

**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO**

Gleddson Fryttys Menezes Leite

**Previsão, por meio de análise de Séries Temporais, de
leitos de Unidades de Terapia Intensiva - UTI na rede de
atendimento do Sistema Único de Saúde - SUS no
município de Goiânia.**

São Carlos

2021

Gleddson Fryttys Menezes Leite

Previsão, por meio de análise de Séries Temporais, de leitos de Unidades de Terapia Intensiva - UTI na rede de atendimento do Sistema Único de Saúde - SUS no município de Goiânia.

Trabalho de conclusão de curso apresentado ao Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, como parte dos requisitos para conclusão do MBA em Ciências de Dados.

Área de concentração: Ciências de Dados

Orientador: Prof. Dr. Antônio Castelo Filho

Versão original

São Carlos

2021

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

L533p Leite, Glleddson Fryttys Menezes
Previsão, por meio de análise de Séries
Temporais, de leitos de Unidades de Terapia
Intensiva - UTI na rede de atendimento do Sistema
Único de Saúde - SUS no município de Goiânia. /
Glleddson Fryttys Menezes Leite; orientador Antônio
Castelo Filho. -- São Carlos, 2021.
52 p.

Trabalho de conclusão de curso (MBA em Ciência
de Dados) -- Instituto de Ciências Matemáticas e de
Computação, Universidade de São Paulo, 2021.

1. Ciência de Dados. 2. Aprendizado de Máquina.
3. Séries Temporais. 4. Redes Neurais. I. Castelo
Filho, Antônio, orient. II. Título.

Dedico este trabalho à minha esposa Aline, amiga, parceira e inspiração em todos os momentos.

AGRADECIMENTOS

Agradeço à Controladoria-Geral da União pela oportunidade de realização do curso, órgão este que tem conduzido com excelência o combate à corrupção no Brasil.

Agradeço, também, aos professores do ICMC/USP pela maestria na condução das disciplinas.

E, por fim, agradeço ao meu orientador por toda a compreensão e paciência.

“E, o mais importante, tenha a coragem de seguir seu coração e sua intuição. De alguma forma, eles já sabem o que você realmente quer se tornar. Tudo o resto é secundário.”

Steve Jobs

RESUMO

LEITE, G.F.M. **Previsão, por meio de análise de Séries Temporais, de leitos de Unidades de Terapia Intensiva - UTI na rede de atendimento do Sistema Único de Saúde - SUS no município de Goiânia.** 2021. 52p. Monografia (MBA em Ciências de Dados) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2021.

A previsão da demanda sobre a sobre a taxa de ocupação hospitalar do SUS é um importante fator de impacto no desempenho operacional dos hospitais públicos, sendo que fragilidades no planejamento de leitos pode ocasionar prejuízos sociais e financeiros ao Estado. Neste sentido, este trabalho elabora um modelo baseado em séries temporais para previsão das diárias de leitos das Unidade de Tratamento Intensivo – UTI no município de Goiânia em determinado mês. Para tanto, foram elaborados e comparados modelos probabilísticos, tais como ARIMA Sazonal, e modelos dinâmicos, como as arquiteturas de redes LSTM.

Palavras-chave: Ciência de Dados. Aprendizado de Máquina. Séries Temporais. Long Short-term Memory. ARIMA. Sistema de Informações Hospitalares. Sistema Único de Saúde. Unidade de Terapia Intensiva.

ABSTRACT

LEITE, G.F.M. **Forecasting, through Time Series analysis, of beds of Intensive Care Units - ICU in the service network of the Brazilian Unique Health System in the city of Goiânia.** 2021. 52p. Monografia (MBA em Ciências de Dados) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2021.

Forecasting demand on the Brazilian Unique Health System hospital occupancy rate is an important impact factor on the operational performance of public hospitals, and weaknesses in bed planning can cause social and financial losses to the country. In this sense, this work elaborates a model based on time series to forecast the bed rates of the Intensive Care Units - ICU in the city of Goiânia in a given month. For that, probabilistic models, such as ARIMA Sazonal, and dynamic models, such as LSTM network architectures, were developed and compared.

Keywords: Data Science. Machine Learning. Time Series. Long Short-term Memory. ARIMA. Hospital Information System. Brazilian Unique Health System. Intensive Care Unit.

LISTA DE FIGURAS

Figura 1 – Série de diárias de UTI Adulto Tipo I, II e III no município de Goiânia com observações mensais de janeiro de 2009 até dezembro de 2019.	34
Figura 2 – Decomposição Série Temporal nas componentes de tendência, sazonalidade e ruído.	35
Figura 3 – Correlograma das Função de Autocorrelação e de Autocorrelação Parcial da Série.	36
Figura 4 – Média Móvel Simples com janela de 12 meses e Naive Forecasting comparados com os dados observados no ano de 2019.	37
Figura 5 – Previsão de diárias de UTI no município de Goiânia para o ano de 2019 como os modelos de Suavização Exponencial de Holt e Sazonal de Holt-Winters comparados com os dados observados no ano de 2019.	38
Figura 6 – Previsão de diárias de UTI no município de Goiânia para o ano de 2019 com o modelo <i>XGBoost (Gradient Boosting)</i> comparados com os dados observados no ano de 2019.	39
Figura 7 – Execução do passo-a-passo no PMDARIMA (2020) com vistas a minimizar o AIC para um modelo do tipo ARIMA com sazonalidade - SARIMA.	40
Figura 8 – Previsão de diárias de UTI no município de Goiânia para o ano de 2019 com Modelo SARIMA(1,1,0)x(3,1,2) ₁₂ comparados com os dados observados no ano de 2019.	41
Figura 9 – Raiz do Erro Quadrático Médio entre as previsões geradas pelos modelos SEH, HW, MMS-12, XGB, SARIMA(1,1,0)x(3,1,2) ₁₂ e Naive Forecasting e os dados de teste (observações do ano de 2019)	42
Figura 10 – Comparativo, com dados escalados, entre a quantidade de Diárias de UTI e a População Estimada de Goiânia dos anos de 2009 até 2019.	42
Figura 11 – Execução do passo-a-passo no PMDARIMA (2020) com vistas a minimizar o AIC para um modelo do tipo ARIMA com sazonalidade - SARIMA e inclusão de variável exógena.	43
Figura 12 – Previsão de diárias de UTI no município de Goiânia para o ano de 2019 com Modelo SARIMA(0,1,1)x(3,1,0) ₁₂ , considerando a intervenção da população do município, comparados com os dados observados no ano de 2019.	44
Figura 13 – Raiz do Erro Quadrático Médio entre as previsões geradas pelos modelos LSTM e os dados de teste (observações do ano de 2019)	44
Figura 14 – Previsão de diárias de UTI no município de Goiânia para o ano de 2019 com Modelo LSTM comparados com os dados observados no ano de 2019.	45

Figura 15 – Raiz do Erro Quadrático Médio entre as previsões pelos modelos e os dados de teste (observações do ano de 2019) 47

LISTA DE ABREVIATURAS E SIGLAS

AIC	<i>Akaike Information Criterion</i>
AIH	Autorização de Internação Hospitalar
API	<i>Application Programming Interface</i>
ARIMA	Modelo Autorregressivo Integrado de Médias Móveis
ANS	Agência de Saúde Suplementar
CID	Classificação Internacional de Doenças e Problemas Relacionados à Saúde
CNRAC	Centro Nacional de Regulação da Alta Complexidade
DATASUS	Departamento de Informática do SUS
FAEC	Fundo de Ações Estratégicas e Compensação
HW	Suavização Exponencial Sazonal de Holt-Winters
IBGE	Instituto Brasileiro de Geografia e Estatística
LSTM	<i>Long Short-Term Memory</i>
MMS	Média Móvel Simples
MS	Ministério da Saúde
RNN	Rede Neural Recorrente
SARIMA	Modelo Autorregressivo Integrado de Médias Móveis com Sazonalidade
SIGTAP	Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS
SEH	Suavização Exponencial de Holt
SIHSUS	Sistema de Informações Hospitalares do SUS
SUS	Sistema Único de Saúde do Brasil
TABNET	Sistema de Informações de Saúde do Datasus
UTI	Unidade de Terapia Intensiva

SUMÁRIO

	Sumário	17
1	INTRODUÇÃO	19
1.1	Contextualização e Motivação.	19
1.2	Objetivos.	21
1.3	Organização do trabalho de conclusão de curso	22
2	REVISÃO BIBLIOGRÁFICA	23
2.1	Trabalhos correlacionados à análise de séries temporais para previsão na área da saúde pública.	23
2.2	Conclusões.	25
3	MÉTODOS, TÉCNICAS E CONCEITOS	27
3.1	Séries Temporais.	27
3.2	Modelos Probabilísticos.	27
3.3	Modelos Dinâmicos.	28
3.3.1	Long Short-Term Memory - LSTM.	29
3.4	Modelos Comparativos.	29
3.4.1	Média Móvel Simples - MMS.	29
3.4.2	Naive Forecasting.	30
3.4.3	Suavização Exponencial de Holt - SEH.	30
3.4.4	Suavização Exponencial Sazonal de Holt-Winters - HW.	30
3.4.5	Gradient Boosting.	30
4	DESENVOLVIMENTO	31
4.1	Sistema de Informações Hospitalares do SUS.	31
4.2	Coleta dos Dados.	32
4.2.1	Escopo dos Dados.	32
4.2.2	Do Dataset.	33
4.3	Análise da Série Temporal por Meio de Técnicas Descritivas.	33
4.3.1	Decomposição da Série.	34
4.3.2	Estacionariedade da Série.	34
4.3.3	Correlograma.	35
4.3.4	Conclusões.	36
4.4	Previsões Comparativas.	36
4.4.1	Média Móvel Simples e Naive Forecasting.	37
4.4.2	SEH e HW.	37

4.4.3	Gradient Boosting.	37
4.5	Previsões com modelo ARIMA Sazonal - SARIMA.	38
4.6	Previsões com modelo ARIMA Sazonal - SARIMA incluindo variável exógena	39
4.7	Previsões com LSTM.	41
5	CONCLUSÕES	47
	REFERÊNCIAS	49

1 INTRODUÇÃO

Neste capítulo será apresentada a motivação pela escolha do tema da pesquisa e uma introdução ao assunto.

1.1 Contextualização e Motivação.

O Sistema Único de Saúde - SUS é a denominação do sistema público de saúde no Brasil e foi instituído pela Constituição Federal de 1988. Com o surgimento do SUS, a população brasileira passou a ter direito à saúde universal e gratuita, financiada com recursos provenientes dos orçamentos da União, dos Estados, do Distrito Federal e dos Municípios. Apesar de a União ser o principal financiador da saúde pública no país, arcando com metade dos gastos, a outra metade fica por conta de estados e municípios, estabelecendo assim um modelo de financiamento tripartite.

A Lei 8.080, de 07/09/1990, conhecida como Lei Orgânica da Saúde, ainda dispõe que o conjunto de ações e serviços de saúde, prestados por órgãos e instituições públicas federais, estaduais e municipais, da Administração direta e indireta e das fundações mantidas pelo Poder Público, constitui o SUS. Fazem, então, parte do SUS os centros e postos de saúde, os hospitais públicos – incluindo os universitários, os laboratórios e hemocentros–, os serviços de Vigilância Sanitária, Vigilância Epidemiológica, Vigilância Ambiental, além de fundações e institutos de pesquisa acadêmica e científica. A Lei ainda dispõe que a iniciativa privada poderá participar do SUS em caráter complementar.

Conforme [CRUZ \(2016\)](#), nos últimos anos, o SUS tem passado por mudanças que interferem diretamente em seu funcionamento, citando as transições demográficas, nutricionais, tecnológicas e epidemiológicas. O artigo ainda ressalta o surgimento de “novas doenças” que, além de gerar cenário adverso para o financiamento da saúde pública, desperta o temor da população e direcionam a atenção da mídia e até mesmo de organismos internacionais para a administração da saúde pública brasileira. Conforme ressaltado, para dar conta das questões estruturantes que representam as maiores falhas do SUS, é preciso promover mudanças tanto no modelo de atenção à saúde, quanto nos modelos de gestão e de financiamento do sistema, sugerindo um movimento da gestão rumo à eficiência e à maior qualidade, apontando para a necessidade de otimizar a utilização dos recursos existentes.

Neste sentido, abordando-se a rede hospitalar, [CESCONETTO, LAPA e CALVO \(2008\)](#), com foco em estabelecer prioridades e políticas para controle dos gastos, informa que o custo médio das internações é cem vezes maior que o custo médio dos atendimentos ambulatoriais. Sendo assim, o impacto sobre estudos das redes hospitalares no controle e

melhor alocação de recursos públicos com a assistência à saúde deve ser maior que estudos similares efetuados com dados dos atendimentos ambulatoriais. O autor ainda informa que avaliação de serviços de saúde é, em geral, efetuada com o uso de indicadores parciais, citando, a exemplo, a taxa de ocupação de leitos, que permite uma avaliação imediata de algumas características do hospital.

NETO e MALIK (2007) reconhece que o hospital é uma das organizações mais complexas operadas pelo homem, portanto faz-se necessário um instrumento que permita tornar mais rapidamente disponíveis os dados a respeito do seu funcionamento, de maneira a estruturar a gestão de maneira sistemática. Neste sentido, CABRAL (2012) afirma que, dentre as instituições existentes, o hospital se destaca por sua complexidade e por ser vital para sobrevivência humana. Complementa, ainda, que esta complexidade decorre do fato de unir inúmeras empresas dentro de uma só (hotel, restaurante industrial, laboratório, centro cirúrgico, Unidade de Terapia Intensiva - UTI, serviço de imagem etc.) e funcionar ininterruptamente.

RAFFA, MALIK e PINOCHET (2016) afirma que o gerenciamento de leitos é estratégia, e que a gestão deve estar na base do planejamento estratégico do hospital, uma vez que exerce impacto financeiro nas organizações de saúde. Ressalta, ainda, que, com relação aos indicadores epidemiológicos, alguns hospitais observam somente as sazonalidades.

Segundo JUNIOR (2007), *“disponibilidade de sistemas adequados de informações epidemiológicas e administrativas é considerado o fator crítico de sucesso para o gerenciamento ético do cuidado em saúde”*, o que permite, dentre outros, a avaliação econômica, o que aponta para oportunidades de melhoria do processo e resultados.

Sobre os indicadores hospitalares, RAMOS et al. (2015) informa que a taxa de ocupação hospitalar varia de acordo com determinadas características da entidade de saúde, o que determina, inclusive, a performance no atendimento. Desta forma, percebe-se que a atividade de levantamento da taxa de ocupação, principalmente quando do direcionamento para número de casos acima do esperado de determinada enfermidade, não é uma tarefa trivial, todavia essencial para a otimização de recursos de todo o sistema de saúde.

Conforme mencionado por ANS (2012), há uma recomendação de uma taxa de ocupação entre 75 e 85%. Sendo que, a taxa de ocupação acima do preconizado está relacionada com um aumento de casos adversos, infecção hospitalar e diminuição da segurança do ambiente assistencial. Já a taxa de ocupação abaixo de 75% indica baixa utilização e ineficiência na gestão do hospital. MS (2002a) enfatiza que a taxa de ocupação esperada para cada tipo de leito depende do porte do hospital ou pool de leitos, mas nunca deve ser superior a 90%, pois números superiores a esse valor provocam taxas de recusa muito altas.

Sendo assim, o conhecimento sobre a taxa de ocupação hospitalar do SUS é fator

essencial para planejamento das ações e serviços de saúde do Governo Federal. A pandemia enfrentada pelo COVID-19 mostra que o Estado deve possuir plano de contingência de forma a otimizar a atual estrutura do sistema de saúde, realizando as devidas correções com base nas estimativas de propagação de uma doença, sejam surtos, pandemias, endemias ou epidemias. A exemplo, de acordo com [MS \(2020\)](#), em relação aos leitos de UTI, a taxa de ocupação média é de 78%, sendo necessária a otimização de recursos, principalmente quanto à utilização de leitos ociosos.

Portanto, de forma a ter controle sobre a taxa de ocupação hospitalar, a previsão da demanda é um importante fator de impacto no desempenho operacional dos hospitais. Percebe-se que oscilações de demanda não previstas causam sérios impactos na eficiência da gestão hospitalar, podendo gerar ociosidade de leitos ou taxas de recusas elevadas, sendo que, dos dois modos, geram prejuízos financeiro ao SUS e prejuízos à sociedade por um sistema de saúde ineficiente. Por esse motivo, a previsão de demanda de leitos de internação visa assistir o gestor hospitalar quanto ao planejamento de oferta dos serviços de saúde à população.

1.2 Objetivos.

A programação dos leitos hospitalares necessários para uma dada população é uma tarefa complexa, envolvendo tanto a análise da oferta desses serviços como a sua demanda. Tradicionalmente, há dois métodos utilizados para calcular o número de leitos hospitalares necessários para uma dada população em determinado período de tempo ([MS, 2002a](#)):

- a) Baseado em uma relação leito por habitante;
- b) Adotando uma fórmula/modelo em que a quantidade de leitos é função do número de internações esperadas, do tempo médio de permanência, considerando-se uma taxa de ocupação de 80-85%.

Ao contrário dos métodos tradicionais utilizados, este trabalho se propõe a elaborar um modelo baseado em séries temporais através da análise de dados reais das internações hospitalares, para fazer previsões da demanda futura do número de leitos.

Diante da vasta classificação de leitos ([MS, 2002b](#)), o trabalho focará apenas nos Leitos de Unidade de Tratamento Intensivo – UTI. Os leitos de UTI são destinados ao tratamento de paciente grave e de risco que exija assistência médica e de enfermagem ininterruptas, além de equipamentos e recursos humanos especializados, sendo reconhecidos por terem o custo mais elevado na gestão hospitalar.

Para tanto, serão utilizados dados relativos das Autorizações Hospitalares de Internação – AIH, coletadas mediante o Sistema de Informação Hospitalar do SUS – SIHSUS. Conforme menciona [CERQUEIRA et al. \(2019\)](#), a utilização da base de dados do SIHSUS é bastante frequente, seja em estudos de aplicações nas áreas de saúde

coletiva, quando na discussão da qualidade da informação disponibilizada, seja nas análises comparativas entre as bases de dados existentes.

Este trabalho reconhece que a regionalização na análise da saúde é essencial para melhores previsões envolvendo a saúde pública. Sendo assim, com relação ao espaço demográfico da análise, serão analisados os dados hospitalares referentes aos leitos de UTI localizados no município de Goiânia, capital do estado de Goiás, que concentra grande parte do atendimento hospitalar do estado. Neste sentido, o presente trabalho se concentrará apenas nos aspectos gerais da produção hospitalar, com foco principal na oferta de leitos de UTI por meio de informações extraídas do SIHSUS.

Para a previsão de leitos de UTI, conforme descrito adiante, serão utilizados modelos probabilísticos, como o modelo autorregressivo integrado e de média móvel – ARIMA, onde serão analisados aspectos de tendência, sazonalidade e estacionariedade da série, e também modelos dinâmicos.

1.3 Organização do trabalho de conclusão de curso

Esta monografia está organizada com uma breve pesquisa bibliográfica no Capítulo 2, a apresentação dos métodos, técnicas e tecnologias utilizadas no Capítulo 3, a descrição do desenvolvimento do trabalho com a exposição do problema, as atividades realizadas e os resultados obtidos no Capítulo 4 e as conclusões do trabalho no Capítulo 5.

2 REVISÃO BIBLIOGRÁFICA

Nesta seção, é apresentado o referencial teórico para o desenvolvimento de um modelo de series temporais, sendo discutido os modelos existentes e as técnicas para avaliar a qualidade das previsões.

2.1 Trabalhos correlacionados à análise de séries temporais para previsão na área da saúde pública.

A análise de séries temporais contempla a preocupação em derivar conhecimentos sobre a movimentação recente das medidas de interesse em saúde, prever resultados e reconhecer fatores que interferem sobre eles (ANTUNES; CARDOSO, 2015). Na programação de iniciativas de promoção da saúde, as variáveis são influenciadas por diversos fatores, dentro deles por condições socioeconômicas, saneamento básico, provisão e acesso a serviços de saúde.

ANTUNES e CARDOSO (2015) discorre sobre as variações sazonais e cíclicas que afetam a medida de muitas doenças, inclusive sobre a taxa de mortalidade estar ligada a uma maior amplitude meteorológica entre regiões. Em estudos epidemiológicos, o autor afirma que as previsões podem ser realizadas utilizando regressão linear, métodos ARIMA e através de reconhecimento dos padrões de variação da medida.

O trabalho de SOUZA, ADRIANO e LOPES (2015), que tinha por objetivo prever a taxa de ocupação em dois setores do Hospital Universitário de Santa Maria, utilizou uma amostra mensal de janeiro de 2000 a dezembro 2004, sendo a variável de interesse a taxa de ocupação dos leitos. O trabalho utilizou um modelo ARIMA(1,1,1) para o setor de Pronto Atendimento e ARIMA(1,1,2) para o setor Hospital Geral. Os autores concluem que a metodologia empregada foi capaz de retratar a realidade do hospital e fornecer subsídios para melhor funcionamento e atendimento aos pacientes.

No intuito de compreender o comportamento da taxa de ocupação hospitalar e identificar suas tendências e sazonalidades, CABRAL (2012) utilizou a abordagem metodologia *Box-Jenkins*. Todavia, o objetivo do trabalho não era realizar previsões futuras, mas auxiliar na gestão do hospital sobre a compreensão da taxa de ocupação.

Já GOLDWASSER et al. (2015) utilizou um modelo de séries temporais e um sistema de filas que permitiram realizar as previsões futuras da demanda de leitos de UTI no estado do Rio de Janeiro. Foram utilizados os modelos SARIMA, Holt-Winters e Tendência Amortecida Multiplicativa, porém o último apresentou melhor aderência aos dados. Os dados foram obtidos diretamente na central de regulação do estado e organizados cronologicamente. Com isso, para diferentes cenários, foi possível definir o número mínimo

de leitos, os tempos máximos de espera e as curvas de probabilidades. Além das estimativas, a análise permitiu observar a influência que alguns aspectos gerenciais podem exercer sobre a regulação de leitos.

Em [BERGS, HEERINCKX e VERELST \(2014\)](#), utilizou-se a análise de séries temporais para previsão do número de pacientes mensais no departamento de emergência de quatro hospitais belgas. Para tanto, foram coletados dados do número de pacientes mensais entre 2005 e 2011. Os autores utilizaram um modelo de suavização exponencial automatizada para prever os pacientes mensais durante o ano de 2011. O modelo foi capaz de prever as visitas mensais com um erro percentual absoluto médio variando de 2,64% a 4,8%, indicando acurácia na previsão.

Em [TANDBERG e QUAILS \(1993\)](#), um modelo de médias móveis simples foi utilizado para realizar previsões do volume de pacientes e do tempo de permanência sobre dados de dois anos. A análise de séries temporais forneceu previsões de curto prazo poderosas e precisas do volume futuro do departamento de emergência. Foram utilizados quatro modelos para realização das previsões, incluindo ARIMA, todavia, de acordo com as medidas de avaliação adotadas, os modelos mais simples tiveram melhor desempenho no estudo.

Já em [KADRI et al. \(2014\)](#), um modelo ARMA foi utilizado para a previsão de chegadas no departamento de emergência de um hospital na França. Os resultados indicaram que o desempenho de previsão dos modelos propostos é aceitável, podendo ajudar outros hospitais regionais e seus departamentos de emergência a realizarem suas próprias análises para auxiliar no planejamento. A série temporal foi essencialmente linear e, portanto, a modelagem ARMA ofereceu previsões robustas em muitos casos. Os autores informam que o modelo poderia ser melhorado utilizando métodos que incluem variáveis exógenas (ARMAX), medições meteorológicas ou eventos epidêmicos, por exemplo.

Já no âmbito da pandemia ocasionada pelo COVID-19, um importante estudo foi desenvolvido pela Harvard Global Health Institute ([HARVARD, 2020](#)), verificando a capacidade de cada um dos 306 centros hospitalares dos EUA (Hospital Referral Regions - HRR) em face ao crescente número de pacientes infectados. O modelo desenvolvido fornece a estimativa de leitos disponíveis bem como aqueles necessários para acomodação dos pacientes do COVID-19, auxiliando assim o planejamento hospitalar para enfrentamento da pandemia. O estudo foca em previsões regionalizadas, procurando assim evitar a transferência de pacientes entre hospitais. O estudo foi desenvolvido para acesso à população e apresenta a visualização das estimativas sobre os seguintes cenários: - a taxa de infecção da população (20%, 40% e 60%); - a cada taxa sobre os meses seguintes (6, 12 ou 18 meses). O estudo foi desenvolvido com base em dados empíricos obtidos da disseminação da cidade Wuhan na China e o dicionário de dados apresenta os métodos de cálculo da taxa de hospitalização e quantidade de dias de permanência.

2.2 Conclusões.

Com base na revisão bibliográfica apresentada, verifica-se que a elaboração de modelos para predição em saúde pública é constante e visam não apenas mitigar os riscos das enfermidades, mas também auxiliar no processo decisório na gestão da saúde, seja nas entidades privadas, seja nas entidades públicas.

Por fim, verificou-se que um modelo com vistas a fornecer estimativas de ocupação de leitos é essencial na ocorrência de picos epidêmicos, principalmente para o aproveitamento de recursos regionalizados, evitando assim a remoção de pacientes entre centros hospitalares.

3 MÉTODOS, TÉCNICAS E CONCEITOS

Nesta seção, é apresentada uma breve descrição dos métodos relacionados à previsão com séries temporais que serão utilizados para a implementação do modelo.

3.1 Séries Temporais.

Para este estudo, interessam as observações mensais dos quantitativos de leitos de UTI em utilização nas entidades de saúde localizadas no município de Goiânia, sendo que o objetivo da análise da série temporal gerada será fazer previsões de demanda de leitos de UTI, a princípio de curto prazo, para a série. Embora não seja possível prever exatamente os valores futuros, pode-se prever um comportamento aproximado para as observações.

De forma elementar, uma série temporal é qualquer conjunto de observações ordenadas no tempo (MORETTIN; TOLOI, 2018). Como as observações são tomadas mensalmente, teremos uma série temporal contínua. Conforme afirma EHLERS (2009), uma série temporal X_t é decomposta da seguinte forma:

$$X_t = T_t + C_t + R_t$$

onde T_t é a componente de tendência, C_t é a componente sazonal e R_t é o ruído.

3.2 Modelos Probabilísticos.

Conforme mencionado por MORETTIN e TOLOI (2018), uma metodologia bastante utilizada na análise de modelos paramétricos é conhecida como abordagem de *Box e Jenkins*, consistindo em ajustar modelos autorregressivos integrados de médias móveis, ARIMA(p, d, q), a um conjunto de dados. Em resumo, os componentes dos modelos ARIMA são assim representados:

- a) AR(p): Componentes autorregressivas, utilizam a relação de dependência entre a observação corrente e as observações em um período prévio;
- b) Integrado(d): Diferenças para tornar a série estacionária;
- c) MA(q): Componentes de médias móveis, utilizam a dependência entre uma observação e um erro residual de um modelo de média móvel aplicado a observações em atraso.

Na composição dos modelos ARIMA, quando a série é estacionária, a inclusão de termos autorregressivos e de médias móveis – ARMA(p,q) - pode ser a solução adequada.

Todavia, para tomar evidências da estacionariedade da série, ou de suas diferenças, será executado o teste de *Dickey-Fuller* (DICKEY; FULLER, 1979).

A estratégia para construção do modelo será baseada no ciclo iterativo proposto por MORETTIN e TOLOI (2018), no qual a escolha da estrutura do modelo é baseada nos próprios dados:

- a) uma classe geral de modelos é considerada para a análise, no caso modelos ARIMA (especificação);
- b) há identificação do modelo com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
- c) estimação dos parâmetros do modelo identificado;
- d) verificação ou diagnóstico do modelo ajustado, por meio de uma análise de resíduos, para saber se esse modelo é adequado para fazer a previsão.

Importante ressaltar que, no caso de observação de uma componente sazonal, há a necessidade de se considerar uma sazonalidade estocástica ajustada à série original a um modelo ARIMA com sazonalidade (SARIMA). Neste caso, os componentes de um modelo SARIMA $(p, d, q)x(P, D, Q)m$ são assim descritos:

- a) (p, d, q) : componentes não-sazonais;
- b) $(P, D, Q)m$: componentes sazonais, onde m é o período da sazonalidade.

Para identificação das componentes descritas, é importante a elaboração de correlogramas associados à série. Conforme afirma EHLERS (2009), a interpretação de correlogramas é um dos aspectos mais difíceis da análise de séries temporais. Para tanto, foram utilizados métodos automatizados do pacote PMDARIMA (2020) para a linguagem de programação Python. O pacote processa o método auto-ARIMA que busca identificar, para a série temporal, os parâmetros ideais para um modelo ARIMA/SARIMA. Indicando, através do Critério de Informação de Akaike (*Akaike Information Criterion* - AIC em inglês), os parâmetros que indicam o modelo com melhor qualidade, ou seja, o modelo ARIMA em que o critério é minimizado.

3.3 Modelos Dinâmicos.

O modelo ARIMA é o método estocástico mais conhecido para a realização de previsões em séries temporais. Conforme exposto, para sucesso do modelo, ainda que tenhamos pacotes computacionais que facilitam o trabalho, faz-se necessária a identificação de suas componentes: tendência, sazonalidade e estacionariedade.

Tradicionalmente, a previsão em séries temporais tem sido dominado por métodos lineares, assim como o modelo ARIMA, pois são eficazes para muitos problemas. Todavia,

ao assumir um relacionamento linear, exclui-se distribuições complexas. As técnicas de aprendizado de máquina e os algoritmos de aprendizado profundo tem introduzido novas abordagens onde as relações entre as variáveis são modeladas em uma hierarquia profunda e em camadas (SIAMI-NANINI; TAVAKALI; SIAMI NANIN,). Métodos de aprendizado profundo são capazes de identificar estruturas e padrões de dados, como não linearidade e complexidade, que podem ser utilizados na previsão em séries temporais.

Além disso, conforme afirma GÉRON (2019), quando é utilizada uma rede neural recorrente, aqueles componentes necessários nos modelos ARIMA, tais como tendência e sazonalidade, geralmente não precisam ser identificados, uma vez que o modelo não terá que aprendê-los.

3.3.1 Long Short-Term Memory - LSTM.

A LSTM é uma especialização de uma Rede Neural Recorrente - RNN que adiciona explicitamente a manipulação da ordem entre as observações quando está aprendendo a função de mapeamento da entrada das observações para a saída. Ou seja, é um tipo de rede neural que adiciona, nativamente, suporte para dados de entrada compostos por sequências de observações (BROWNLEE, 2018). Uma importante característica da LSTM é a sua capacidade em preservar e treinar características de dados por um longo período de tempo.

Para a construção e teste do modelo baseado em LSTM, foi utilizada a API *Keras* do pacote TensorFlow do Python (KERAS, 2020), que já possui nativamente a implementação daquele modelo.

3.4 Modelos Comparativos.

A princípio, para a realização das previsões na série temporal, foram selecionados os modelos ARIMA com sazonalidade (SARIMA) e o modelo LSTM. Todavia, para determinar a eficiência do modelo, o trabalho utilizou outros métodos e modelos, comparando as previsões geradas por estes modelos e os dados reais pela métrica da Raiz do Erro Quadrático Médio (*Root Mean Squared Error - RMSE* em inglês).

3.4.1 Média Móvel Simples - MMS.

A MMS consiste na média aritmética tomada a cada período constante. Sendo assim, a observação mais antiga é substituída pela observação mais recente, calculando uma nova média (MORETTIN; TOLOI, 2018). Para o presente trabalho, não foi realizada nenhuma previsão com base na MMS, a comparação foi realizada com a média móvel de toda a série tomada no período de teste.

3.4.2 Naive Forecasting.

No método *Naive Forecasting*, as previsões são simplesmente os valores das últimas observações (HYNDMAN; ATHANASOPOULOS, 2018).

3.4.3 Suavização Exponencial de Holt - SEH.

A SEH é um técnica de suavização recomendada para séries que apresentam tendência linear (MORETTIN; TOLOI, 2018). Num modelo de previsões, é o mais indicado quando o fator predominante na série temporal é a tendência, pois é o diferencial do modelo.

3.4.4 Suavização Exponencial Sazonal de Holt-Winters - HW.

A HW é utilizada em séries que apresentam um comportamento um pouco mais complexo, como sazonalidade. Envolve três equações com três parâmetros de suavização que são associados a cada componente da série: nível, tendência e sazonalidade (MORETTIN; TOLOI, 2018).

3.4.5 Gradient Boosting.

O *Gradient Boosting* é uma técnica de aprendizado de máquina, para problemas de regressão e classificação, que produz um modelo de previsão na forma de um *ensemble* de modelos de previsão fracos, geralmente árvores de decisão WIKIPÉDIA (2020b).

Para o presente trabalho, foi utilizada o pacote *XGBoost* do Python (XGBOOST, 2020), que é uma implementação do *Gradient Boosting*.

4 DESENVOLVIMENTO

Este capítulo descreve como foram realizados os experimentos relacionados a este trabalho. Para isso, apresenta-se inicialmente de onde foram obtidos os dados usados neste estudo, e como foram processados de maneira a obter-se os datasets que mais tarde foram utilizados. A seguir, descreve-se os experimentos realizados, e analisa-se os resultados obtidos.

O código fonte e os dados utilizados nos testes deste trabalho estão no endereço Web da referência [LEITE \(2020\)](#)

4.1 Sistema de Informações Hospitalares do SUS.

Neste estudo foram usados os dados das Autorização de Internação Hospitalar - AIH armazenados pelo Sistema de Informação Hospitalar do SUS - SIHSUS.

O SIHSUS teve início no ano de 1991, mediante a Portaria MS/GM n.º 896/90, e é um dos principais sistemas nacionais de informação em saúde ([CERQUEIRA et al., 2019](#)). A finalidade do SIHSUS é registrar todos os atendimentos provenientes de internações hospitalares que foram financiadas pelo SUS. É através do SIHSUS que o Ministério da Saúde recebe de forma mensal os dados de todas as internações autorizadas (aprovadas ou não para pagamento) para que possam ser repassados às Secretarias de Saúde Municipais e Estaduais os valores de produção de Média e Alta Complexidade - MAC, além dos valores de CNRAC, FAEC e de Hospitais Universitários – em suas variadas formas de contrato de gestão ([DATASUS, 2020c](#)).

A AIH é o documento para identificar o paciente e os serviços prestados sob o regime de internação hospitalar e fornecer informações para o gerenciamento do SIHSUS. Conforme consta em [MS \(2007\)](#), previamente à emissão da AIH, faz-se necessária emissão de laudo por agente de saúde (médico/enfermeiro) com dados "*de identificação do paciente, da anamnese, exame físico, resultados de exames complementares, justificativa da solicitação; identificação do profissional solicitante e autorizador; diagnóstico inicial, códigos de procedimentos de acordo com a Tabela do SUS e CID*". Por fim, a AIH somente é gerada quando ocorre uma internação em um prestador público, privado ou conveniado ao SUS e é enviada ao gestor da Unidade Prestadora de Serviços (Secretarias Estaduais ou Municipais de Saúde). Mensalmente, os gestores enviam ao Ministério da Saúde os dados de todas as internações ocorridas.

O SIHSUS é um relevante instrumento de informação para orientar o gestor na tomada de decisões relacionadas ao planejamento das ações de saúde ([MS, 2007](#)). Conforme afirma [CERQUEIRA et al. \(2019\)](#), os dados da produção do SUS são amplamente utilizadas

na literatura para vários propósitos, "*entre os quais avaliação do sistema de saúde, seus gastos, oferta e demanda de serviços, cobertura, e também para a geração e análise de indicadores.*"

A base do SIHSUS é interessante do ponto de vista em que trata de indicadores e taxas relativas às internações e às comparações geográficas. Apresenta problemas de qualidade quando comparada com outras bases, todavia as informações contidas no SIHSUS serão de qualidade para o presente trabalho (CERQUEIRA et al., 2019).

Apesar de fazer ressalvas sobre o uso da base do SIHSUS, RAMOS et al. (2015) afirma que a abrangência do banco de dados é adequada para a avaliação do desempenho hospitalar e para nortear o processo decisório em saúde.

4.2 Coleta dos Dados.

Conforme já mencionado, no momento em que um paciente ingressa em uma unidade hospitalar para internação, o agente de saúde preenche o laudo para a emissão da AIH. A AIH é formada por uma sequência de treze dígitos correspondente ao número de autorização, que se refere à internação, e não ao paciente (CERQUEIRA et al., 2019). Estes dados são então enviados ao SUS onde são processados, integrados e, posteriormente, liberados para consulta pública de forma tabular, onde há a disseminação de informações sobre a saúde pública.

As informações relacionadas ao paciente, ao motivo principal da internação hospitalar e ao local e período de internação são consolidados na base AIH Reduzida - RD (CERQUEIRA et al., 2019). A RD contém as AIH aprovadas e também os valores efetivamente pagos por mês de competência. Ela inclui os procedimentos processados e validados pelo Ministério da Saúde entre os apresentados por todos os estabelecimentos prestadores de serviços para o SUS. Já o detalhamento das AIH, ou seja, as informações dos serviços profissionais e procedimentos hospitalares realizados no decorrer da internação hospitalar, são registrados na base de Serviços Profissionais - SP.

Por fim, as informações das bases RD e SP são disponibilizadas para consulta e extração no Sistema de Informações de Saúde do Datasus - TABNET (DATASUS, 2020a).

4.2.1 Escopo dos Dados.

Cada serviço profissional ou procedimento hospitalar realizado no decorrer da internação hospitalar é relacionado a um procedimento contido no Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos e OPM do SUS - Sigtap (DATASUS, 2020b), inclusive, a quantidade de diárias nos leitos de internação hospitalar.

O escopo dos dados da série temporal deste trabalho foi definido para apenas as informações do município de Goiânia e a ocupação dos leitos de UTI Adulto de janeiro/2009

até dezembro/2019. Sendo assim, foi necessário, então, relacionar os procedimentos contidos no Sigtap que se referem às diárias de ocupação daquele tipo de leito. O *dataset* final que terá as observações referentes à série temporal mensal da ocupação dos leitos hospitalares será composta apenas dos leitos de UTI Adulto Tipo I, II e III. Leitos estes destinados à assistência de pacientes com idade superior a 15 anos de idade (EBSERH, 2016).

4.2.2 Do Dataset.

O *dataset* foi extraído em 20/11/2020 das informações de Assistência à Saúde do TABNET e foram aplicados os seguintes filtros:

1. Dados Detalhados de AIH (SP), por local de internação, 2008 em diante;
2. Linha: Procedimento;
3. Coluna: Ano/mês atendimento;
4. Abrangência Geográfica: Goiás;
5. Período: de janeiro/2008 até agosto/2020;
6. Município: 520870 - Goiânia;
7. Procedimento: - *0802010105 DIARIA DE UNIDADE DE TERAPIA INTENSIVA DE ADULTO (UTI I)*; - *0802010083 DIARIA DE UNIDADE DE TERAPIA INTENSIVA ADULTO (UTI II)*; - *0802010091 DIARIA DE UNIDADE DE TERAPIA INTENSIVA ADULTO (UTI III)*.

O resultado foi o total de diárias de UTI por procedimento durante o período de janeiro/2008 até agosto/2020 totalizado por mês. Os dados foram tratados de forma a totalizar por mês o quantitativo total de diárias de UTI (I, II e III) para o período em análise (janeiro/2009 - dezembro/2019). No final, o resultado foi a série expressa na Figura 1.

4.3 Análise da Série Temporal por Meio de Técnicas Descritivas.

Por meio de técnicas descritivas, já podem ser identificados componentes do modelo que descreverão propriedades da série, como padrão de tendência, existência de variação sazonal, observações discrepantes, alterações estruturais etc. (EHLERS, 2009). Para tanto, foram realizadas diversas representações gráficas dos dados, assim como a utilização de pacotes da linguagem Python que descrevem o comportamento da série.

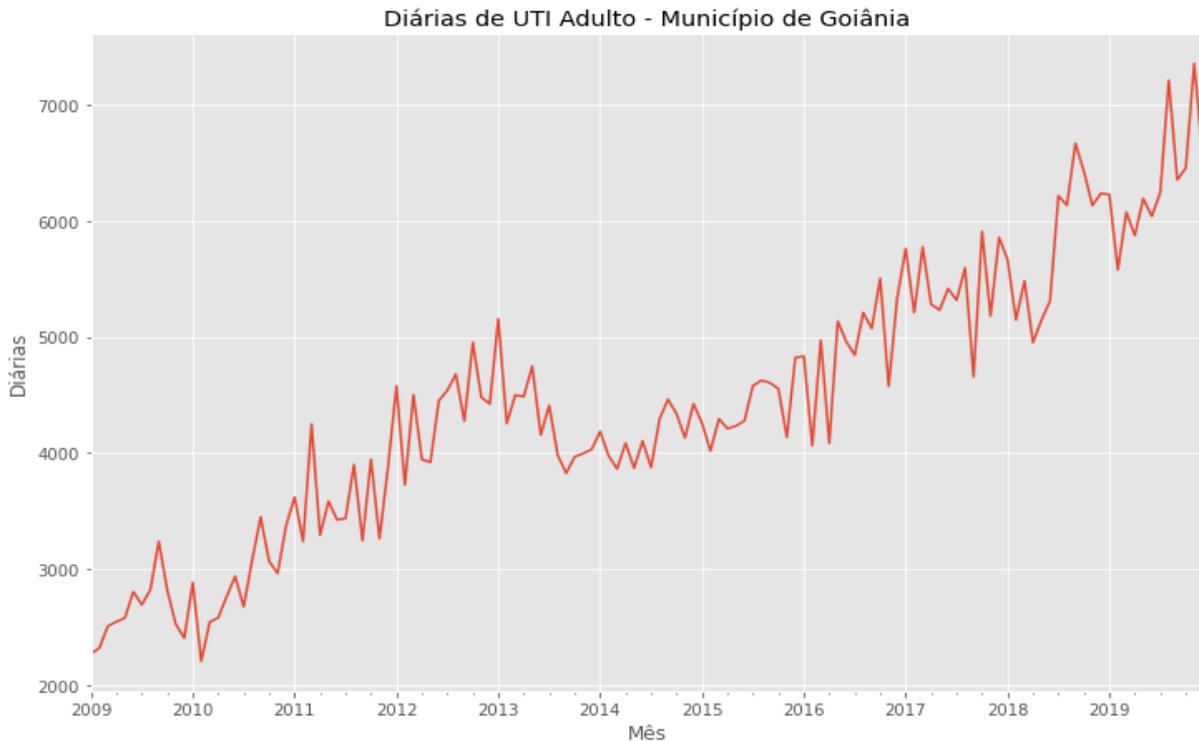


Figura 1: Série de diárias de UTI Adulto Tipo I, II e III no município de Goiânia com observações mensais de janeiro de 2009 até dezembro de 2019.

4.3.1 Decomposição da Série.

Conforme já foi mencionado, uma série pode ser decomposta em tendência, sazonalidade e ruído. A função `seasonal_decompose` do pacote *Statsmodels* da linguagem Python ([STATSMODELS, 2020](#)), de forma simples, decompõe a série nestas três componentes. Para tanto, devem ser definidos o tipo e o período de sazonalidade. A Figura 2 apresenta esta decomposição.

A análise gráfica da decomposição da série já proporciona algumas suposições. Aparentemente há um comportamento aleatório no ruído, levando a presunção de ausência de normalidade. Outro ponto é que já se pode observar uma aparente tendência positiva polinomial da série. Isso tudo posto sobre uma suposição de 12 meses de um padrão de sazonalidade.

4.3.2 Estacionariedade da Série.

Os modelos ARIMA podem ser aplicados em casos em que os dados mostram evidências de não estacionariedade da série. Para tanto, necessita-se determinar o passo de diferenciação, o que corresponde à parte "integrada" do modelo, para se eliminar a não estacionariedade. Conforme menciona [EHLERS \(2009\)](#), independente do seu uso para induzir estacionariedade, a diferenciação pode ser muito útil como ferramenta exploratória.

Para verificar a estacionariedade da série, foi aplicado o teste de *Dickey-Fuller*

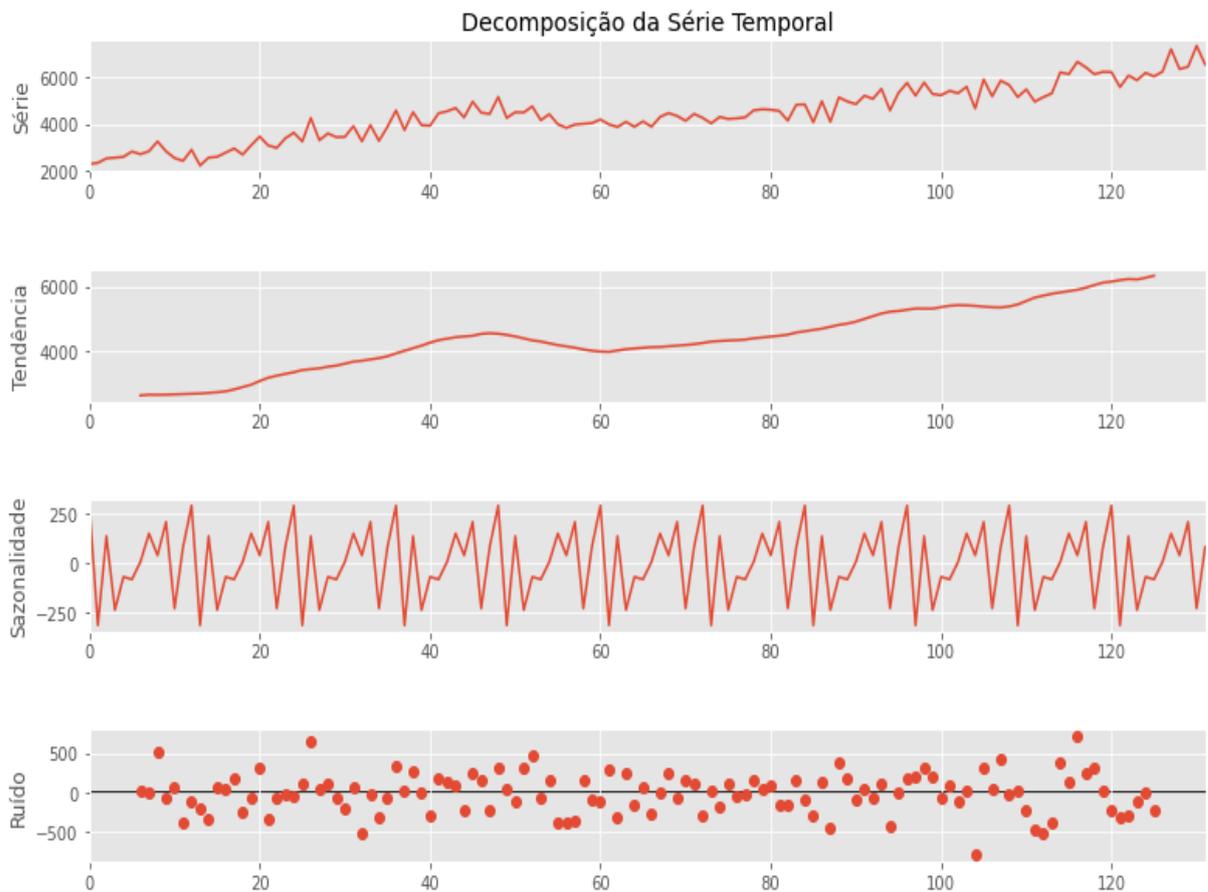


Figura 2: Decomposição Série Temporal nas componentes de tendência, sazonalidade e ruído.

(função *"adfuller"* do pacote *Statmodels*), verificando que, para a primeira diferença, há fortes evidências de estacionariedade da série. O que indica já haver pelo menos uma componente "integrada".

4.3.3 Correlograma.

O correlograma é um gráfico com os k primeiros coeficientes da autocorrelação com função de k . Segundo [EHLERS \(2009\)](#), apesar de não ser uma tarefa simples, o correlograma pode ser uma ferramenta poderosa para identificar características da série temporal. A Figura 3 apresenta os correlogramas das autocorrelações e autocorrelações parciais da série tomadas sobre os primeiros 65 coeficientes.

A primeira interpretação que podemos obter do correlograma da autocorrelação é sobre a aleatoriedade da série temporal. A parte sombreada refere-se ao intervalo de confiança de 95%, sugerindo que os valores de correlação fora desta área são muito provavelmente uma correlação e não um acaso estatístico. Observa-se, portanto, que não se trata de uma série completamente aleatória.

Pela função de autocorrelação, verifica-se que há um decaimento exponencial,

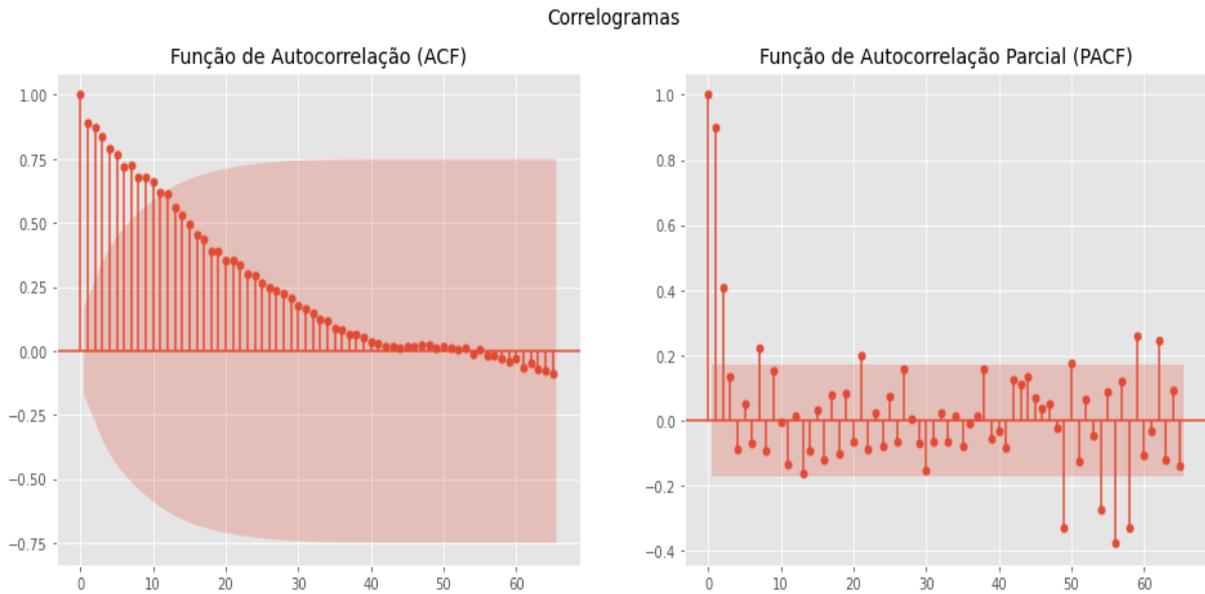


Figura 3: Correlograma das Função de Autocorrelação e de Autocorrelação Parcial da Série.

indicando um processo AR(1). Por outro lado, a função de autocorrelação parcial possui decaimento para com $k \geq 2$, reforçando o processo AR(1).

4.3.4 Conclusões.

Apesar de as técnicas descritivas serem importantes para a compreensão da série temporal, no contexto deste trabalho foi utilizado o método auto-ARIMA do pacote [PMDARIMA \(2020\)](#). No auto-ARIMA, a função automaticamente descobrirá os valores das componentes sazonais (p, d, q) e não-sazonais (P, D e Q) ótimos que são adequados para o conjunto de dados fornecer boas previsões, mediante execução gradual de hiperparâmetros e comparativo através da métrica AIC.

4.4 Previsões Comparativas.

Conforme já mencionado na Seção 3.4, foram produzidas cinco previsões com modelos simples para estabelecer um patamar mínimo de medidas de desempenho para os modelos SARIMA e LSTM, as quais foram comparadas mediante o RMSE.

Para o teste do modelo, tendo em vista que os dados referem-se às diárias de leitos de UTI do município de Goiânia no período de janeiro de 2009 até dezembro de 2019, os modelos foram treinados com dados até dezembro de 2018 e os dados referentes ao ano de 2019 serão utilizados para teste de eficiência.

4.4.1 Média Móvel Simples e Naive Forecasting.

Conforme já mencionado nas Seções 3.4.1 e 3.4.2, foram calculados os valores da média móvel simples, para todo o período da série, considerando uma janela de 12 meses, e, também, para o *Naive Forecasting*, conforme apresentado na Figura 4.

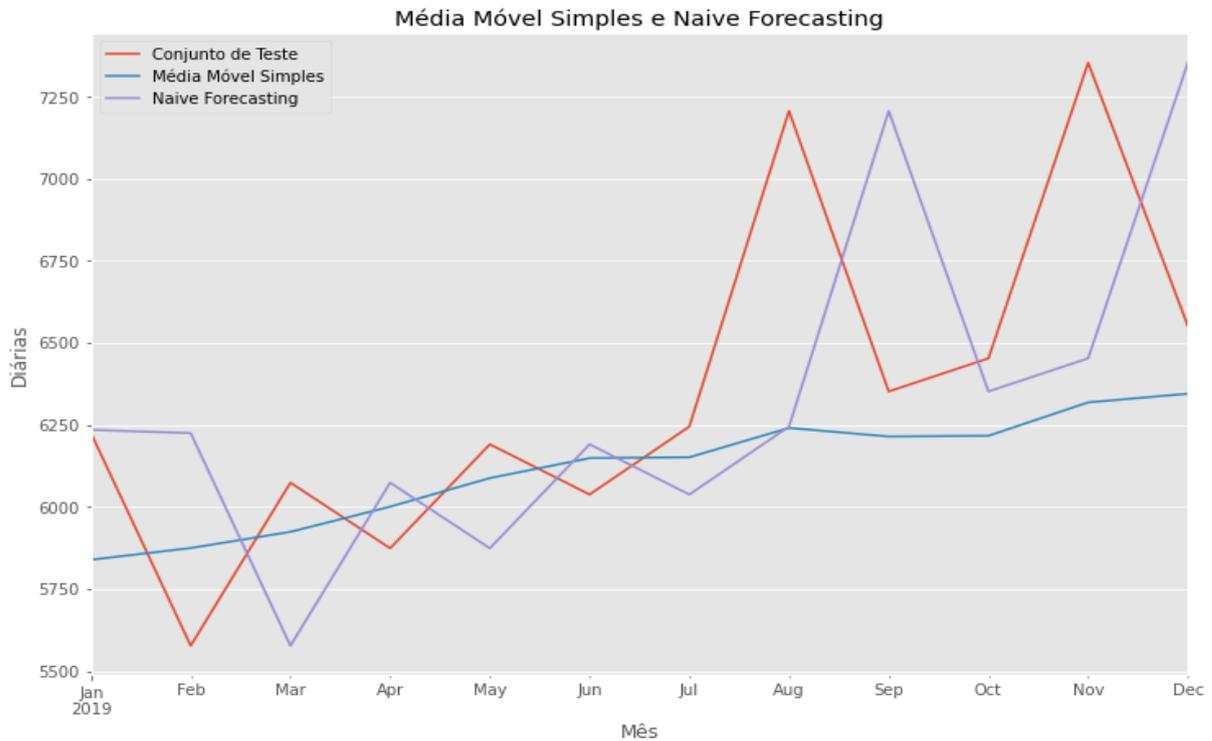


Figura 4: Média Móvel Simples com janela de 12 meses e Naive Forecasting comparados com os dados observados no ano de 2019.

4.4.2 SEH e HW.

Já para a Suavização Exponencial de Holt e Sazonal de Holt-Winters (Seções 3.4.3 e 3.4.4, foram preditos os valores para o ano de 2019, um comparativo entre os valores de teste e a predição está no gráfico da Figura 5.

4.4.3 Gradient Boosting.

O último modelo comparativo foi realizado utilizando o pacote [XGBOOST \(2020\)](#) da linguagem de programação Python, que é uma implementação do *Gradient Boosting* - Seção 3.4.5.

Conforme menciona [BROWNLEE \(2020b\)](#), o *XGBoost* é uma implementação eficiente do *Gradient Boosting* para problemas de classificação e regressão. Afirma, ainda, que o XGBoost pode ser usado para previsão com séries temporais, todavia exige que os dados sejam transformados em um problema de aprendizado supervisionado. Ressalta-se que esta mesma abordagem é utilizada também no modelo LSTM. Por fim, sugere que

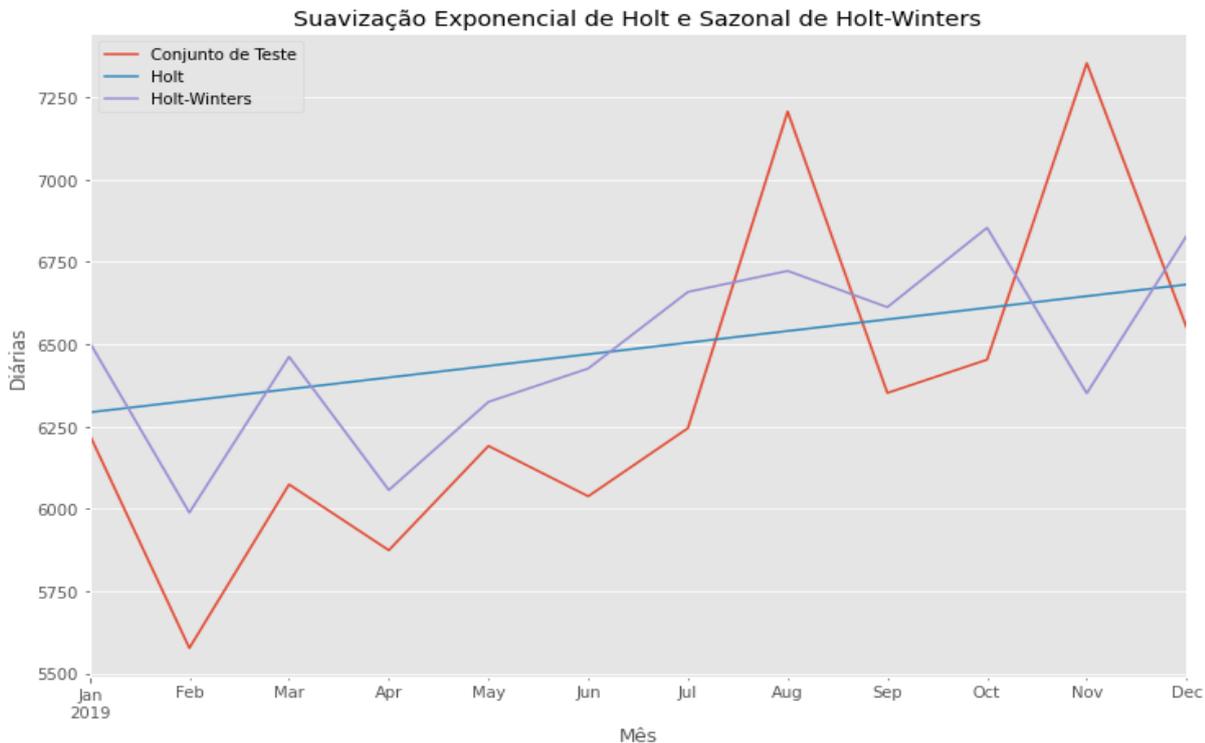


Figura 5: Previsão de diárias de UTI no município de Goiânia para o ano de 2019 como os modelos de Suavização Exponencial de Holt e Sazonal de Holt-Winters comparados com os dados observados no ano de 2019.

a validação siga o método *walk-forward validation*, o que foi realizado para selecionar o melhor tamanho da janela deslissante, que no caso foram conjuntos de cinco observações da série temporal (características) e a observação seguinte como rótulo. A Figura 6 apresenta o comparativo entre os dados de teste e as previsões geradas pelo modelo.

4.5 Previsões com modelo ARIMA Sazonal - SARIMA.

Conforme já mencionado na Seção 3.2, foram utilizados métodos automatizados do pacote [PMDARIMA \(2020\)](#) que processam o método auto-ARIMA buscando identificar, para a série temporal, os parâmetros ideais para um modelo ARIMA/SARIMA.

Em resumo, o método auto-ARIMA é inicializado informando o conjunto de dados de treinamento, os valores iniciais e máximos das componentes sazonais (p , d , q) e não-sazonais (P , D e Q), sobre os quais o algoritmo realizará as validações, e o período de sazonalidade m .

Informando o parâmetro "*random*" como verdadeiro, similarmente ao método de validação *Grid Search* ([WIKIPÉDIA, 2020c](#)), o auto-ARIMA realiza uma escolha aleatória sobre o espaço dos hiperparâmetros. Ao final da execução do método, é indicado o melhor modelo (melhor AIC) que se ajusta aos dados.

A Figura 7 exibe a execução do método auto-ARIMA sobre os dados de treina-

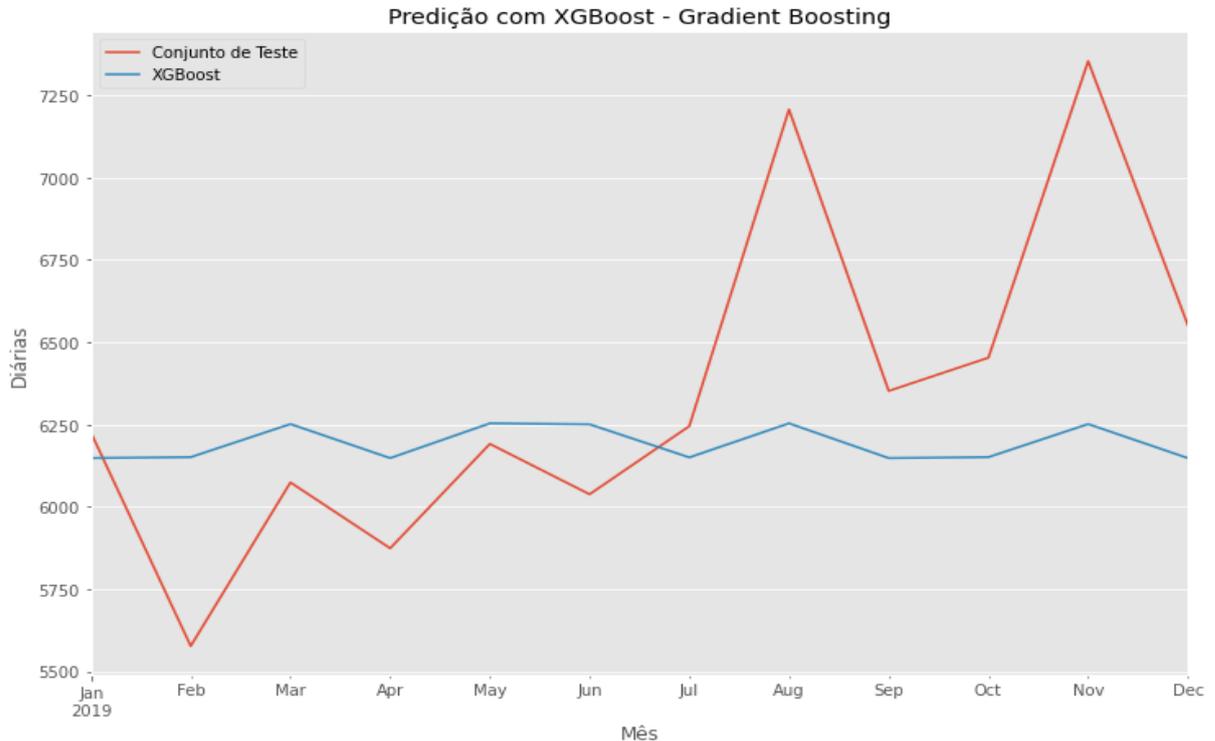


Figura 6: Previsão de diárias de UTI no município de Goiânia para o ano de 2019 com o modelo *XGBoost* (*Gradient Boosting*) comparados com os dados observados no ano de 2019.

mento. Como pode ser observado, o modelo que melhor se ajustou aos dados apresentou componentes sazonais na ordem $(1,1,0)$ e componentes não-sazonais na ordem $(3,1,2)$, com sazonalidade 12.

Posteriormente, foi realizada a análise de resíduos do modelo gerado, e, pelo método de *Shapiro-Wilk* verificou-se que havia fortes indícios de que os mesmos se distribuíam de forma normal, o que, a princípio, apontava para um bom modelo de previsão. Entretanto, observando o gráfico da Figura 8, percebe-se que os valores previstos não conseguem acompanhar os vales e picos das observações originais.

Comparando o RMSE do modelo SARIMA acima em relação ao RMSE dos modelos comparativos (Seção 4.4), verifica-se que a performance das previsões superou apenas o *Naive Forecasting* (Figura 9). Neste caso, observa-se que, considerando apenas o componente tendência expresso pela Suavização Exponencial de Holt, teríamos uma performance melhor.

4.6 Previsões com modelo ARIMA Sazonal - SARIMA incluindo variável exógena

Foge ao escopo deste trabalho a análise das ocorrências de eventos que sejam fontes de ruídos à série temporal estudada. Como a variável do modelo refere-se aos leitos de UTI, entende-se que existam várias intervenções (MORETTIN; TOLOI, 2018) que poderiam

```

Performing stepwise search to minimize aic
ARIMA(0,1,0)(0,1,0)[12] : AIC=1637.854, Time=0.02 sec
ARIMA(1,1,0)(1,1,0)[12] : AIC=1588.814, Time=0.12 sec
ARIMA(0,1,1)(0,1,1)[12] : AIC=inf, Time=0.17 sec
ARIMA(1,1,0)(0,1,0)[12] : AIC=1601.803, Time=0.05 sec
ARIMA(1,1,0)(2,1,0)[12] : AIC=1580.612, Time=0.28 sec
ARIMA(1,1,0)(3,1,0)[12] : AIC=1567.792, Time=0.62 sec
ARIMA(1,1,0)(4,1,0)[12] : AIC=1561.238, Time=1.13 sec
ARIMA(1,1,0)(5,1,0)[12] : AIC=inf, Time=2.40 sec
ARIMA(1,1,0)(4,1,1)[12] : AIC=1558.677, Time=1.98 sec
ARIMA(1,1,0)(3,1,1)[12] : AIC=1558.576, Time=1.56 sec
ARIMA(1,1,0)(2,1,1)[12] : AIC=inf, Time=1.35 sec
ARIMA(1,1,0)(3,1,2)[12] : AIC=1558.145, Time=2.43 sec
ARIMA(1,1,0)(2,1,2)[12] : AIC=inf, Time=1.59 sec
ARIMA(1,1,0)(4,1,2)[12] : AIC=inf, Time=5.00 sec
ARIMA(1,1,0)(3,1,3)[12] : AIC=inf, Time=4.35 sec
ARIMA(1,1,0)(2,1,3)[12] : AIC=1562.061, Time=3.90 sec
ARIMA(1,1,0)(4,1,3)[12] : AIC=inf, Time=6.08 sec
ARIMA(0,1,0)(3,1,2)[12] : AIC=inf, Time=3.19 sec
ARIMA(2,1,0)(3,1,2)[12] : AIC=inf, Time=4.82 sec
ARIMA(1,1,1)(3,1,2)[12] : AIC=inf, Time=4.37 sec
ARIMA(0,1,1)(3,1,2)[12] : AIC=inf, Time=3.17 sec
ARIMA(2,1,1)(3,1,2)[12] : AIC=inf, Time=4.54 sec
ARIMA(1,1,0)(3,1,2)[12] intercept : AIC=1561.136, Time=2.79 sec

Best model: ARIMA(1,1,0)(3,1,2)[12]
Total fit time: 55.934 seconds

```

Figura 7: Execução do passo-a-passo no [PMDARIMA \(2020\)](#) com vistas a minimizar o AIC para um modelo do tipo ARIMA com sazonalidade - SARIMA.

gerar ruídos no modelo, prejudicando, portanto, as previsões geradas.

Todavia, tendo em vista que as previsões geradas pelo modelo SEH, conforme RMSE da Figura 9, apresentou a melhor performance, procurou-se uma variável exógena ([WIKIPÉDIA, 2020e](#)) que determinasse o componente de tendência da série. Sendo assim, conforme mencionado por [MS \(2002a\)](#), que uma das formas de se estimar a quantidade de leitos é mediante a população local, obteve-se em [IBGE \(2020\)](#) as estimativas anuais de população do município de Goiânia dos anos de 2009 até 2019.

O gráfico da Figura 10 apresenta o comparativo, em dados escalados, entre as diárias de UTI do município de Goiânia e a população estimada. Observa-se, visualmente, que ambas as séries possuem similaridade na tendência. Além disso, os dados de ambas as variáveis possuem um coeficiente de correlação de *Pearson* de 0.9, ou seja, fortemente correlacionadas.

Novamente, foi executado o método auto-ARIMA do pacote [PMDARIMA \(2020\)](#) com os mesmos parâmetros e dados de treinamento da Seção 4.5, incluindo-se o parâmetro X , referente à variável exógena, ou seja, as estimativas populacionais de Goiânia, tomadas mensalmente, através de uma interpolação linear, de janeiro de 2009 até dezembro de 2018.

A Figura 11 exibe a execução do método auto-ARIMA sobre os dados de treinamento. Como pode ser observado, o modelo que melhor se ajustou aos dados apresentou

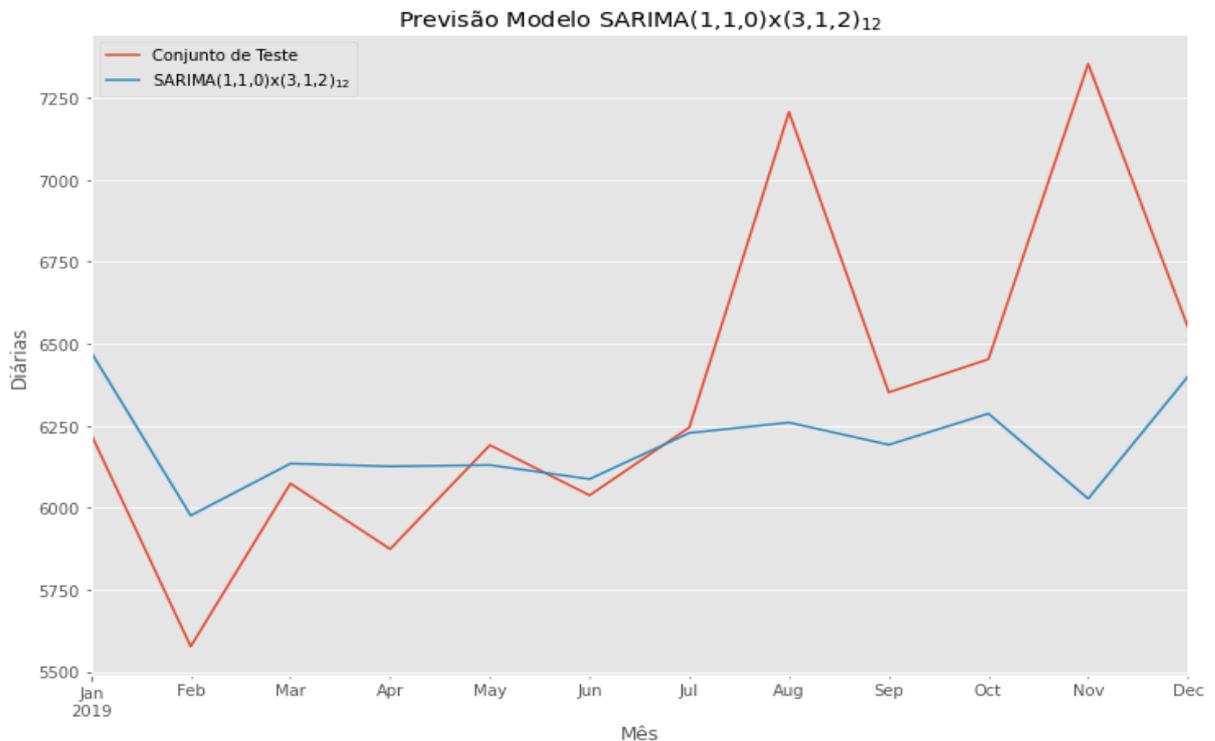


Figura 8: Previsão de diárias de UTI no município de Goiânia para o ano de 2019 com Modelo SARIMA(1,1,0)x(3,1,2)₁₂ comparados com os dados observados no ano de 2019.

componentes sazonais na ordem (0,1,1) e componentes não-sazonais na ordem (3,1,0), com sazonalidade 12.

No gráfico da Figura 12, pode-se observar que as previsões geradas pelo modelo com variável exógena, a princípio, tende a ter uma performance melhor se comparado com o modelo sem a variável exógena (Figura 8). Todavia, na análise dos resíduos do modelo, pelo método de *Shapiro-Wilk* verificou-se que há indícios de que os mesmos não se distribuíam de forma normal, apontando para um modelo não apropriado. Ainda assim, o modelo apresentou o RMSE de 410,01, ou seja, menor que os demais modelos apresentados.

4.7 Previsões com LSTM.

Conforme já mencionado, um importante atributo da LSTM é a sua capacidade em preservar e treinar características de dados por um longo período de tempo. Além disso, para aplicação de tal modelo, os componentes necessários nos modelos ARIMA, tais como tendência e sazonalidade, geralmente não precisam ser identificados, uma vez que o modelo não terá que aprendê-los. Sendo assim, o objetivo com este modelo é, a partir de única variável, desconsiderando possíveis intervenções, verificar se o modelo é capaz de produzir melhores previsões que os modelos mencionados nas seções anteriores.

	RMSE
Modelo	
SEH	435.78
HW	439.44
MMS-12	450.04
XGB	493.37
SARIMA(1,1,0)x(3,1,2)₁₂	502.36
Naive Forecasting	577.02

Figura 9: Raiz do Erro Quadrático Médio entre as previsões geradas pelos modelos SEH, HW, MMS-12, XGB, SARIMA(1,1,0)x(3,1,2)₁₂ e Naive Forecasting e os dados de teste (observações do ano de 2019)

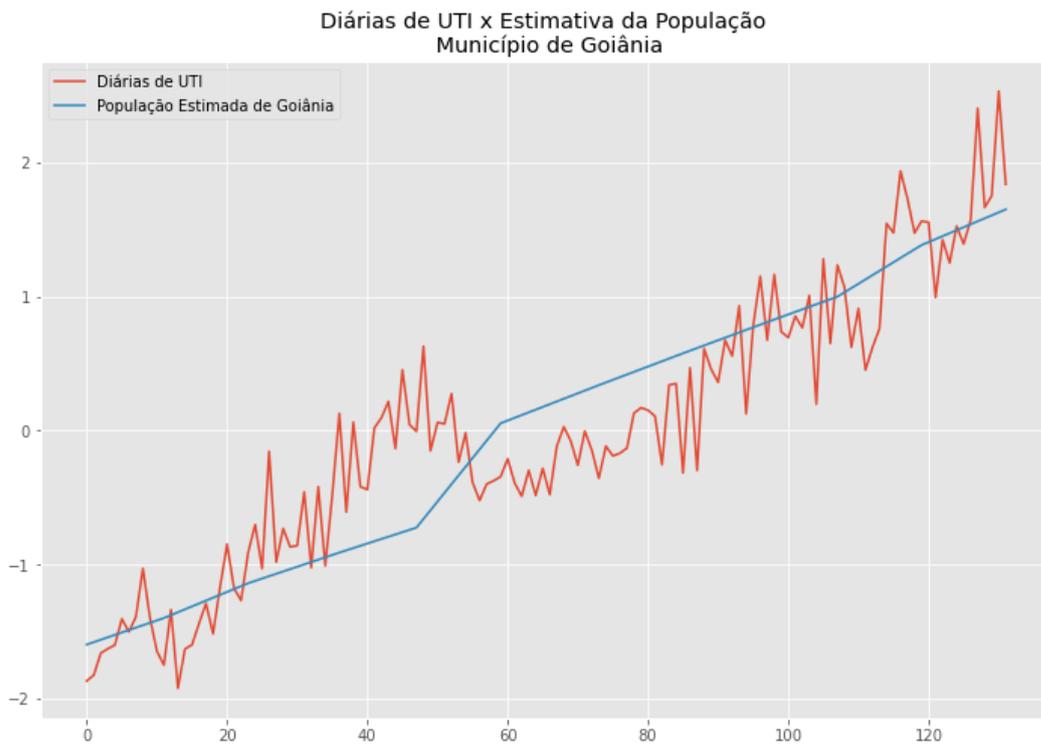


Figura 10: Comparativo, com dados escalados, entre a quantidade de Diárias de UTI e a População Estimada de Goiânia dos anos de 2009 até 2019.

Da mesma forma que o XGBoost, para o modelo LSTM, os dados devem ser transformados em um problema de aprendizado supervisionado. O pacote [KERAS \(2020\)](#) facilita tal transformação com a classe *TimeseriesGenerator*, que gera lotes de dados temporais. Para o presente caso, foram gerados lotes sequenciais de tamanho doze, ou seja, as características são as diárias de UTI referentes a doze meses e os rótulos serão as diárias do mês subsequente.

```

Performing stepwise search to minimize aic
ARIMA(0,1,0)(0,1,0)[12] : AIC=1641.401, Time=0.14 sec
ARIMA(1,1,0)(1,1,0)[12] : AIC=1591.528, Time=0.42 sec
ARIMA(0,1,1)(0,1,1)[12] : AIC=inf, Time=0.69 sec
ARIMA(1,1,0)(0,1,0)[12] : AIC=1604.042, Time=0.10 sec
ARIMA(1,1,0)(2,1,0)[12] : AIC=1586.173, Time=0.92 sec
ARIMA(1,1,0)(3,1,0)[12] : AIC=1576.481, Time=2.84 sec
ARIMA(1,1,0)(4,1,0)[12] : AIC=1577.217, Time=4.80 sec
ARIMA(1,1,0)(3,1,1)[12] : AIC=1576.572, Time=3.88 sec
ARIMA(1,1,0)(2,1,1)[12] : AIC=1576.934, Time=2.21 sec
ARIMA(1,1,0)(4,1,1)[12] : AIC=1578.022, Time=4.72 sec
ARIMA(0,1,0)(3,1,0)[12] : AIC=1610.609, Time=1.13 sec
ARIMA(2,1,0)(3,1,0)[12] : AIC=1579.750, Time=1.47 sec
ARIMA(1,1,1)(3,1,0)[12] : AIC=1580.322, Time=2.84 sec
ARIMA(0,1,1)(3,1,0)[12] : AIC=1575.837, Time=2.64 sec
ARIMA(0,1,1)(2,1,0)[12] : AIC=1587.357, Time=1.56 sec
ARIMA(0,1,1)(4,1,0)[12] : AIC=1583.241, Time=4.03 sec
ARIMA(0,1,1)(3,1,1)[12] : AIC=1583.089, Time=2.94 sec
ARIMA(0,1,1)(2,1,1)[12] : AIC=1583.183, Time=1.37 sec
ARIMA(0,1,1)(4,1,1)[12] : AIC=1584.264, Time=4.79 sec
ARIMA(0,1,2)(3,1,0)[12] : AIC=1583.673, Time=2.84 sec
ARIMA(1,1,2)(3,1,0)[12] : AIC=1583.376, Time=2.87 sec
ARIMA(0,1,1)(3,1,0)[12] intercept : AIC=inf, Time=nan sec

Best model: ARIMA(0,1,1)(3,1,0)[12]
Total fit time: 49.235 seconds

```

Figura 11: Execução do passo-a-passo no [PMDARIMA \(2020\)](#) com vistas a minimizar o AIC para um modelo do tipo ARIMA com sazonalidade - SARIMA e inclusão de variável exógena.

Para as previsões com LSTM, foram treinados cinco modelos descritos a seguir:

1. Modelo com uma camada LSTM com 100 unidades (dimensionalidade do espaço de saída), treinada por 100 épocas;
2. Modelo com uma camada LSTM com 100 unidades e uma camada de *Dropout* ([WIKIPÉDIA, 2020a](#)) treinada por 100 épocas;
3. Modelo com uma camada LSTM com 100 unidades e uma camada de *Dropout* treinada por 100 épocas e taxa de aprendizado com decaimento a partir da época 20, exponencial a -0,2;
4. Modelo com uma camada LSTM com 100 unidades, treinada por 1000 épocas;
5. Modelo com uma camada LSTM com 100 unidades, treinada por 100 épocas, com taxa de aprendizado de 0,0001.

Para todos os modelos, foi utilizada a função de ativação ReLU ([WIKIPÉDIA, 2020d](#)) na camada LSTM, e a rede foi treinada utilizando o algoritmo de otimização Adam ([BROWNLEE, 2020a](#)) com taxa de aprendizado padrão de 0,001, exceto no modelo 5, e com a perda computada pelo Erro Quadrático Médio.

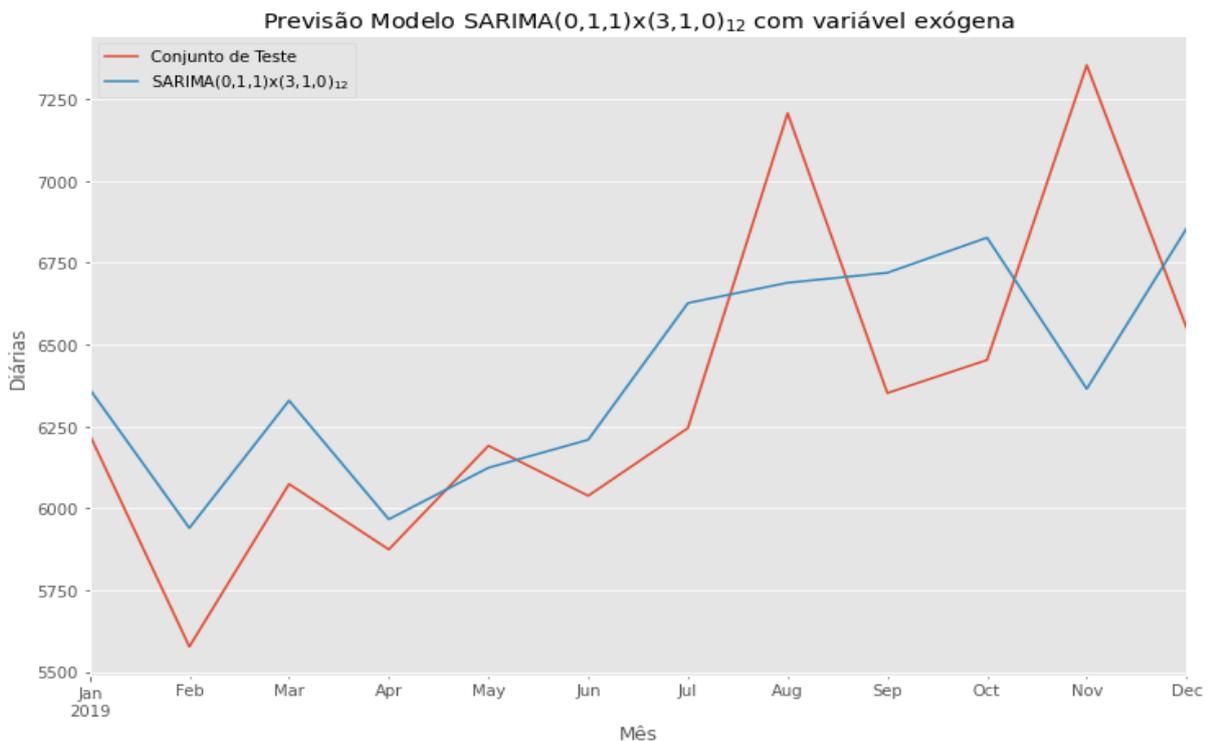


Figura 12: Previsão de diárias de UTI no município de Goiânia para o ano de 2019 com Modelo SARIMA(0,1,1)x(3,1,0)₁₂, considerando a intervenção da população do município, comparados com os dados observados no ano de 2019.

A Figura 13 apresenta o RMSE de cada um dos cinco modelos treinados com a LSTM. Já a Figura 14 apresenta o gráfico comparando os dados observados no ano de 2019 e as previsões geradas pelo Modelo 1.

RMSE	
LSTM	
Modelo 1	355.427605
Modelo 5	431.287381
Modelo 3	440.003188
Modelo 4	523.334248
Modelo 2	607.945507

Figura 13: Raiz do Erro Quadrático Médio entre as previsões geradas pelos modelos LSTM e os dados de teste (observações do ano de 2019)

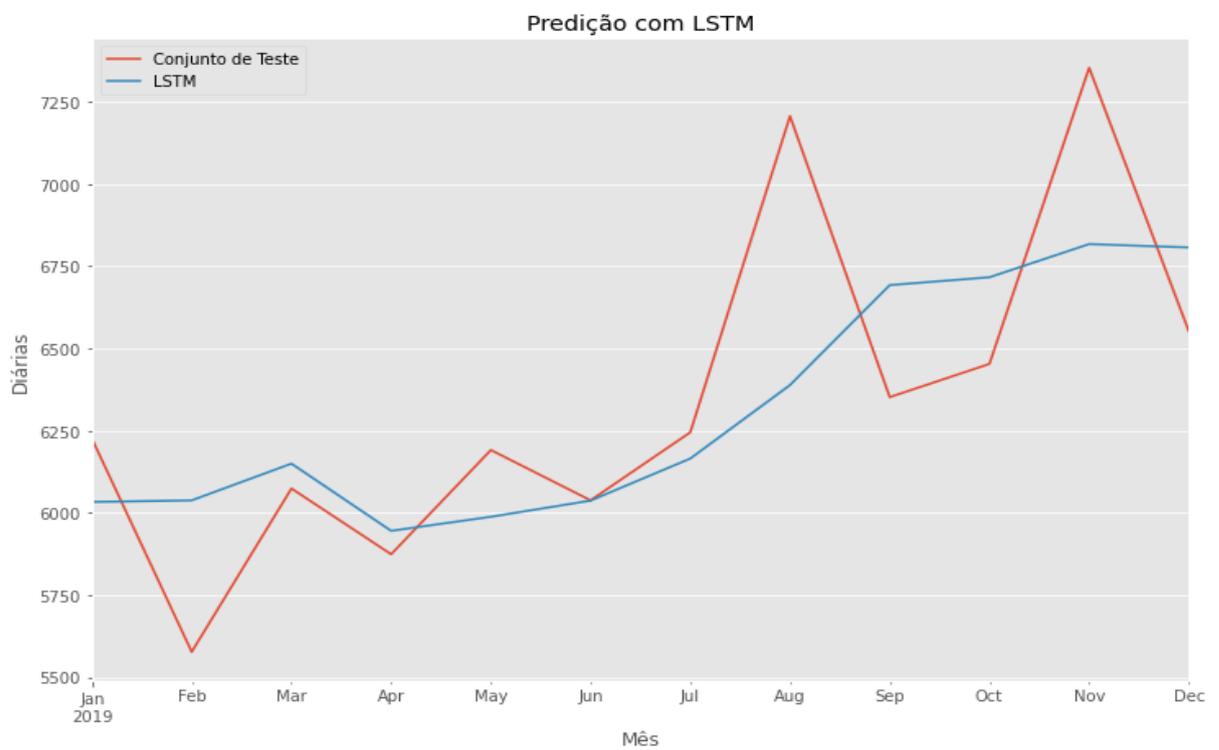


Figura 14: Previsão de diárias de UTI no município de Goiânia para o ano de 2019 com Modelo LSTM comparados com os dados observados no ano de 2019.

5 CONCLUSÕES

Conforme mencionado na Seção 1.2, o objetivo foi concentrar-se apenas nos aspectos gerais da produção hospitalar, com foco principal na oferta de leitos de UTI. Desta forma, considera-se que os objetivos foram alcançados, pois, através apenas da série das diárias de UTI do município de Goiânia, obteve-se um modelo simplificado que fosse capaz de realizar previsões mais confiáveis do que a utilização de modelos de suavização exponencial.

A Figura 15 consolida os resultados do cálculo da Raiz do Erro Quadrático Médio de todos os modelos produzidos. Percebe-se que as previsões produzidas pelo modelo SARIMA sem a variável exógena (estimativa da população de Goiânia) somente foi melhor do que as previsões do *Naive Forecasting*. Isso leva à conclusão de que os ruídos tem forte interferência na série e isso não foi captado pelo modelo. Por outro lado, quando a variável exógena referente à população de Goiânia foi incluída, o modelo SARIMA apresentou a segunda melhor performance acompanhado pela SEH.

	RMSE
Modelo	
LSTM	355.427605
SARIMA-ex(0,1,1)x(3,1,0)[12]	410.013385
Holt	435.784856
Holt Winters	439.442241
MMS-12	450.044552
XGB	493.371938
SARIMA(1,1,0)x(3,1,2)[12]	502.364932
Naive Forecasting	577.020725

Figura 15: Raiz do Erro Quadrático Médio entre as previsões pelos modelos e os dados de teste (observações do ano de 2019)

Observou-se que a série relativa à população de Goiânia tem forte característica de tendência linear. Levando em consideração que as previsões pelo modelo SEH apresentaram boa performance em relação aos demais modelos, constata-se que, nos modelos probabilísticos, as demais componentes (sazonalidade e ruídos) não foram captadas de forma efetiva.

Por fim, os testes demonstraram que o modelo LSTM foi o que apresentou melhor performance preditiva, sendo aproximadamente 15% melhor do que o segundo ranqueado, o que demonstra que houve uma melhor percepção do ruído R_t . Nos experimentos realizados, verificou-se que a quantidade de épocas elevadas e uma taxa de aprendizado muito baixa não trouxeram melhor performance no treinamento, concluindo que o modelo mais simples

foi o que apresentou o menor RMSE. Sendo assim, o modelo LSTM é adequado nas situações onde o Gestor em Saúde Pública conte tão somente com os dados de diárias de UTI, e não haja análise de intervenção que explique as oscilações abruptas na série, o que vem ao encontro dos objetivos do trabalho.

Todavia, consultas realizadas no [DATASUS \(2020a\)](#) demonstraram que há vários grupos de enfermidades que ocasionam ruídos nos quantitativos de utilização dos leitos de UTI, o que levanta várias questões sobre a série. A dificuldade dos modelos na percepção dos ruídos demonstra que a compreensão destas variações em um modelo univariado é limitada, portanto a inclusão de outras variáveis explicativas poderia melhorar a performance das previsões geradas. Neste sentido, a inclusão de técnicas de análise de intervenção com auxílio de especialista da área de saúde pública tende a gerar modelos com performance preditiva mais acurada, sendo um ponto de evolução do presente trabalho. Além disso, uma escala de tempo menor, como dias, por exemplo, beneficiaria o planejamento de leitos, já que há sazonalidades que ocorrem em períodos inferiores a um mês. Entretanto, trata-se de uma limitação do SIHSUS, o que demandaria a busca de outras fontes de dados.

REFERÊNCIAS

ANTUNES, J. L. F.; CARDOSO, M. R. A. Uso da análise de séries temporais em estudos epidemiológicos. **Revista Epidemiologia e Serviços de Saúde**, Brasília, p. 565–576, 2015.

BERGS, J.; HEERINCKX, P.; VERELST, S. Knowing what to expect, forecasting monthly emergency department visits: A time-series analysis. **International Emergency Nursing**, Belgium, n. 22, p. 112–115, november 2014.

BRASIL. Agência Nacional de Saúde Suplementar. **Taxa de ocupação operacional geral**. V1.01. [S.l.], 2012. Disponível em: <<http://ans.gov.br/images/stories/prestadores/E-EFI-01.pdf>>. Acesso em: 20 set. 2020.

BRASIL. Instituto Brasileiro de Geografia e Estatística. **Estimativas de População**. 2020. Disponível em: <<https://www.ibge.gov.br/estatisticas/downloads-estatisticas.html>>. Acesso em: 20 nov. 2020.

BRASIL. Ministério da Educação. Empresa Brasileira de Serviços Hospitalares. **Manual de Conceitos e Nomenclaturas de Leitos Hospitalares**. 1. ed. Brasília, 2016. Disponível em: <http://www2.ebserh.gov.br/documents/695105/1744025/Manual+Leitos+25_01_17+Publicado+Port+16-2017.pdf/571691d5-ccc3-4a70-b403-e024fbe1a7c9>. Acesso em: 12 out. 2020.

BRASIL. Ministério da Saúde. **Atenção hospitalar (leitos e internações)**. [S.l.], 2002. Disponível em: <<http://portalsaude.saude.gov.br/images/pdf/2015/outubro/02/Se----o-A---Aten----o-Hospitalar--Leitos-e-Interna----es.pdf>>. Acesso em: 08 abr. 2020.

_____. **Padronização de nomenclatura do censo hospitalar**. 2. ed. Brasília, 2002.

_____. Estados terão R\$ 432 milhões para enfrentar COVID-19. Brasília, 2020. Disponível em: <<https://www.saude.gov.br/noticias/agencia-saude/46547-estados-terao-r-432-milhoes-para-enfrentar-covid-19>>. Acesso em: 08 abr. 2020.

BRASIL. Ministério da Saúde. DATASUS. Brasília, 2020. Disponível em: <<http://tabnet.datasus.gov.br>>. Acesso em: 04 out. 2020.

_____. **Sistema de Gerenciamento da Tabela de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais do SUS**. Brasília, 2020. Disponível em: <<http://sigtap.datasus.gov.br>>. Acesso em: 12 out. 2020.

_____. Sistema de informações hospitalares do SUS. Brasília, 2020. Disponível em: <<http://datasus1.saude.gov.br/sistemas-e-aplicativos/hospitalares/sihsus>>. Acesso em: 04 out. 2020.

BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. **Manual Técnico do Sistema de Informação Hospitalar**. 1. ed. Brasília, 2007. Disponível em: <http://bvsmis.saude.gov.br/bvs/publicacoes/07_0066_M.pdf>. Acesso em: 04 out. 2020.

BROWNLEE, J. **Deep Learning for Time Series Forecasting**. 1.4. ed. [s.n.], 2018. Disponível em: <<https://machinelearningmastery.com/deep-learning-for-time-series-forecasting/>>. Acesso em: 08 dez. 2020.

_____. **Gentle Introduction to the Adam Optimization Algorithm for Deep Learning**. 2020. Disponível em: <<https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>>. Acesso em: 08 dez. 2020.

_____. **How to Use XGBoost for Time Series Forecasting**. 2020. Disponível em: <<https://machinelearningmastery.com/xgboost-for-time-series-forecasting/>>. Acesso em: 08 dez. 2020.

CABRAL, R. P. **Efeitos da sazonalidade na ocupação de leitos de um hospital filantrópico no município de Fortaleza-CE**. 2012. Dissertação (Mestrado) — Universidade Federal do Ceará, 2012.

CERQUEIRA, D. R. C. et al. **Uma análise da base de dados do Sistema de Informação Hospitalar entre 2001 e 2018: dicionário dinâmico, disponibilidade dos dados e aspectos metodológicos para a produção de indicadores sobre violência**. Rio de Janeiro, RJ: IPEA, 2019. Disponível em: <https://www.ipea.gov.br/portal/images/stories/PDFs/livros/epub/200305_abdsih_2001_2018.epub>. Acesso em: 02 jun. 2020.

CESCONETTO, A.; LAPA, J. S.; CALVO, M. C. M. Avaliação da eficiência produtiva de hospitais do SUS de Santa Catarina, Brasil. **Revista Saúde Pública**, 2008.

CRUZ, A. Agenda de eficiência para o SUS. IV, n. 18, p. 15–23, 2016. Disponível em: <www.conass.org.br/consensus>.

DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American Statistical Association**, v. 74, n. 366, p. 427–431, 1979.

EHLERS, R. **Análise de Séries Temporais**. São Carlos, 2009. Disponível em: <<https://sites.icmc.usp.br/ehlers/stemp/stemp.pdf>>. Acesso em: 12 out. 2020.

GOLDWASSER, R. S. et al. Dificuldades de acesso e estimativas de leitos públicos para unidades de terapia intensiva no estado do Rio de Janeiro. **Revista Saúde Pública**, 2015.

GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor Flow**. 2. ed. Sebastopol - CA: O'Reilly Media, Inc, 2019.

Harvard Global Health Institute. Pandemics explained unlocking evidence for better decision making. 2020. Disponível em: <<https://globalepidemics.org/hospital-capacity>>. Acesso em: 30 jun. 2020.

HYNDMAN, R.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2. ed. [s.n.], 2018. Disponível em: <<https://otexts.com/fpp2/>>. Acesso em: 08 dez. 2020.

JUNIOR, E. Uso da informação na gestão de hospitais públicos. **Ciência & Saúde Coletiva**, v. 12, 2007.

KADRI, F. et al. Time series modelling and forecasting of emergency department overcrowding. **J Med Syst**, n. 38, p. 107, september 2014.

KERAS. [S.l.], 2020. Disponível em: <<https://www.tensorflow.org/guide/keras>>. Acesso em: 08 dez. 2020.

LEITE, G. F. M. **Previsão, por meio de análise de Séries Temporais, de leitos de Unidades de Terapia Intensiva - UTI na rede de atendimento do Sistema Único de Saúde - SUS no município de Goiânia - Código Fonte**. 2020. Disponível em: <https://github.com/Glledson/notebook_analise_leitos_uti_goiania>. Acesso em: 13 dez. 2020.

MORETTIN, P. A.; TOLOI, C. M. **Análise de séries temporais**. 3. ed. São Paulo, SP: Blucher, 2018. v. 1.

NETO, G. V.; MALIK, A. M. Tendências na assistência hospitalar. **Ciência e Saúde Coletiva**, 2007.

PMDARIMA. [S.l.], 2020. Disponível em: <<http://alkaline-ml.com/pmdarima>>. Acesso em: 08 dez. 2020.

RAFFA, C.; MALIK, A. M.; PINOCHET, L. H. C. O desafio de mapear variáveis na gestão de leitos em organizações hospitalares privadas. **Revista de Gestão em Sistemas de Saúde**, 2016.

RAMOS, M. C. A. et al. Avaliação de desempenho de hospitais que prestam atendimento pelo sistema público de saúde Brasil. **Revisata Saúde Pública**, 2015.

SIAMI-NANINI, S.; TAVAKALI, N.; SIAMI NANIN. In: . [S.l.: s.n.].

SOUZA, M. F.; ADRIANO, M. F.; LOPES, L. F. D. Previsão da demanda de leitos hospitalares por meio da análise de série temporais. **Ciência e Natura**, Santa Maria, Brasil, v. 31, n. 1, p. 34–47, 2015.

STATSMODELS. 2020. Disponível em: <<https://www.statsmodels.org/>>. Acesso em: 08 dez. 2020.

TANDBERG, D.; QUAILS, C. Time series forecasts of emergency department patient volume, length of stay, and acuity. **Ann Emerg Med**, Albuquerque, n. 23, p. 299–306, february 1993.

WIKIPÉDIA. **Dilution (neural networks)** — **Wikipédia, a enciclopédia livre**. 2020. Disponível em: <[https://en.wikipedia.org/wiki/Dilution_\(neural_networks\)](https://en.wikipedia.org/wiki/Dilution_(neural_networks))>. Acesso em: 08 dez. 2020.

_____. **Gradient boosting** — **Wikipédia, a enciclopédia livre**. 2020. Disponível em: <https://pt.wikipedia.org/wiki/Gradient_boosting>. Acesso em: 08 dez. 2020.

_____. **Hyperparameter optimization** — **Wikipédia, a enciclopédia livre**. 2020. Disponível em: <https://en.wikipedia.org/wiki/Hyperparameter_optimization>. Acesso em: 08 dez. 2020.

_____. **Rectifier (neural networks)** — **Wikipédia, a enciclopédia livre**. 2020. Disponível em: <[https://en.wikipedia.org/wiki/Rectifier_\(neural_networks\)](https://en.wikipedia.org/wiki/Rectifier_(neural_networks))>. Acesso em: 08 dez. 2020.

_____. **Variáveis exógenas e endógenas** — Wikipédia, a enciclopédia livre. 2020. Disponível em: <https://pt.wikipedia.org/wiki/Variáveis_exógenas_e_endógenas>. Acesso em: 08 dez. 2020.

XGBOOST. 2020. Disponível em: <<https://xgboost.readthedocs.io/en/latest/>>. Acesso em: 08 dez. 2020.