que resultados acima deste valor serão classificados como pertencentes à classe “manuais - com necessidades”, enquanto valores menores que esse, serão classificados como pertencentes à classe “automáticas - sem necessidades”.

A acurácia é calculada como a proporção de previsões corretas, positivas ou negativas, em relação à quantidade total de previsões. A Sensibilidade ou Recall indica a Taxa de Verdadeiros Positivos, ou seja, é a proporção de casos positivos reais que foram corretamente identificados pelo modelo, neste trabalho a classe “automáticas - sem necessidades”.

A Especificidade é a proporção de casos negativos reais que foram corretamente identificados como negativos pelo modelo, neste caso relacionada à classe “manuais - com necessidades”.Área sob a Curva ROC (AUC - ROC): AUC - ROC é uma medida de desempenho para problemas de classificação em vários limiares de ponto de corte. Ela indica quão bem o modelo é capaz de distinguir entre classes. Um valor AUC de 1 indica um modelo perfeito, enquanto um valor de 0.5 indica um modelo que não tem capacidade de classificação melhor do que um modelo aleatório.

Destaca-se ainda que o modelo com melhor desempenho foi selecionado para a realização dodeployem um serviço de computação em nuvem da AmazonWeb Service (AWS) e atualmente encontra-se em funcionamento.

4. RESULTADOS

Nas Figuras 1 a 3 são indicados os valores das métricas obtidas, respectivamente para os Modelos 1 a 3.

Figura 1– Desempenho do Modelo 1.

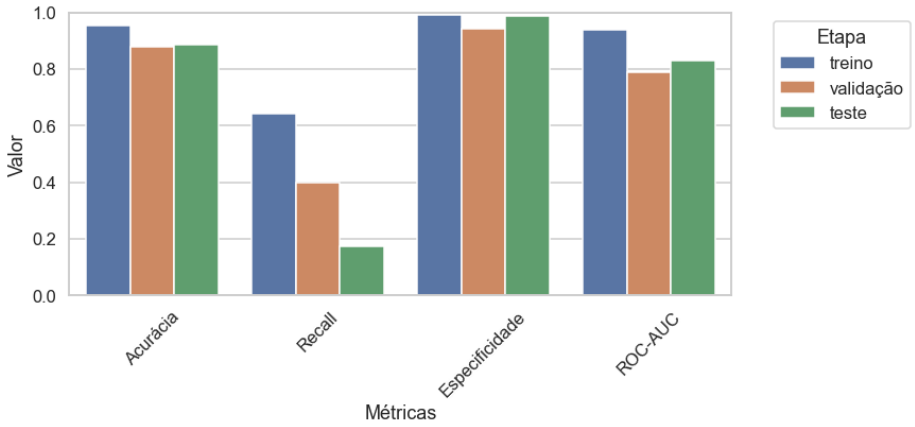


Figura 2 – Desempenho do Modelo 2.

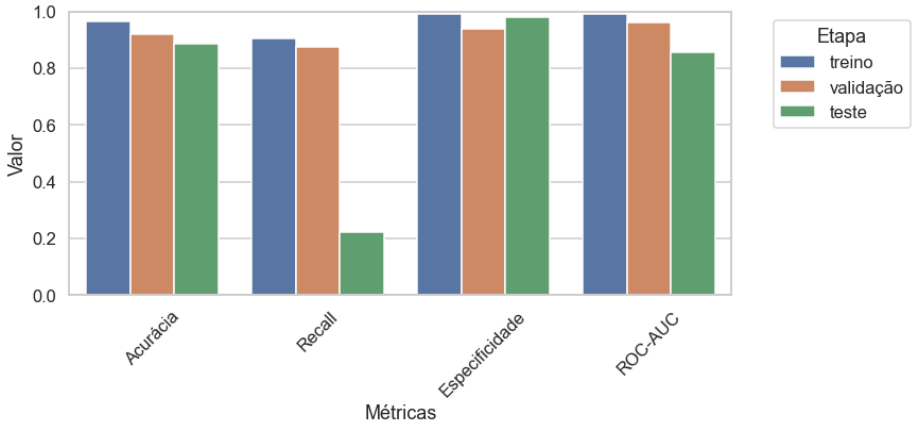
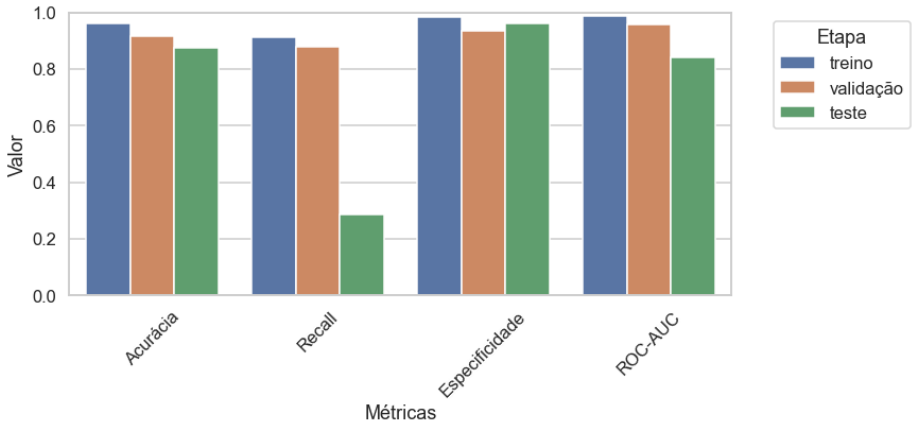


Figura 3 – Desempenho do Modelo 3.



Considerando a acurácia os três modelos são similares, com acurácia de treino, teste e validação acima de 0,85 (Figuras 1 a 3). Ressalta-se que o Modelo 1 tem certa tendência a sobreajuste, ou overfitting, visto que o valor da acurácia de treino é ligeiramente superior que o valor da acurácia de validação, o que indicaria a necessidade de efetuar ajustes no modelo.

Entretanto, quando se observam outras métricas, o modelo 1, treinado na base de dados original, que possui maior desbalanceamento de mensagens entre classes obteve pior desempenho, enquanto os demais, que tiveram o data augmentation com dados do ChatGPT, têm comportamento similar e superior ao obtido pelo modelo1.

O modelo 1 (Figura 1) possui recallda ordem de 0,6 na fase de treino, 0,4 na fase de validação e 0,17 na fase de teste. Isso indica sua baixa capacidade em classificar as mensagens “automáticas - sem necessidade”, (classe positiva) o que era de se esperar, em decorrência da pequena quantidade e, talvez qualidade das mensagens disponíveis desta classe.Há um significativo desbalanceamento entre classes, sendo11% de mensagens da

classe minoritária, o que explica bem tais resultados. Os outros dois modelos (Figuras 2 e 3) possuem recall de treino e validação acima de 0,85, o que representa um bom desempenho. Entretanto, no teste com dados inéditos para todos os modelos, ambos também apresentaram baixa especificidade, de 0,2 a 0,3, embora com melhores condições que o Modelo 1. Isso requer certa atenção ao desempenho esperado para esses modelos, além de uma análise específica das mensagens contidas no banco de dados do teste e suas características, para identificar os motivos que levaram tais modelos a terem um desempenho tão diferente no teste em comparação com o desempenho obtido na etapa de validação.

Quanto à especificidade, os três modelos têm essa métrica acima de 0,9 nas três fases, indicando boa capacidade em classificar as mensagenscom necessidades (classe negativa), em decorrência da qualidade e quantidade adequada de mensagens com este rótulo (Figuras 1 a 3).

A área abaixo da curva ROC para os três modelos têm comportamento esperado, sendo mais baixa para o Modelo 1 em comparação com os demais nas três fases. Servindo como um bom indicador da qualidade dos modelos.

Em linhas gerais, os Modelos 1 e 2 possuem comportamentos semelhantes, tendo contribuição significativa do data augmentation com uso do ChatGPT para a melhoria geral do seu desempenho em comparação com o desempenho do Modelo 1. Em números absolutos o Modelo 2 obteve valores de algumas métricas superiores aos do Modelo 3, sendo selecionado pela empresa para ser colocado em prática.

5. CONCLUSÕES E CONSIDERAÇÕES FINAIS

Foi feita uma classificação binária de mensagens de pacientes de uma operadora de saúde quanto à presença de necessidades de saúde. Como a base de dados de mensagens estava com quantidades de amostras em classes desbalanceadas, foi realizado um processo de data augmentation com o uso do ChatGPT para reduzir tal discrepância.

O modelo com melhor desempenho foi aquele que se utiliza de dados sintéticos gerados pelo ChatGPT para reduzir o desequilíbrio entre quantidades de amostras de classes desbalanceadas. Mesmo ainda havendo espaço para aprimoramentos do modelo, observou-se que o data augmentation feito com esta ferramenta teve contribuições significativas para a melhoria do seu desempenho, mostrando ser uma ferramenta eficaz para a construção de dados sintéticos de mensagens.

A implantação do modelo ampliou o atendimento de pessoas com necessidades de saúde.

6. REFERÊNCIAS BIBLIOGRÁFICAS

Aubaid, A. M., & Mishra, A. (2020). A Rule-Based Approach to Embedding Techniques for Text Document Classification. Applied Sciences.

Brasil. (2018). Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). DiárioOficial da União.

Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br). (2023). Pesquisasobre o uso das tecnologias de informação e comunicaçãonosdomicíliosbrasileiros: TIC Domicílios 2022. https://cetic.br/media/docs/publicacoes/2/20230825143720/tic_domicilios_2022_livro_eletronico.pdf

Cronin, R. M., Fabbri, D., Denny, J. C., & Jackson, G. P. (2015). Automated Classification of Consumer Health Information Needs in Patient Portal Messages. AMIA Annual Symposium Proceedings, 2015, 1861-1870. PMID: 26958285; PMCID: PMC4765690.

Lu, G., Liu, Y., Wang, J., & Wu, H. (2023). CNN-BiLSTM-Attention: A multi-label neural classifier for short texts with a small set of labels. Information Processing & Management, 60(3), 103320. https://doi.org/10.1016/j.ipm.2023.103320.

Sulieman, L., Gilmore, D., French, C., Cronin, R. M., Jackson, G. P., Russell, M., &Fabbri, D. (2017). Classifying patient portal messages using Convolutional Neural Networks. Journal of Biomedical Informatics, 74, 59-70. https://doi.org/10.1016/j.jbi.2017.08.014

Huh, J., Yetisgen-Yildiz, M., & Pratt, W. (2013). Text classification for assisting moderators in online health communities. Journal of Biomedical Informatics, 46(6), 998-1005. https://doi.org/10.1016/j.jbi.2013.08.011

Sapozhnikova, L. E., & Gordeeva, O. A. (2019). Text classification using convolutional neural network.

Wang, C., Deng, C., & Wang, S. (2020). Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. Pattern Recognition Letters, 136, 190-197.

Zhang, Y., & Chen, L. (2021). A Study on Forecasting the Default Risk of Bond Based on XGBoost Algorithm and Over-Sampling Method. Theoretical Economics Letters, 11(02), 258-267.

Kotsiantis, S., Kanellopoulos, D., &Pintelas, P. (2006). Handling imbalanced datasets: A review. GESTS international transactions on computer science and engineering, 30(1), 25-36.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., ... & Zhou, T. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4.