

Predição de Sepsis a partir de Dados do Atendimento Pré-Hospitalar

Naira Kaieski¹, Priscila Schmidt Lora², Cristiano André da Costa¹

¹ Software Innovation Laboratory (SOFTWARELAB)
Programa de Pós-graduação em Computação Aplicada (PPGCA)
Universidade do Vale do Rio dos Sinos (UNISINOS)
São Leopoldo -- RS — Brasil

²Programa de Pós-graduação em Enfermagem
Universidade do Vale do Rio dos Sinos (UNISINOS)
Porto Alegre -- RS — Brasil

naira.kaieski@gmail.com, {plora, cac}@unisinos.br

Abstract. *Sepsis is a syndrome that affects millions of people each year, causing approximately 20% of deaths worldwide. Early recognition of sepsis symptoms makes possible to start the appropriate treatment in time to provide better outcomes. Thus, it is relevant to develop tools that enable early identification of sepsis in the prehospital environment, such as the qSofa heuristic score. This article presents the results of tests performed with machine learning models trained with data from the first attendance provided by SAMU. Even with a restricted data set, the models developed showed improvements in the order of 7.7% in accuracy and 17.7% in AUC in relation to the results of qSofa.*

Resumo. *A sepsis é uma síndrome que acomete milhões de pessoas a cada ano causando aproximadamente 20% das mortes no mundo. O reconhecimento precoce dos sintomas de sepsis possibilita o início do tratamento adequado a tempo de proporcionar melhores desfechos para os pacientes. Dessa forma, é relevante desenvolver ferramentas que possibilitem a identificação da sepsis já no ambiente pré-hospitalar como o escore heurístico qSofa. Este artigo apresenta os resultados de testes realizados com modelos de machine learning treinados com dados do primeiro atendimento prestado pelo SAMU. Mesmo com um conjunto de dados restrito, os modelos desenvolvidos apresentaram melhorias na ordem de 7,7% na acurácia e 17,7% na AUC em relação aos resultados do qSofa.*

1. Introdução

A sepsis é uma síndrome associada a anormalidades fisiológicas, patológicas e biológicas causadas por uma resposta desregulada do hospedeiro a uma infecção causando um sério risco de sequelas graves ou mesmo óbito [Singer et al. 2016]. Milhares de pessoas no mundo desenvolvem sepsis todos os anos ocasionando perdas humanitárias e econômicas significativas [Buchman et al. 2020]. Estimativas mundiais acerca da incidência de sepsis apontam que em 2017 foram registrados cerca de 48,9 milhões de casos e 11,0 milhões de mortes relacionadas à sepsis, representando 19,7% de todas as mortes globais [Rudd et al. 2020].

A sepse é reconhecidamente uma séria ameaça à saúde em todo o mundo apresentando uma alta taxa de incidência e mortalidade em países de média e baixa renda como o Brasil e em regiões onde os recursos hospitalares são limitados [World Health Organization 2020]. Contudo, este tema se torna ainda mais relevante à luz de recentes pesquisas [Elezkurtaj et al. 2021, Zafer et al. 2021, Li et al. 2020] que investigam a ocorrência de sintomas análogos à sepse viral em pacientes com manifestações mais graves de COVID-19 (doença causada pelo vírus *Severe Acute Respiratory Syndrome Coronavirus 2* [SARS-CoV-2]).

A identificação precoce e o tratamento adequado podem mitigar as graves consequências da sepse [Reyna et al. 2020]. Neste contexto, o presente artigo apresenta uma abordagem baseada em técnicas de aprendizado de máquina (*machine learning* - ML) para prever o desenvolvimento de sepse baseado apenas em dados coletados no primeiro atendimento prestado ao paciente já no ambiente pré-hospitalar. A evolução do poder de processamento, a grande disponibilidade de dados de saúde e a capacidade de “aprendizado” dos algoritmos de inteligência artificial possibilitam que estes sejam explorados amplamente em áreas críticas como a saúde [Pucchio et al. 2021].

No Brasil, o atendimento pré-hospitalar pode ser realizado pelo Serviço de Atendimento Móvel de Urgência (SAMU) disponibilizado pelo Sistema Único de Saúde (SUS) [Brasil 2003]. O serviço do SAMU pode ser solicitado gratuitamente por qualquer pessoa e é composto por uma unidade reguladora (Central Médica de Regulação - CMR) e um corpo assistencial (equipe das ambulâncias). Neste trabalho estamos analisando essencialmente os dados coletados pelas equipes de atendimento do SAMU que se deslocam com uma ambulância e uma equipe de profissionais de saúde até o local onde o paciente se encontra. Esta equipe é responsável por estabilizar o paciente, coletar informações preliminares como sinais vitais e dados demográficos e então conforme a indicação do regulador procede com a remoção do paciente para a unidade de saúde adequada [O’Dwyer et al. 2017].

Uma vez que o diagnóstico precoce de sepse pode proporcionar desfechos mais favoráveis aos pacientes, justificam-se os esforços da comunidade científica para desenvolver ferramentas que auxiliem na identificação da sepse já neste atendimento pré-hospitalar realizado pelo SAMU. Assim, o objetivo deste trabalho foi analisar o potencial de predição precoce das variáveis coletadas durante este atendimento pré-hospitalar que consiste em um ambiente de baixa complexidade e com poucos recursos. Os algoritmos de aprendizado de máquina podem ser úteis nesta tarefa de predição fornecendo indícios que permitam um direcionamento mais assertivo dos pacientes para o atendimento nas unidades de saúde. Neste trabalho foram analisados diferentes algoritmos de ML para identificar qual possui a melhor capacidade de identificar sepse precocemente apenas com dados coletados no atendimento pré-hospitalar.

Os resultados encontrados foram comparados com o escore qSOFA (*Quick Sequential [Sepsis-related] Organ Failure Assessment*) [Singer et al. 2016]. Este escore é bastante simples, sendo calculado apenas com informações do nível de consciência do paciente, frequência respiratória e pressão arterial sistólica. Consequentemente, este escore pode ser utilizado à beira do leito e também no ambiente pré-hospitalar. Os resultados encontrados neste trabalho apontam uma melhoria significativa nas métricas de avaliação comparando os modelos de ML e o escore qSOFA. Foi encontrado um ganho de 7,7% na

acurácia utilizando um modelo treinado com *Gradient Boosting* e de 17,7% na AUC (*Area Under the Receiver Operating Characteristic Curve*) utilizando uma implementação com Regressão logística. Os resultados são detalhados na seção 3.

O escore qSofa tem sido avaliado em outros trabalhos como uma ferramenta para identificação precoce dos sinais de sepse no ambiente pré-hospitalar [Lane et al. 2020, Jouffroy et al. 2019, Koyama et al. 2019]. Da mesma forma, modelos de ML já foram desenvolvidos com o objetivo de proporcionar um diagnóstico precoce de sepse [Goh et al. 2021, Shiraishi and Marujo 2020, Barton et al. 2019, Mao et al. 2018]. Contudo, é necessária atenção ao se comparar os resultados obtidos por diferentes estudos tendo em vista os diferentes ambientes e principalmente a disponibilidade de dados. Os trabalhos mencionados sobre ML utilizaram bases de dados muito maiores em relação aquela disponível no momento da elaboração deste trabalho.

2. Métodos

Esta seção apresenta a metodologia utilizada para o desenvolvimento do trabalho detalhando a origem dos dados, tratamento e processamento dos dados.

2.1. Contexto

Este trabalho foi realizado em colaboração com o SAMU de Porto Alegre, Grupo Hospitalar Conceição (GHC) de Porto Alegre e Universidade do Vale do Rio dos Sinos, como parte do projeto Modelagem Semântica de Sepse. Este projeto foi submetido e aprovado pelo comitê de ética da Universidade e das instituições parceiras. Neste artigo foram utilizados dados coletados pela equipe de atendimento móvel do SAMU e do Hospital Nossa Senhora da Conceição (HNSC) do GHC. A Figura 1 apresenta um diagrama bastante simplificado do fluxo de atendimento do SAMU. Contudo, a compreensão deste fluxo de atendimento é relevante para o entendimento da origem dos dados utilizados neste trabalho.

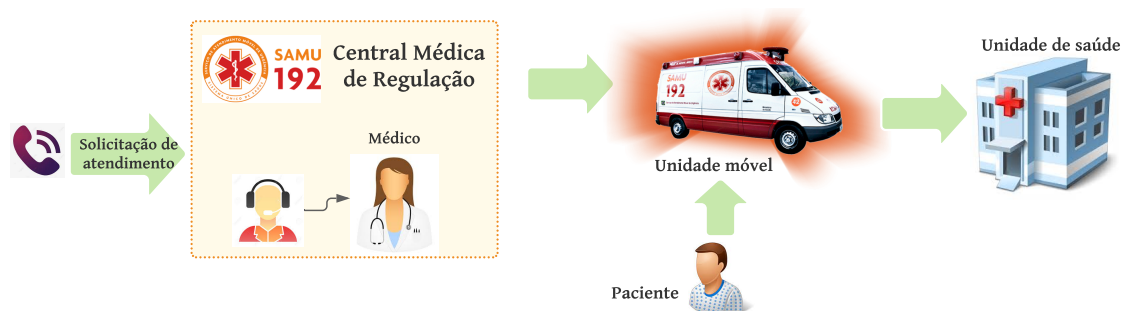


Figura 1. Diagrama do fluxo de atendimento do SAMU

As solicitações de atendimento que chegam ao SAMU são direcionadas para um profissional na CMR que avalia as informações preliminares e define sobre a necessidade de deslocamento de uma equipe de atendimento móvel. A equipe móvel analisa a situação do paciente, procede com a coleta e registro de informações do paciente e em comunicação com a CMR transporta o paciente para a unidade de saúde indicada para prosseguimento do atendimento. Durante este fluxo de trabalho diversas informações sobre o atendimento são registradas, sendo as variáveis de interesse deste estudo aquelas coletadas pela equipe da unidade móvel.

Cabe aqui destacar um problema complexo que consiste na falta de integração entre os sistemas computacionais do SAMU e das unidades de saúde para onde os pacientes são transportados. Este é um desafio em termos de análise de dados deste tipo de atendimento pré-hospitalar. Uma vez que a equipe do SAMU presta apenas o atendimento inicial é necessário obter o desfecho deste atendimento para treinar os algoritmos de aprendizado supervisionado.

Os dados utilizados neste artigo correspondem ao período de 01 de janeiro de 2017 até 31 de agosto de 2019. Dessa forma, o conjunto de dados analisado neste trabalho ainda não contempla casos de pacientes com COVID-19.

2.2. Obtenção dos dados

A obtenção do conjunto de dados utilizado neste artigo deriva de um processamento complexo necessário para superar a barreira de falta de integração entre os sistemas do SAMU e dos hospitais. Considerando que ambas instituições realizam muitos atendimentos a cada dia, o processo de cruzamento dos dados e identificação dos pacientes não pode ser totalmente automatizado. A Figura 2 apresenta as etapas do processamento e o resultado final, onde o conjunto de dados ficou restrito a 741 registros completos (todas as variáveis de interesse com valores válidos) para análise.

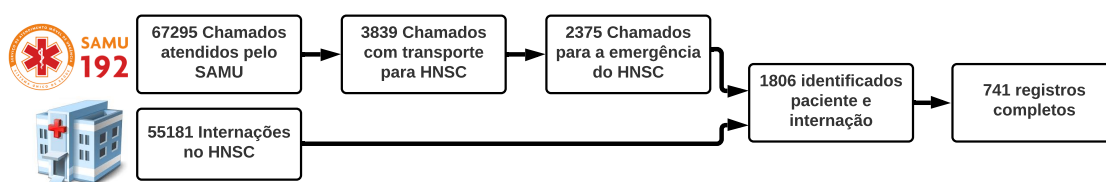


Figura 2. Processamento para obtenção de conjunto de dados

Neste estudo é importante identificar se o paciente atendido pelo SAMU desenvolveu sepse em algum momento durante o atendimento que se inicia com o SAMU e é concluído após a alta ou óbito na instituição de saúde. Assim, é necessário manter a rastreabilidade deste paciente através dos sistemas do SAMU e do hospital. Logo, foram efetuados filtros para selecionar os pacientes de interesse. Considerando todos os 3839 atendimentos que o SAMU realizou com transporte do paciente para o HNSC foram selecionados apenas aqueles direcionados para o setor de emergência (2375) descartando atendimentos como os obstétricos. Por consequência, foi necessário rastrear o 2375 através da base de dados do hospital para identificar se em algum momento do atendimento este paciente foi diagnosticado com sepse. Na ausência de uma chave compartilhada entre os sistemas foi elaborado inicialmente uma espécie de chave composta utilizando o nome completo do paciente, idade, data do chamado e data da internação para localizar este paciente em ambas as bases de dados. Porém, como existiam diversos erros tanto na grafia dos nomes quanto nas idades foi necessário uma segunda etapa manual do cruzamento dos dados baseada nas datas do chamado e da internação e em partes dos nomes.

O processo de cruzamento dos dados dos sistemas resultou em um conjunto de 1806 registros que foi reduzido para 741 devido aos dados faltantes nas variáveis de interesse. O atendimento do SAMU prevê que sempre sejam coletados dados dos pacientes

atendidos, no entanto em algum momento deste processo houve uma falha grave no registro destes dados no sistema eletrônico acarretando em prejuízos na análise dos dados. A Figura 3 apresenta um gráfico com os dados completos e faltantes.

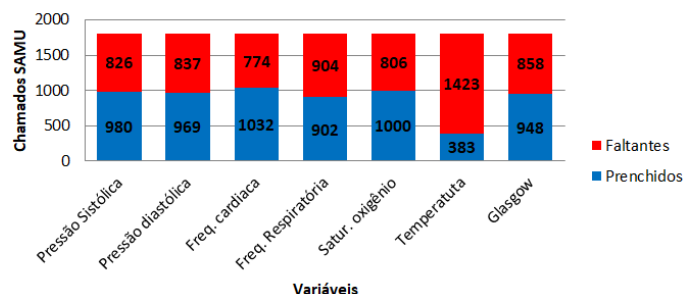


Figura 3. Registro das variáveis pelo SAMU

Em virtude do grande número de dados faltantes, a variável temperatura foi excluída da análise. A idade do paciente foi calculada a partir da data de nascimento armazenada no sistema do hospital. O desfecho de sepse é mais uma variável que foi obtida a partir da base de dados da instituição de saúde, porém esta informação foi extraída do conjunto de dados não estruturados do hospital que são compostos por textos de evolução e nota de alta do paciente. Para identificar o diagnóstico de sepse foi efetuado um processamento no texto da evolução e da nota de alta utilizando expressões regulares em busca de critérios como “sepse”, “choque séptico”, “septicemia” e os CID (Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde) relacionados com sepse.

2.3. Pré-processamento

Após a obtenção dos dados foi executada uma etapa de pré-processamento que consistiu na análise dos dados a fim de identificar as variáveis que poderiam ser utilizadas para treinar os modelos de ML. Nesta etapa também foi realizada a identificação dos dados com valores potencialmente incorretos e faltantes. Considerando que a base de dados é composta por uma única coleta de dados do paciente no atendimento pré-hospitalar a utilização de técnicas de imputação de dados faltantes torna-se bastante restrita. Dessa forma optou-se pela exclusão das variáveis que apresentavam uma grande quantidade de dados faltantes (colunas e linhas).

Alguns algoritmos de ML necessitam que as variáveis numéricas sejam padronizadas ou normalizadas para alterar a escala e evitar o comprometimento das métricas de avaliação dos modelos. Assim, foi realizada a padronização dos dados com a técnica de *z-score* que consiste em subtrair a média e dividir o valor pelo desvio padrão. Todo o pré-processamento, treinamento e avaliação dos modelos foi realizado utilizando a linguagem de programação Python, Scikit-Learn [Scikit-learn 2021], Jupyter notebook e o ambiente do Google Colab para executar algumas rotinas com maior demanda de capacidade de processamento.

2.4. Variáveis analisadas

Na Tabela 1 é apresentado um detalhamento das variáveis utilizadas para treinar os algoritmos de ML. Todas as variáveis preditoras selecionadas são numéricas e correspondem

basicamente aos sinais vitais coletados pela equipe do SAMU no local do atendimento. A variável de interesse ou desfecho é o diagnóstico de sepse que é extraído da base de dados do hospital.

Tabela 1. Detalhamento das variáveis utilizadas

Variável	Nome	Tipo de dado	Origem	Utilização no algoritmo
Frequência cardíaca	fc	Inteiro	SAMU	Preditora
Frequência respiratória	fr	Inteiro	SAMU	Preditora
Glasgow	glasgow	Inteiro	SAMU	Preditora
Idade	idade	Inteiro	Hospital	Preditora
Pressão arterial diastólica	paDiast	Inteiro	SAMU	Preditora
Pressão arterial sistólica	paSist	Inteiro	SAMU	Preditora
Saturação de oxigênio	satO2	Inteiro	SAMU	Preditora
Sepse	sepsse	Catégorica	Hospital	Desfecho

No total foram utilizadas sete variáveis preditoras e uma variável de desfecho. Como a variável de interesse possui apenas dois valores possíveis (0 que representa não sepse e 1 que representa sepse) caracteriza-se uma tarefa de classificação binária a ser processada e aprendida pelos algoritmos de ML.

2.5. Algoritmos de *machine learning*

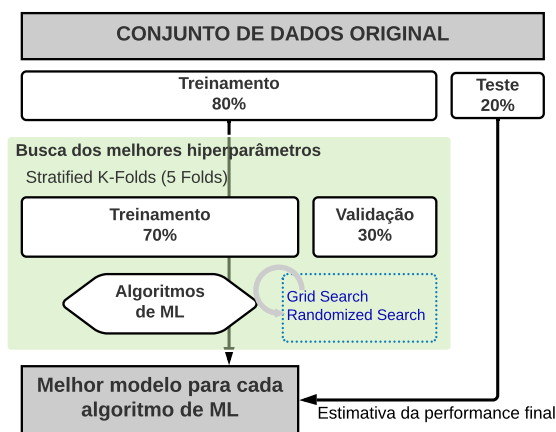
Para os testes realizados, foram selecionados algoritmos de ML com aprendizado supervisionado para executar a tarefa de classificação binária do problema (desenvolver ou não sepse). Optou-se por explorar diferentes algoritmos com o intuito de encontrar aquele com capacidade de gerar as melhores métricas de avaliação e cujos resultados pudessem ser mais facilmente interpretados. Além dos algoritmos de ML foi calculado o escore qSofa para ser utilizado como uma base de comparação uma vez que outros trabalhos relacionados já mencionados adotam a mesma prática. A lista abaixo visa apenas apresentar os algoritmos treinados sem a intenção de detalhar o funcionamento dos mesmos.

- Regressão logística: um algoritmo bastante conhecido na literatura e consiste em um modelo linear para classificação.
- Árvore de decisão: utiliza a estratégia de divisão de um problema complexo em vários mais simples.
- *Random Forest*: pode ser considerado uma evolução a árvore de decisão uma vez que emprega um comitê de árvores.
- *Gradient Boosting*: baseado na construção de vários modelos mais simples de árvore de decisão que agrupados tornam os resultados mais robustos.
- XGBoost: algoritmo baseado em comitê que combina as predições de um conjunto de classificadores. Este algoritmo tem se destacado em competições de ML apresentando bons resultados em problemas diversos [Chen and Guestrin 2016].
- LightGBM: algoritmo é mais rápido e escalonável que o XGBoost com a vantagem de processar variáveis categóricas sem a necessidade de conversão [Microsoft 2016].
- CatBoost: algoritmo desenvolvido recentemente e que tem apresentado resultados interessantes. É capaz de processar variáveis categóricas sem a necessidade de transformação [CatBoost 2021].

Conforme apresentado foram treinados modelos conhecidos e consolidados em diversos problemas de classificação binária como regressão logística e árvore de decisão e também alternativas mais recentes XGBoost, LightGBM e CatBoost. Os códigos desenvolvidos em Python com Jupyter Notebook para este trabalho podem ser acessados no repositório ¹.

2.6. Estratégia para treinamento e avaliação dos modelos

Devido ao conjunto pequeno de dados é necessário buscar formas de melhorar o desempenho dos algoritmos evitando o sobreajuste (*overfitting*) e subajuste (*underfitting*) do modelo. Para este estudo foi adotada a divisão do conjunto de dados original em treinamento e teste conforme apresentado na Figura 4. Para o teste final do modelo foi reservado 20% do conjunto original de dados.



Fonte: Adaptado de [Raschka and Mirjalili 2017].

Figura 4. Divisão do conjunto de dados em treinamento, validação e teste

Uma etapa importante do treinamento dos algoritmos consiste em encontrar a melhor configuração dos parâmetros de cada modelo, com o objetivo de o otimizar para melhor resolver o problema em questão. Para esta busca dos melhores parâmetros foram executadas as técnicas de *grid search* e *randomized search*. O *grid search* faz uma busca exaustiva com o conjunto de valores definidos para cada parâmetro. Já o *randomized search* faz um busca randomizada. Para esta investigação da melhor configuração foi executada uma validação cruzada estratificada (*Stratified K-Folds*) com a divisão do conjunto de dados de treinamento em cinco dobras (*fold*). Encerrada a pesquisa, o melhor modelo identificado (segundo critérios estabelecidos previamente) é treinado novamente com todo o conjunto de treinamento para então ser executada a avaliação final de desempenho com o conjunto de dados de teste.

2.7. Métricas utilizadas para avaliar os modelos

Para avaliar cada um dos modelos gerados e produzir subsídios para comparação, foram calculados algumas métricas bastante conhecidos para problemas de classificação binária:

- Matriz de confusão: esta matriz indica os erros e acertos do modelo.

¹<https://bitbucket.org/uhsospital/sepse-risk-assessment>

- Acurácia: indica quanto o modelo acertou nas classificações gerais.
- Precisão: dos pacientes que o modelo indicou que desenvolveriam sepse quantos realmente desenvolveram.
- Sensibilidade (*recall*): considerando os pacientes que desenvolveram sepse quanto o modelo acertou.
- Escore F1: é uma média harmônica entre a precisão e a sensibilidade.
- Especificidade: considerando os pacientes que não desenvolveram sepse quanto o modelo acertou.
- AUC: consiste em uma representação gráfica da sensibilidade versus especificidade.

3. Resultados

Após as etapas de preparação e pré-processamento, restaram 741 linhas no conjunto de dados que correspondem a chamados de pacientes com idade ≥ 18 . Cada linha corresponde a um chamado diferente atendido pelo SAMU com encaminhamento do paciente para o HNSC. Deste conjunto de dados, 149 registros são reservados a avaliação final do modelo e 592 são utilizados para a seleção dos melhores parâmetros e posterior treinamento do melhor modelo. A Tabela 2 apresenta algumas características dos conjuntos de dados após a divisão entre treinamento e teste. É perceptível que as classes são desbalanceadas, sendo a classe minoritária os casos de sepse.

Tabela 2. Conjunto de dados de treinamento e teste

Variável	Conjunto de dados	
	Treinamento	Teste
Sepse	241 (40,7%)	53 (35,6%)
Idade	65.4 ¹ [17.2] ²	65.7 ¹ [17.6] ²
Total	592	149

¹ Média das idades.

² Desvio padrão das idades.

Na Tabela 3, são apresentados os resultados da avaliação final dos melhores modelos treinados para cada algoritmo. As métricas do qSofa serão utilizadas como uma base de comparação para os modelos de ML.

Tabela 3. Resultados da avaliação final dos modelos

Modelo	F1	AUC	Acurácia	Precisão	Sensibilidade	Especificidade
qSofa	0,52	0,62	0,65	0,57	0,48	0,77
Árvore de decisão	0,45	0,7	0,64	0,5	0,42	0,77
CatBoost	0,46	0,7	0,68	0,59	0,38	0,85
Gradient Boosting	0,42	0,71	0,7	0,67	0,3	0,92
LightGBM	0,58	0,69	0,68	0,55	0,62	0,72
Random Forest	0,55	0,71	0,65	0,51	0,6	0,68
Regressão logística	0,45	0,73	0,68	0,59	0,36	0,86
XGBoost	0,52	0,71	0,69	0,58	0,47	0,81

O qSofa é um escore proposto em 2016 durante a última reunião de especialistas para atualizar a definição de sepse [Singer et al. 2016]. Este escore é simples suficiente para ser calculado por um profissional de saúde à beira do leito. O escore é baseado em

apenas três variáveis (frequência respiratória ≥ 22 ipm; pressão arterial sistólica ≤ 100 mmHg; alteração do nível de consciência). O qSofa é considerado positivo quando a soma da pontuação for ≥ 2 .

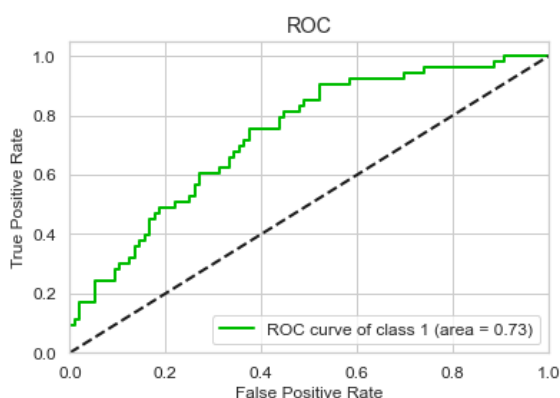
No entanto uma análise apenas das métricas de avaliação dos modelos pode não ser informativa suficiente, dessa forma na Tabela 4 são expostos os valores da matriz de confusão encontrada para cada um dos melhores modelos de ML.

Tabela 4. Matriz de confusão da avaliação final dos modelos

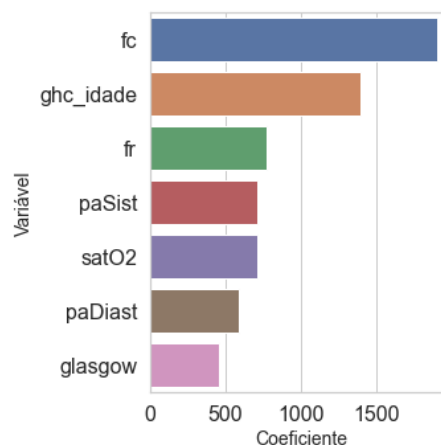
Modelo	VN ¹	FP ²	FN ³	VP ⁴
Árvore de decisão	74	22	31	22
CatBoost	82	14	33	20
Gradient Boosting	88	8	37	16
LightGBM	69	27	20	33
Random Forest	65	31	21	32
Regressão logística	83	13	34	19
XGBoost	78	18	28	25

¹ Verdadeiro Negativo. ² Falso Positivo. ³ Falso Negativo. ⁴ Verdadeiro Positivo.

Na Figura 5(a) é exposto a curva ROC do modelo de Regressão Logística que apresentou o melhor AUC entre os modelos testado. A Figura 5(b) apresenta um gráfico onde são exibidos os coeficientes calculados com o modelo LightGBM. Esta representação visual facilita a interpretação dos resultados gerados pelo modelo uma vez que é possível identificar a influência de cada variável no resultado final. Esta é uma característica comum de todos os modelos testados neste trabalho.



(a) Curva ROC gerada pelo modelo de Regressão Logística



(b) Coeficientes de relevância das variáveis para o modelo LightGBM

Figura 5. Representações gráficas das avaliações dos modelos de ML

4. Discussão

Um aspecto relevante deste trabalho é que o mesmo utiliza dados reais de um serviço de atendimento móvel pré-hospitalar que retrata a realidade brasileira. Os resultados aqui discutidos devem ser extrapolados para além da análise das métricas dos modelos de ML,

haja vista o longo caminho até obter os dados para processamento. Os algoritmos de ML são sensíveis à quantidade e qualidade dos dados disponíveis sendo necessária uma maior conscientização por parte dos profissionais que efetuam a coleta e o registro das informações.

Os dados retratados anteriormente na Tabela 3 através das métricas de avaliação elucidam que os modelos baseados em ML podem gerar resultados que superam o escore qSofa mesmo utilizando um conjunto restrito de variáveis numéricas e uma amostra bastante limitada. Quando analisamos a acurácia, o modelo baseado em *Gradient Boosting* apresenta um resultado 7.7% melhor em relação ao qSofa. Em termos de AUC a regressão logística apresenta um ganho de 17,7%. Todos os modelos baseados em ML produziram AUC superiores aos apresentados pelo escore qSofa. Estes resultados são influenciados pelas técnicas utilizadas para seleção dos melhores hiperparâmetros para cada um dos algoritmos de ML testados. Quando executamos a busca destes melhores hiperparâmetros para os algoritmos utilizamos as técnicas de *grid search* e *randomized search* sendo definida a AUC como métrica a ser maximizada na busca.

Sempre é importante considerar a natureza do problema para avaliar de forma clara os ganhos obtidos com os diferentes modelos. Analisando a matriz de confusão da Tabela 4 temos as informações quantitativas sobre os erros e acertos dos modelos. Sob este ponto de vista é relevante avaliar os valores de VP, VN e com especial atenção de FN. Embora um valor alto de FP não seja desejável, do ponto de vista do cuidado com a saúde o dano é menor quando o modelo erra indicando que um paciente possa desenvolver sepse quando na verdade ele não desenvolve. O início do tratamento de sepse geralmente envolve a administração de antibiótico e é notória a preocupação dos profissionais de saúde com o seu uso precipitado, pois pode agravar o problema de bactérias resistentes à medicação.

Já o valor de FN significa que o modelo indicou que o paciente não desenvolveria sepse enquanto que na verdade ele desenvolveu. Novamente sob a ótica do cuidado com a saúde, este é o erro mais grave, pois os profissionais de saúde podem não dispender a atenção que particularmente a sepse demanda com intervenções mais rápidas. Os ambientes hospitalares no Brasil geralmente possuem um grande número de pacientes aguardando atendimento e a identificação tardia dos sinais de sepse aumenta as chances de um desfecho desfavorável. Considerando a natureza do problema estudado, é necessário analisar para além das métricas puramente e buscar os ganhos em termos de maximizar os VP, VN e minimizar os valores de FN mesmo que isso implique em um FP maior. Assim, os modelos desenvolvidos com LightGBM e *Random Forest* apresentam métricas de VP e FN interessantes na matriz de confusão que de certa forma são sumarizados na métrica F1. Comparando F1 do LightGBM e do qSofa este ganho é de 11,5%. Qualquer melhora nos indicadores de validação é representativa, pois significa sistemas mais robusto de apoio à decisão para os profissionais de saúde.

Uma das limitações deste trabalho é o reduzido número de variáveis e principalmente de amostras que impactam negativamente na capacidade de treinamento dos modelos de ML. A falta de integração dos sistemas torna o processo de extração de informações moroso, o que justifica a escolha dos pesquisadores pela utilização de bases de dados públicas que muitas vezes foram geradas em ambientes onde a disponibilidade de recursos é ampla porém não representam a realidade do sistema de saúde brasileiro.

5. Conclusão

No presente trabalho foi possível demonstrar que mesmo com um conjunto reduzido de dados os algoritmos de ML apresentaram resultados melhores em relação ao escore heurístico empregado em situações análogas.

A partir do entendimento de que a sepse é uma síndrome tempo-dependente que implica em desfechos de saúde piores na medida em que o início do apropriado tratamento é retardado, a investigação de ferramentas de diagnóstico precoce de sepse mais precisas já no ambiente pré-hospitalar é justificável e desejável. O aprofundamento das pesquisas referentes à ocorrência de sintomas de sepse viral em pacientes com manifestações mais severas de COVID-19 abre novas possibilidades de pesquisas sobre a temática.

As próximas etapas da pesquisa envolvem a análise mais detalhada dos dados hospitalares disponíveis com o objetivo de treinar os mesmos algoritmos de ML para identificação precoce de sepse comunitária. Outra possibilidade interessante consiste na análise de dados hospitalares mais recentes que contemplem o atendimento dos paciente com COVID-19 e sintomas análogos à sepse viral.

6. Agradecimentos

Gostaríamos de registrar nosso agradecimento à empresa True Information Technology e ao Grupo Hospitalar Conceição por viabilizar o acesso aos dados analisados e pela disponibilidade em discutir novas ideias.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Processo 309537/2020-7).

Referências

- Barton, C., Chettipally, U., Zhou, Y., Jiang, Z., Lynn-Palevsky, A., Le, S., Calvert, J., and Das, R. (2019). Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Computers in biology and medicine*, 109:79–84.
- Brasil (2003). Portaria Nº 1.864, de 29 de setembro de 2003. *Ministério da Saúde*.
- Buchman, T. G., Simpson, S. Q., Sciarretta, K. L., Finne, K. P., Sowers, N., Collier, M., Chavan, S., Oke, I., Pennini, M. E., Santhosh, A., Wax, M., Woodbury, R., Chu, S., Merkeley, T. G., Disbrow, G. L., Bright, R. A., MaCurdy, T. E., and Kelman, J. A. (2020). Sepsis Among Medicare Beneficiaries: 3. The Methods, Models, and Forecasts of Sepsis, 2012-2018. *Critical Care Medicine*, 48:302–318.
- CatBoost (2021). Catboost. <https://catboost.ai/docs/concepts/about.html>. Acesso em 19-04-2021.
- Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM.
- Elezkurtaj, S., Greuel, S., Ihlow, J., Michaelis, E. G., Bischoff, P., Kunze, C. A., Sinn, B. V., Gerhold, M., Hauptmann, K., Ingold-Heppner, B., Miller, F., Herbst, H., Corman, V. M., Martin, H., Radbruch, H., Heppner, F. L., and Horst, D. (2021). Causes of death and comorbidities in hospitalized patients with COVID-19. *Scientific Reports*, 11(4263).
- Goh, K. H., Wang, L., Yeow, A. Y. K., Poh, H., Li, K., Yeow, J. J. L., and Tan, G. Y. H. (2021). Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nature communications*, 12(1).

- Jouffroy, R., Saade, A., Carpentier, A., Ellouze, S., Philippe, P., Idialisoa, R., Carli, P., and Vivien, B. (2019). Triage of Septic Patients Using qSOFA Criteria at the SAMU Regulation: A Retrospective Analysis. *Prehospital Emergency Care*, 22(1).
- Koyama, S., Yamaguchi, Y., Gibo, K., Nakayama, I., and Ueda, S. (2019). Use of prehospital qSOFA in predicting in-hospital mortality in patients with suspected infection: A retrospective cohort study. *PLoS One*, 14(5).
- Lane, D. J., Wunsch, H., Saskin, R., Cheskes, S., Lin, S., Morrison, L. J., and Scales, D. C. (2020). Screening strategies to identify sepsis in the prehospital setting: a validation study. *Canadian Medical Association journal*, 192(10):E230—E239.
- Li, H., Liu, L., Zhang, D., Xu, J., Dai, H., Tang, N., Su, X., and Cao, B. (2020). SARS-CoV-2 and viral sepsis: observations and hypotheses. *The Lancet*, 395:1517–1520.
- Mao, Q., Jay, M., Hoffman, J. L., Calvert, J., Barton, C., Shimabukuro, D., Shieh, L., Chettipally, U., Fletcher, G., Kerem, Y., Zhou, Y., and Das, R. (2018). Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ open*, 8(1).
- Microsoft (2016). Lightgbm. <https://www.microsoft.com/en-us/research/project/lightgbm>. Acesso em 19-04-2021.
- O'Dwyer, G., Konder, M. T., Reciputti, L. P., Macedo, C., and Lopes, M. G. M. (2017). O processo de implantação do Serviço de Atendimento Móvel de Urgência no Brasil: estratégias de ação e dimensões estruturais. *Cadernos de Saúde Pública*, 33(7).
- Pucchio, A., Eisenhauer, E. A., and Moraes, F. Y. (2021). Medical students need artificial intelligence and machine learning training. *Nature Biotechnology*, 39:388—389.
- Raschka, S. and Mirjalili, V. (2017). *Python Machine Learning*. Packt Publishing, second edition.
- Reyna, M. A., Josef, C. S., Jeter, R., Shashikumar, S. P. B., Westover, B., Nemati, S., Clifford, G. D. D., and Sharma, A. (2020). Early Prediction of Sepsis From Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. *Critical Care Medicine*, 48:210–217.
- Rudd, K. E., Johnson, S. C., Agesa, K. M., Shackelford, K. A., Tsoi, D., Kievlan, D. R., Colombara, D. V., Ikuta, K. S., Kissoon, N., Finfer, S., Fleischmann-Struzek, C., Machado, F. R., Reinhart, K. K., Rowan, K., Seymour, C. W., Watson, R. S., West, T. E., Marinho, F., Hay, S. I., Lozano, R., Lopez, A. D., Angus, D. C., Murray, C. J. L., and Naghavi, M. (2020). Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet*, 395:200–211.
- Scikit-learn (2021). Scikit-learn. <https://scikit-learn.org/>. Acesso em 19-04-2021.
- Shiraishi, D. and Marujo, E. (2020). Predição de sepsis em unidade de terapia intensiva: uma abordagem de aprendizado de máquina. In *Anais do XX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 416–421, Porto Alegre, RS, Brasil. SBC.
- Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G. R., Chiche, J.-D., Coopersmith, C. M., Hotchkiss, R. S., Levy, M. M., Marshall, J. C., Martin, G. S., Opal, S. M., Rubenfeld, G. D., van der Poll, T., Vincent, J.-L., and Angus, D. C. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8):801–810.
- World Health Organization (2020). Global report on the epidemiology and burden of sepsis: current evidence, identifying gaps and future directions. <https://apps.who.int/iris/bitstream/handle/10665/334216/9789240010789-eng.pdf>, Geneva.
- Zafer, M. M., El-Mahallawy, H. A., and Ashour, H. M. (2021). Severe COVID-19 and Sepsis: Immune Pathogenesis and Laboratory Markers. *Microorganisms*, 9(1).