

# Machine Learning aplicado à atenção domiciliar para predição de condição de óbito

Machine Learning applied to home care for predicting passing away conditions

Machine Learning aplicado a la atención domiciliar para predecir las condiciones de muerte

Recebido: 03/10/2022 | Revisado: 18/10/2022 | Aceitado: 20/10/2022 | Publicado: 25/10/2022

**Daniel Henrique Cordeiro Silva**

ORCID: <https://orcid.org/0000-0001-8981-3953>

Universidade Federal de Ouro Preto, Brasil

E-mail: [daniel.hcs@aluno.ufop.edu.br](mailto:daniel.hcs@aluno.ufop.edu.br)

**Elisa Maria do Nascimento Timo**

ORCID: <https://orcid.org/0000-0003-2307-4822>

Captamed, Brasil

E-mail: [elisa.nascimento@captamed.com.br](mailto:elisa.nascimento@captamed.com.br)

## Resumo

Nos processos de atenção domiciliar, onde pacientes são cuidados em casa por equipes de saúde multidisciplinares, diversos são os desafios para o gerenciamento e monitoramento à distância, não sendo raros os casos em que os recursos não são empregados nas situações realmente prioritárias. O advento da tecnologia, a disponibilidade de dados nos diversos sistemas de gestão e bem como as novas ferramentas de suporte à tomada de decisão trazem enormes possibilidades, retorno financeiro e maior conforto para pacientes e famílias. Este trabalho tem o objetivo de apresentar a aplicação de aprendizado de máquina, utilizando-se da metodologia CRISP-DM, para identificação de pacientes com maior chance de hospitalização ou óbito domiciliar.

**Palavras-chave:** Atenção domiciliar à saúde; Gestão em saúde; Ciência de dados; Aprendizado de máquina; Inteligência artificial.

## Abstract

In home care processes, where multidisciplinary health teams take care of their patients at home, there are several challenges for resource management and remote monitoring, where, sometimes, resources are not used in main priority situations. The advent of technology, the availability of data in management systems and the new decision-making support tools bring enormous possibilities, financial return and greater comfort for patients and families. This work aims to present the application of machine learning, using the CRISP-DM methodology, to identify patients with a greater chance of hospitalization or to pass away at home.

**Keywords:** Home care; Healthcare management; Machine learning; Data science; Artificial intelligence.

## Resumen

En los procesos de atención domiciliar, donde los pacientes son atendidos en el domicilio por equipos sanitarios multidisciplinares, existen varios retos para la gestión y seguimiento a distancia, no siendo raros los casos en los que no se utilizan recursos en situaciones realmente prioritarias. El advenimiento de la tecnología, la disponibilidad de datos en los distintos sistemas de gestión, así como las nuevas herramientas de apoyo a la decisión, traen enormes posibilidades, retorno económico y mayor cumplimiento para pacientes y familiares. Este trabajo tiene como objetivo presentar la aplicación del aprendizaje automático, utilizando la metodología CRISP-DM, para identificar pacientes con mayor probabilidad de hospitalización o muerte en el hogar.

**Palabras clave:** Atención domiciliar; Gestión de la salud; Ciencia de datos; Aprendizaje automático, Inteligencia artificial.

## 1. Introdução

Com as modificações do perfil demográfico populacional, o aumento crescente da expectativa de vida e dos custos em saúde surge a necessidade de modificar os modelos de atenção à saúde. Com isso, estratégias que sejam mais viáveis e com melhor resultados são então estruturadas, não só focadas no processo saúde/doença e na condição individual, mas trabalhando uma visão holística e descentralizada de atendimento.

Segundo o Ministério da Saúde, “A Atenção Domiciliar (AD) é uma forma de atenção à saúde, oferecida na moradia do paciente e caracterizada por um conjunto de ações de promoção à saúde, prevenção e tratamento de doenças e reabilitação,

com garantia da continuidade do cuidado e integrada à Rede de Atenção à Saúde.

As primeiras unidades de AD surgiram em 1947 nos Estados Unidos e tiveram como motivações iniciais reduzir a sobrecarga de internações hospitalares e proporcionar aos pacientes e familiares um ambiente psicológico mais favorável, conforme pode ser visto nos trabalhos de Ramallo (1998) e Rehem (2005).

Sob esta perspectiva, nota-se o atendimento domiciliar como uma estratégia transformadora, coerente e viável das práticas em saúde, onde é possível aperfeiçoar e se aproximar dos custos em saúde de maneira sólida; e ao mesmo tempo prestar uma assistência individualizada e humanizada ao paciente e familiares.

Com este recurso é possível aperfeiçoar e se aproximar dos custos em saúde de maneira sólida; e ao mesmo tempo prestar uma assistência individualizada e humanizada ao usuário, que consiste muito mais do que disponibilizar ao cliente um atendimento médico no âmbito do seu domicílio, tratando do envolvimento, planejamento, e coordenação de vários serviços multidisciplinares. Assim, o planejamento dos programas de atenção domiciliar, deve ser realizado através de uma equipe multiprofissional, considerando a realidade em que o paciente está inserido. Devem-se considerar as limitações e possibilidades do paciente, pois a atenção no espaço domiciliar visa à promoção, manutenção e/ou reabilitação da saúde e o desenvolvimento e adaptação de suas funções de maneira a favorecer o restabelecimento de sua independência e sua autonomia.

A implantação da AD no Brasil, como uma atividade planejada, ocorreu primeiramente no setor público, com a implantação do SAD do Hospital de Servidores Públicos do Estado de São Paulo, que funciona desde 1963. Entretanto a expansão da AD tem ocorrido principalmente no setor privado com cerca de 150 empresas prestadoras de AD cadastradas somente até o ano 2000 (Mendes Junior, 2000)

A utilização do domicílio como espaço de atenção à saúde atende a lógica para a racionalização de recursos e a constante busca da gestão em saúde para se obter uma assistência devida, com custos viáveis e tangíveis. Ao mesmo tempo esta prática se demonstra inovadora, conseguindo em muitas situações ser uma nova opção ao sistema hospitalocêntrico, que tem a lógica centrado na doença e no cuidado biomédico.

Vale ressaltar, que o gerenciamento e acompanhamento dos pacientes a distância, é um grande desafio e ferramentas de apoio à tomada de decisão, sustentadas pela transformação digital e técnicas de inteligência artificial tornam-se cada vez mais necessárias para priorização de pacientes, planejamento de recursos e melhoria no atendimento.

Como descreve Panesar (2019), provedores de saúde, indivíduos e organizações possuem uma fonte de dados que pode ser usado plenamente usado para suportar aplicações de aprendizado de máquina, e modelos de Machine Learning podem ser desenvolvidos para identificar padrões, anormalidades e destacar áreas que precisam de atenção, aumentando assim a precisão de diversos processos em saúde, levando a melhores resultados e incentivando o sucesso e melhora a satisfação das partes interessadas dos pacientes e profissionais de saúde.

Chen (2019), Rubinger (2022) e Malekloo (2022) apresentaram formas de implementar modelos de previsão em saúde, através da utilização de Machine Learning e Inteligência Artificial, todavia com maior foco em diagnósticos, enquanto Anderson (2022) mostra como resultados financeiros devem ser considerados.

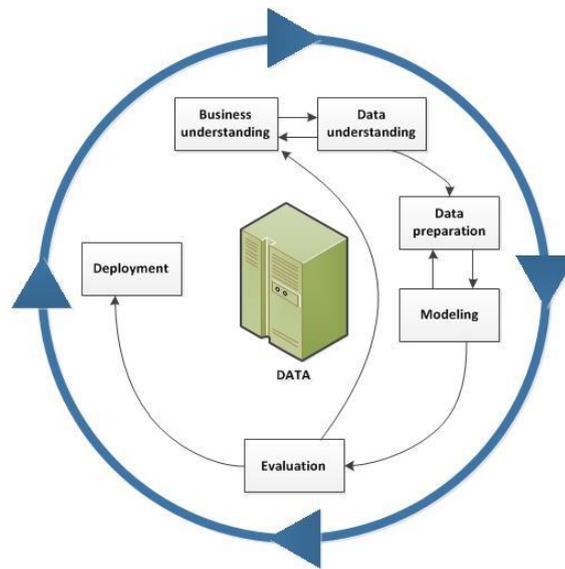
Neste trabalho, por sua vez, o problema a ser analisado será classificar os pacientes com probabilidade incorrer em óbito domiciliar ou hospitalar, com objetivo de proporcionar melhor priorização de recursos de atendimento e maior conforto às famílias e pacientes.

## 2. Metodologia

Este trabalho trata de um estudo observacional e baseia-se na extração de conhecimento de dados históricos de pacientes de uma dada empresa de atenção domiciliar, utilizando-se de algoritmos de aprendizado de máquina. Sendo assim, seu desenvolvimento seguiu a metodologia mais difundida para modelagem em ciência de dados, o Processo Padrão Inter-Indústrias

para Mineração de Dados, ou CRISP-DM (acrônimo para Cross-Industry Standard Process for Data Mining), conforme apresentado por Mariscal (2010) e atualmente amplamente documentada pela IBM (2022). O CRISP-DM, segundo descreve Niaksu (2015), é uma metodologia de uso geral que é independente da indústria, neutra em termos de tecnologia e é considerada padrão de fato para Mineração de Dados, conforme comprovado pelo próprio autor, que já aplicou o mesmo método em outras áreas (Silva, 2022). O objetivo final dos fundadores do CRISP-DM era criar um modelo de processo padrão não proprietário e disponível gratuitamente para engenharia de aplicação. A versão atual inclui a metodologia, o modelo de referência e o guia do usuário de implementação. A metodologia define fases, tarefas, atividades e saídas dessas tarefas. Conforme mostrado na Figura 1, trata-se de um fluxo de processo iterativo, com loops não estritamente definidos entre as fases.

**Figura 1:** Metodologia CRISP-DM.



Fonte: IBM.

Assim como fizeram Tavares e outros (2022), todos os dados foram analisados após um procedimento de desidentificação para garantir a privacidade dos dados em conformidade com as melhores práticas e legislação local.

Cada fase da metodologia é apresentada a seguir, já detalhando o respectivo desenvolvimento neste trabalho.

## 2.1 Entendimento do Negócio

A fase inicial consiste em entender o contexto do problema a ser abordado e definir os objetivos do projeto. Um resumo desta fase foi apresentado na primeira seção

### *Modalidades de Atenção Domiciliar*

Os serviços de atenção domiciliar são divididos em diferentes modalidades de atendimento que variam de acordo com a necessidade dos usuários, seja esta permanente ou provisória. Estes em sua maioria são divididos em cuidados crônicos, agudos ou intensivos.

Para atender a estas modalidades são criados pelas empresas prestadoras de atendimento domiciliar programas de atenção à saúde, onde os pacientes são divididos de acordo com o seu perfil epidemiológico. Na empresa foco desta pesquisa, os clientes têm a possibilidade de se enquadrar nas seguintes modalidades de assistência:

- Gerenciamento de Doenças

- Gerenciamento de Casos crônicos
- Assistência Domiciliar:
- Internação Domiciliar
- Cuidado Paliativo

Uma vez inseridos em seus programas de atenção os pacientes recebem através de suas equipes, um plano de visitas domiciliares de acordo com suas necessidades. Este por sua vez especifica a frequência da visita de cada especialidade e dimensiona o que o paciente necessita receber no tange a medicamentos, materiais e equipamentos.

### ***O Desafio***

Uma importante métrica para este negócio é o índice de hospitalizações (ou seja, paciente teve que deixar o cuidado domiciliar e ser encaminhado a um hospital). Outra situação monitorada é o índice de óbitos domiciliares, uma vez que é uma situação de maior humanização para o paciente e família.

A classificação dos pacientes com maior probabilidade de falecimento em domicílio, além de trazer benefícios às famílias e ao paciente, conforme supracitado, é informação estratégica para as operadoras de saúde, representando menores custos operacionais e demonstra efetividade no atendimento domiciliar.

### **2.2 Entendimento dos Dados**

Esta fase consiste em verificar o universo de informações disponíveis, sua origem e formato, bem como coletar, organizar e documentar todos os dados que se encontram disponíveis para realizar a análise exploratória.

No problema em questão, os dados são obtidos no sistema de gerenciamento da operação do prestador dos serviços de atenção domiciliar. Desta ferramenta, foram extraídas algumas informações de cadastro, diagnóstico, ocorrências diversas registradas, hospitalizações e outros eventos.

A Figura 2 apresenta as características (variáveis ou features) de cada registro da base de dados.

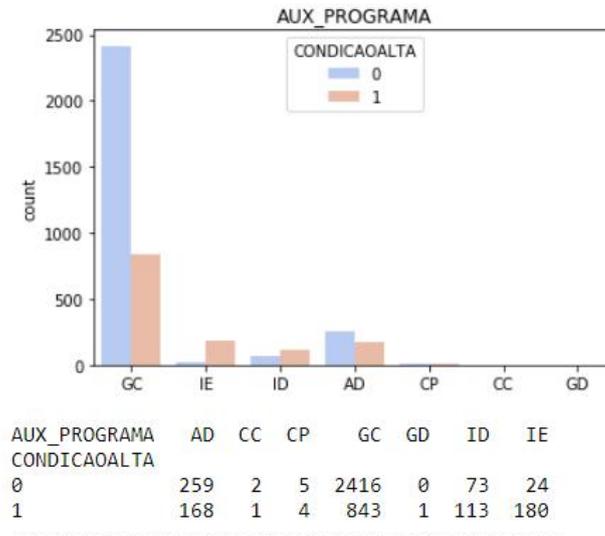
**Figura 2:** Features Disponíveis.

Feature	Descrição
SEXO	-
TEMPO ATEND	-
IDADE	-
AUX_PROGRAMA	Modalidade do programa de atenção (Categórica)
DSGNÓSTICO	Diagnóstico informado no cadastro em texto livre (nominal)
COD_REGIONAL	Resumo do CEP (Categórica)
TIPOSERVICO	Tipo de serviço (Categórica)
AUX_SEGURADORA	Seguradora de saúde (Categórica)
AUX_SCDIAGNOSTIC1NAME	Inicial do CID do diagnóstico principal (Categórica)
Administrativo	Numero de Ocorrências deste tipo.
Cancelamento de serviço	Numero de Ocorrências deste tipo.
Colocar texto do indice novo	Numero de Ocorrências deste tipo.
Consulta com Especialista	Numero de Ocorrências deste tipo.
Equipamento	Numero de Ocorrências deste tipo.
Exteriorização de dispositivo	Numero de Ocorrências deste tipo.
Falta de suprimentos	Numero de Ocorrências deste tipo.
Orientações	Numero de Ocorrências deste tipo.
Ouvidoria	Numero de Ocorrências deste tipo.
Profissional captamed - atraso	Numero de Ocorrências deste tipo.
Profissional captamed - falta	Numero de Ocorrências deste tipo.
Profissional captamed - reclamação	Numero de Ocorrências deste tipo.
Profissional captamed - solicitação	Numero de Ocorrências deste tipo.
Recusa de visita	Numero de Ocorrências deste tipo.
Remoção com prestador captamed	Numero de Ocorrências deste tipo.
Técnicos de enfermagem - atrasos	Numero de Ocorrências deste tipo.
Técnicos de enfermagem - reclamação	Numero de Ocorrências deste tipo.
Técnicos de enfermagem - solicitação	Numero de Ocorrências deste tipo.
Telemonitoramento	Numero de Ocorrências deste tipo.
Internação hosp. clínica médica urg	Numero de Ocorrências deste tipo.
Internação hosp. clínica médica ele	Numero de Ocorrências deste tipo.
Internação hosp. clínica cirúrgica	Numero de Ocorrências deste tipo.
Internação hosp. clínica cirúrgica	Numero de Ocorrências deste tipo.
Internação domiciliar	Numero de Ocorrências deste tipo.
Internação hosp. clínica CTI	Numero de Ocorrências deste tipo.
INTERC PRONT	Numero de Ocorrências deste tipo.
INTERC ATEND	Numero de Ocorrências deste tipo.
CONDICAOALTA	Indica óbito domiciliar (1 = obito dom.)
dtypes: float6	(28), object(7)

Fonte: Autores.

Observa-se a presença de features numéricas e categóricas relativas ao cadastro dos pacientes, bem como outras relativas ao número de ocorrências registradas para cada tipo. A Figura 3 apresenta como são distribuídos os óbitos de cada tipo por modalidade de programa de atenção domiciliar.

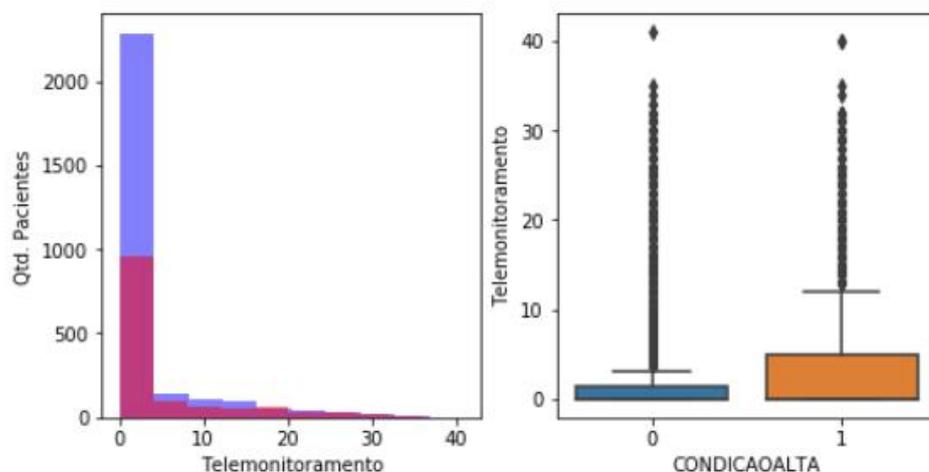
**Figura 3:** Distribuição dos óbitos por Programa de AD.



Fonte: Autores.

Nesta fase exploratória, as variáveis categóricas foram analisadas em diversas tabelas cruzadas e pela contagem de ocorrências de cada tipo de óbito por categoria. Da mesma forma, as features numéricas foram analisadas com o uso de histogramas e gráficos *box plot*. Na Figura 4, por exemplo, pode-se observar entre os pacientes que tiveram óbito domiciliar (CONDICAOALTA = 1) maior número de ocorrências de telemonitoramento (ou seja, chamadas realizadas pelos times de backoffice para os domicílios, no intuito de coleta de informações e orientação). Esta informação, por si só, já agrega valor ao negócio, antes mesmo da conclusão dos modelos, o que reforça que projetos de análise estatística e Machine Learning podem ter um caráter incremental.

**Figura 4:** Análise Mostra Influência do Telemonitoramento.



Fonte: Autores.

### 2.3 Preparação dos Dados

É nesta fase em que é criada a base de dados final. Ou seja, é feita a escolha dos dados que serão trabalhados, de como eles serão cruzados para resolver o problema da empresa, a preparação das databases e a definição do formato que será necessário para a análise.

Aqui é realizada a mineração de dados, considerando fatores como a relevância, a qualidade e as restrições técnicas, como limites no volume ou tipo de dados. Além disso, é preciso definir métodos de avaliação de desempenho para que seja realizada nas próximas fases.

Neste trabalho, foram seguidos os seguintes passos nesta fase:

- Tratamento de dados faltantes: Dados faltantes são representados pelo `NaN` (*Not a Number*). Quando isso ocorreu para uma dada features durante esta preparação, o referido registro teve o valor da feature substituído pela média ou valor mais frequente.
- Higienização: retiradas features categóricas com única categoria com mais de 90% das ocorrências e variáveis numéricas com baixa variação ou excessivo número de outliers.
- Divisão em entradas (features) e saída (CONDICAOALTA).
- Palavras-chave no campo DIAGNÓSTICO, criando nova feature categórica com grupos identificados.
- Conversão de variáveis categóricas em variáveis numéricas binárias.
- Unificação de categoria PROGRAMA\_ID (Internação Domiciliar) e PROGRAMA\_IE
- Divisão entre treinamento e teste.
- Normalização unitária de dados numéricos.

### 2.4 Modelagem

Aqui são selecionadas e aplicadas as técnicas de Data Mining mais apropriadas, com base nos objetivos identificados na primeira fase. A partir de agora, a mineração de dados pode ser associada a análises preditivas para que a empresa preveja as tendências dos negócios. É de bom costume escolher mais de um modelo para medir seu desempenho de performance computacional e o resultado esperado (ex: acurácia).

Independentemente da técnica de aprendizado de máquina utilizada, a abordagem se baseará em três fases: treinamento, validação e teste. Cada etapa, contando com um subconjunto de dados de cada um dos três principais conjuntos de dados.

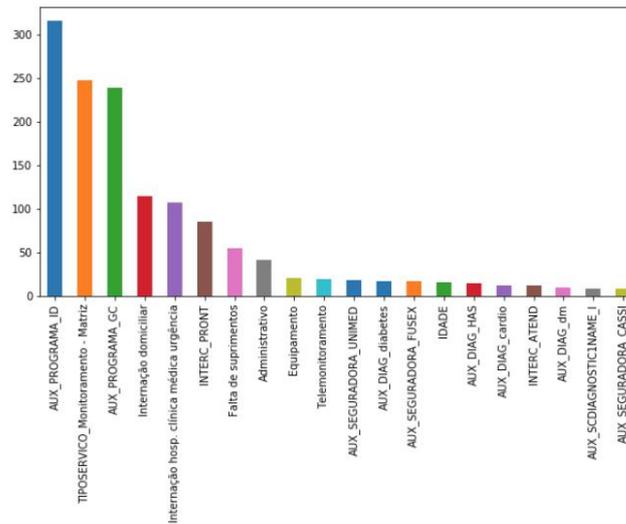
Em geral, na primeira etapa, os parâmetros do modelo são ajustados usando um algoritmo de otimização que visa minimizar a medição de erros, usando a diferença entre a produção estimada e a esperada, também chamada de técnica de aprendizado supervisionado.

O segundo passo é aplicado para ajustar os hiperparâmetros, a fim de superar comportamentos tendenciosos ou superajustados. É seguido por uma comparação entre os resultados de cada tentativa e uma seleção dos que melhor se ajustam aos dados. É possível ainda mesclar os subconjuntos de treinamento e validação para usar a técnica de validação cruzada.

Por fim, o desempenho do modelo treinado é avaliado com dados novos, que simulam um contexto real, em um conjunto de dados menor, mas que segue a mesma distribuição de probabilidade da fase de treinamento.

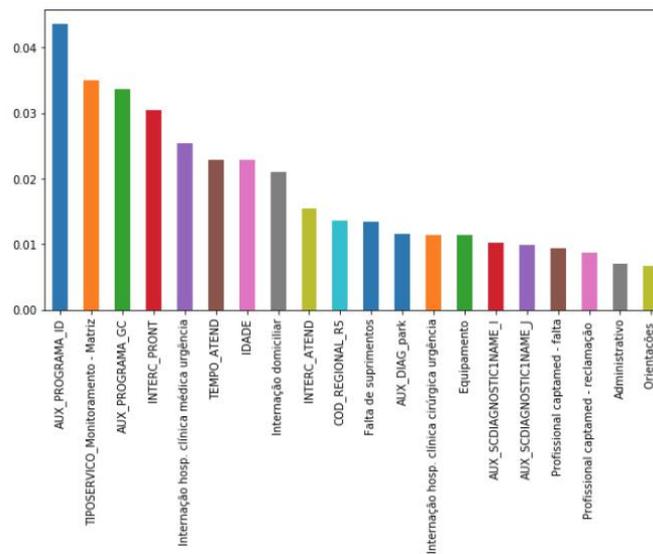
Antes da construção dos modelos, foi realizada análise de seleção de features, aplicando-se métodos que avaliam a importância de cada variável em relação ao target CONDICAOALTA. Foi utilizado um método de seleção univariada, que funciona selecionando os melhores recursos com base em testes estatísticos, bem como o mutual test (Pedregosa, 2011), que detecta qualquer tipo de relação entre as variáveis, incluindo as não lineares conforme apresentado nas Figuras 5 e 6.

**Figura 5:** Feature selection univariada.



Fonte: Autores.

**Figura 6:** Feature Selection Mutual Information.



Fonte: Autores.

Após o trabalho de seleção das variáveis, os algoritmos Random Forest e Multi-Layer Perceptron foram escolhidos para tentar resolver o Problema de Previsão de Óbito Domiciliar.

Os modelos utilizados, baseados nos algoritmos citados, apesar de serem descritos por equações matemáticas ou por operações condicionantes, pode ser de alta complexidade descrevê-los, principalmente para o leigo, ou até mesmo o cliente de negócio em questão, e acabam sendo tratados como “caixas-pretas”. Todavia, como mostrou McCoy (2022), e também London (2019) em seu texto “*Believing in black boxes*”, o valor da explicabilidade do aprendizado de máquina não é intrínseco, mas sim um instrumento para alcançar objetivos maiores, como desempenho e confiança.

### Random Forest

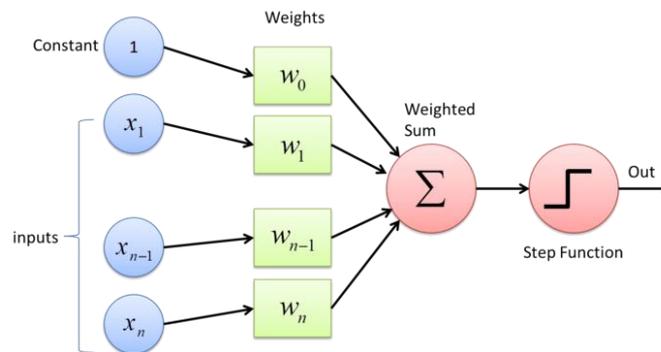
A floresta aleatória, ou Random Forest, é uma técnica de aprendizado de máquina que também pode ser usada para

problemas de regressão e classificação. Ele constrói várias árvores de decisão no tempo de treinamento e produz a previsão média (regressão) das árvores individuais. Conforme apresentado na documentação da biblioteca *Scikit Learn* para desenvolvimento de modelos na linguagem de programação Python (Pedregosa, 2011), “consiste em um número escolhido de árvores de decisão. Cada um dos modelos de árvore de decisão é aprendido em um conjunto diferente de linhas (registros) e em um conjunto diferente de colunas (descrevendo atributos), pelo qual o último também pode ser um descritor de vetor de bits ou de vetor de bytes (por exemplo, impressão digital molecular). Os conjuntos de linhas para cada árvore de decisão são criados pela inicialização e têm o mesmo tamanho da tabela de entrada original. Para cada nó de uma árvore de decisão, um novo conjunto de atributos é determinado com uma amostra aleatória do tamanho  $\sqrt{m}$  em que  $m$  é o número total de atributos. O modelo de saída descreve uma floresta aleatória e é aplicado no nó preditor correspondente usando um voto majoritário simples.”

### Rede Neural Artificial

O Perceptron foi introduzido por McCulloch (1943) como uma versão computacional de um neurônio biológico, que pode ser entendida como uma soma ponderada das entradas usadas em uma função de ativação para calcular uma saída, como mostra a Figura 7.

**Figura 7:** Neurônio Artificial (Perceptron).



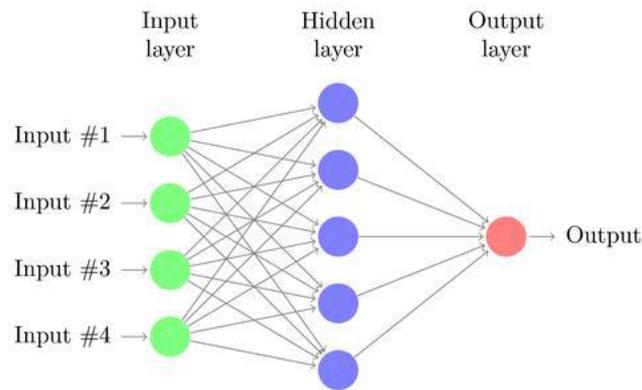
Fonte: [https://medium.com/@tech\\_savvy/ml-scientist-day-3-notes-neural-networks-4f63dfe5be15](https://medium.com/@tech_savvy/ml-scientist-day-3-notes-neural-networks-4f63dfe5be15)

Com base nessa abstração, Minsky e Papert (1969) propuseram perceptrons organizados em camadas interconectadas e confirmaram as Redes Neurais Artificiais (RNA) com mais de uma camada oculta como uma ferramenta poderosa capaz de resolver problemas de separação não linear.

Uma rede Perceptron Multicamadas (MLP), é constituída por uma camada de entrada para receber o sinal (ou seja, receber as diversas variáveis disponíveis para o modelo), camadas escondidas e a camada de saída. Cada camada escondida possui determinado número de neurônios, como visto anteriormente (composto por peso, bias e função de ativação). Uma rede MLP é inicializada com um dado vetor de entrada, que é multiplicado pelos pesos de cada camada, e reavaliado pela função de ativação, propagando para a camada seguinte até que a camada de saída seja atingida (Goodfellow, 2016). Uma possível rede MLP pode ser ilustrada pela Figura 8. Nela, a camada de entrada possui quatro elementos, há somente uma camada escondida formada por cinco neurônios e saída converge para um elemento.

A escolha da topologia da rede está relacionada à complexidade do problema e por isso é um desafio definir tais parâmetros. Neste trabalho, foi selecionada uma arquitetura com duas camadas escondidas, com 96 e 32 unidades (neurônios), utilizando-se função de ativação sigmoid e taxa de dropout de 20%.

**Figura 8:** Multi-Layer Perceptron.



Fonte: [https://medium.com/@tech\\_savvy/ml-scientist-day-3-notes-neural-networks-4f63dfe5be15](https://medium.com/@tech_savvy/ml-scientist-day-3-notes-neural-networks-4f63dfe5be15)

## 2.5 Avaliação dos Modelos

A avaliação dos modelos foi feita dando continuidade aos passos seguintes da metodologia CRISP-DM. Nessa fase, é preciso avaliar os resultados e levantar todas as possibilidades de variações que os dados possam ter, analisando se existe algum fator que tenha sido negligenciado e até que ponto o modelo atende aos objetivos de negócios. Caso o modelo não apresente um resultado satisfatório, devemos retomar a primeira etapa para entender melhor o negócio e os dados, ciclo que se repetiu algumas vezes neste trabalho.

Os resultados obtidos para cada modelo são apresentados a seguir, nos resultados, utilizando-se da avaliação da métrica de acurácia, além das métricas de precisão e *recall*, que podem ser lembradas no *Google Machine Learning Crash Course* (2022).

## 2.6 Implementação

Após validar e testar um modelo, caso as métricas de desempenho atinjam patamares que realmente tenham potencial negócio, são implementadas soluções para que as previsões sejam executadas de forma contínua para que os modelos possam realmente agregar valor ao processo produtivo.

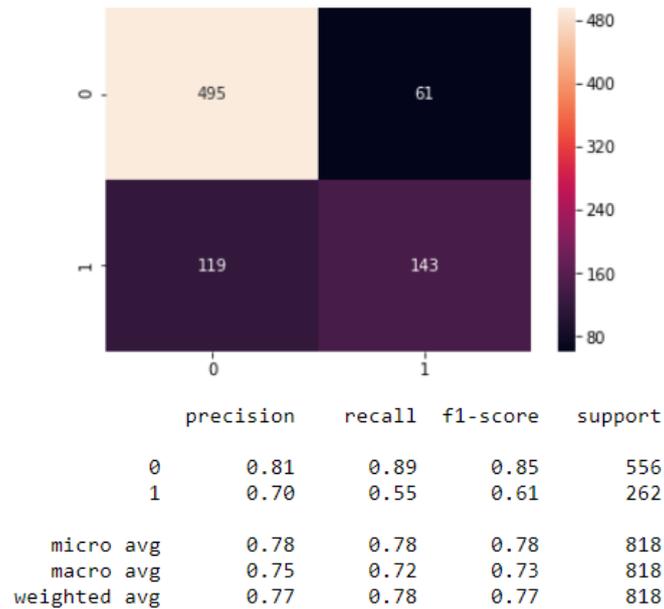
No caso deste trabalho, esta solução poderia ser uma arquitetura simples que faria a leitura dos dados de entrada, diretamente do sistema PIMS, executaria o script Python com o modelo, e retornaria a previsão em forma de escrita no próprio históricos. Todavia, como se trata de uma prova de conceito, a solução, no momento da escrita deste artigo, ainda não fora implementada.

## 3. Resultados e Discussão

### *Resultados para Random Forest*

A aplicação do algoritmo de classificação por floresta randômica resultou em uma acurácia de 78% de acerto na previsão de óbitos domiciliares. A Figura 9 apresenta a matriz de confusão que avalia o resultado do modelo para os dados de teste, onde “1” representa o óbito domiciliar e, conseqüentemente, “0” indica óbito hospitalar.

**Figura 9:** Resultados Modelo Random Forest.



Fonte: Autores.

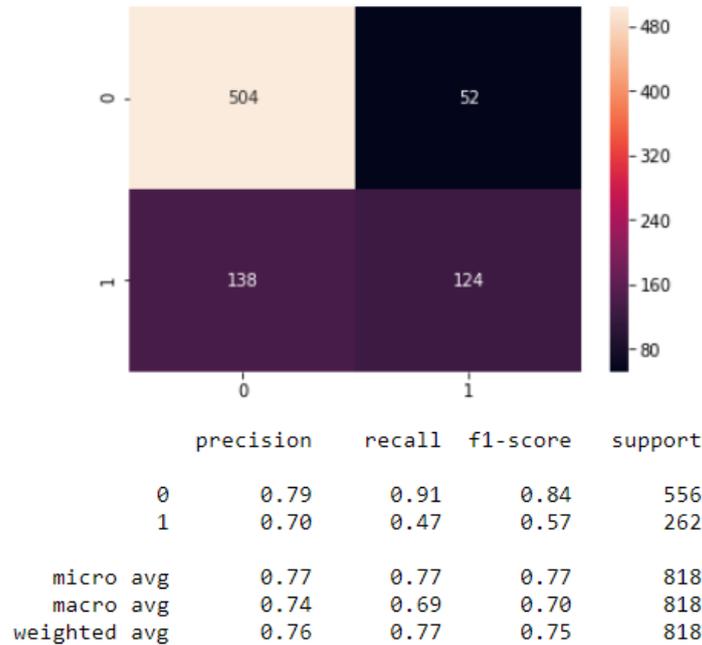
### **Resultados para Rede Neural**

A rede neural treinada com os dados em questão resultou em uma acurácia de 76% de acerto na predição de óbitos domiciliares. A Figura 10 apresenta a matriz de confusão que avalia o resultado do modelo para os dados de teste, onde “1” representa o óbito domiciliar e, conseqüentemente, “0” indica óbito hospitalar.

O desempenho do modelo com RNA ficou abaixo do indicador obtido com modelo Random Forest e aquém do que se esperava, conforme apresenta a Figura 11.

De qualquer forma, as métricas de recall e precision observadas, especialmente para a situação alvo de óbito hospitalar (CONDICAOALTA = 0), apresentam altos valores (próximos ou acima de 80%).

**Figura 10:** Resultados Modelo Redes Neurais.



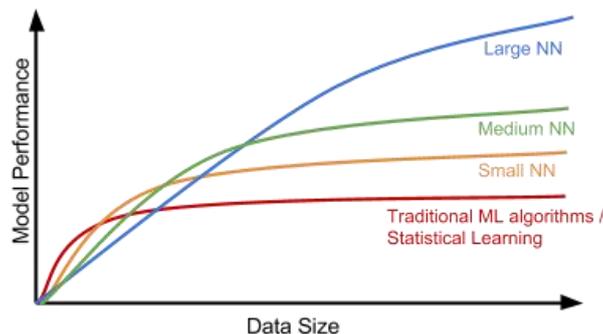
Fonte: Autores.

#### 4. Considerações Finais

Considerando que as métricas para identificação dos pacientes com maior probabilidade de óbito hospitalar possuem valores razoáveis (recall e precision próximos a 90%), este modelo pode ser de grande utilidade, uma vez que esta situação pode indicar tendência de hospitalização do paciente. Ou seja, pode ser implementado um modelo preditivo, a ser processado semanalmente, por exemplo, para suporte à tomada de decisão, indicando os pacientes com maior probabilidade de hospitalização, sendo aqueles para os quais deveria ser priorizado o atendimento naquele período, focando seus profissionais nas famílias que mais precisam do cuidado.

Acredita-se que o processo de modelagem pode ter sido prejudicado pelo fato de a base de dados utilizada possuir registros com entrada manual e ainda incluir campos de texto livre. Além disso, após análise mais aprofundada, percebeu-se que não foi possível obter melhor performance com o modelo de redes neurais, uma vez que este tipo de modelo atinge seu potencial para grandes massas de dados. Neste trabalho, tinha-se um universo de aproximadamente 4 mil registros de óbito, o que pode ser considerada uma base de dados pequena, justificando a melhor performance do modelo Random Forest. A Figura 11 ilustra o desempenho de diversos modelos pela quantidade de dados.

**Figura 11:** Performance X Quantidade de Dados.



Fonte: <https://lilianweng.github.io/posts/2017-06-21-overview/>

Como estudo futuro, a inclusão de novas features e outras ferramentas de feature selection, podem ajudar a melhorar a performance dos modelos nos próximos ciclos da metodologia.

De qualquer forma, é preciso definir uma estratégia para a implantação de tais projetos, visando realmente melhorar o lado operacional do negócio. Analisar novas estratégias para desenvolver modelos que ajudem a prever intercorrências e hospitalizações também seriam de imenso valor para o gerenciamento dos serviços.

## Referências

- Chen, P. H. C., Liu, Y., & Peng, L. (2019) How to develop machine learning models for healthcare. *Nat. Mater.* 18, 410–414. <https://doi.org/10.1038/s41563-019-0345-0>,
- IBM (2022) *Introduction to CRISP-DM*. <<https://www.ibm.com/docs/en/spss-modeler/18.2.0?topic=guide-introduction-crisp-dm>>
- Panesar, A. (2019) *Machine learning and AI for healthcare*. Coventry, UK: Apress, 2019.
- Mariscal, G., Marban, O., & Fernandez, C. (2010). *A survey of data mining and knowledge discovery process models and methodologies*. *The Knowledge Engineering Review*, 25(2), 137-166.
- Rehem, T. C. M. S. B., & Trad, L. A. B. (2005). Assistência domiciliar em saúde: subsídios para um projeto de atenção básica brasileira. *Ciência & Saúde Coletiva*, 10, 231-242. REHEM & TRAD, 2005.
- Mendes Júnior, W. V. (2000). Assistência domiciliar: uma modalidade de assistência para o Brasil? Dissertação de Mestrado, Universidade Estadual do Rio de Janeiro, Rio de Janeiro, Brasil.
- Ramallo, V. J. G., & Tamayo, M. I. P. (1998). Historia de la hospitalización a domicilio, pp. 13-22. In MDD Glez (coord.). *Hospitalización a domicilio*. Hoechst Marion Roussel, Espanha.
- Google (2022) *Machine Learning Crash Course*. <<https://developers.google.com/machine-learning/crash-course/>>
- Pedregosa. et al.,(2011). *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research*, *Journal of Machine Learning Research*, 12, pp. 2825-2830.
- Minsky, M., & Papert, S. (1969). *Perceptrons*. M.I.T. Press. EUA.
- McCulloch, W. S., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. *Bulletin of Mathematical Biophysics* 5, 115–133.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press. EUA.
- Niaksu, O. (2015). CRISP Data Mining Methodology Extension for Medical Domain. *Baltic J. Modern Computing*. 3. 92-109.
- Tavares, L. D., Manoel, A., Donato, T. H. R., Cesena, F., Minanni, C. A., Kashiwagi, N. M., & Szlejf, C. (2022). Prediction of metabolic syndrome: A machine learning approach to help primary prevention. *Diabetes Research and Clinical Practice*, 191, 110047.
- Malekloo, A., Ozer, E., AlHamaydeh, M., & Girolami, M. (2022) Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Structural Health Monitoring*. 21(4):1906-1955.
- McCoy, L. G., Brenna, C. T., Chen, S. S., Vold, K., & Das, S. (2022). Believing in black boxes: Machine learning for healthcare does not need explainability to be evidence-based. *Journal of clinical epidemiology*, 142, 252-257.
- Anderson, D., Bjarnadottir, M. V., & Nenova, Z. (2022). Machine learning in healthcare: Operational and financial impact. In *Innovative Technology at the Interface of Finance and Operations* (pp. 153-174). Springer, Cham.
- Rubinger, L., Gazendam, A., Ekhtiari, S., & Bhandari, M. (2022). Machine learning and artificial intelligence in research and healthcare. *Injury*. ISSN 0020-1383
- Silva, D. H. C., Alves, V. K., & Savio, E. (2022). Redes neurais artificiais aplicadas à moagem de minério de ferro combinadas a modelos empíricos. *Research, Society and Development*, 11(13), e84111332329-e84111332329.
- London, A. J. (2019). Artificial intelligence and black-box medical decisions: accuracy versus explainability. *Hastings Center Report*, 49(1), 15-21.