
**UNIVERSIDADE FEDERAL DO PARÁ**

**INSTITUTO DE TECNOLOGIA**

**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

**SAULO JOEL OLIVEIRA LEITE**

DM: 01/2022

# PREDIÇÃO DE SÉRIES TEMPORAIS DA COVID-19: UMA AVALIAÇÃO DO USO DOS MODELOS SUAVIZAÇÃO EXPONENCIAL, ARIMA, MLP & LSTM

**DISSERTAÇÃO DE MESTRADO**

**BELÉM**

**2022**

**SAULO JOEL OLIVEIRA LEITE**

# PREDIÇÃO DE SÉRIES TEMPORAIS DA COVID-19: UMA AVALIAÇÃO DO USO DOS MODELOS SUAVIZAÇÃO EXPONENCIAL, ARIMA, MLP & LSTM

DM: 01/2022

Dissertação apresentada como requisito parcial à obtenção do título Mestre em Engenharia Elétrica com ênfase em Computação Aplicada, do Programa de Pós-Graduação em Engenharia Elétrica, da Universidade Federal do Pará.

Orientador: Prof. Dr. Roberto Célio Limão de Oliveira

**BELÉM**

**2022**

**Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBDSistema de Bibliotecas da Universidade Federal do Pará**

**Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)**

---

L533p Leite, Saulo Joel Oliveira.

PREDIÇÃO DE SÉRIES TEMPORAIS DA COVID-19 : UMA AVALIAÇÃO DO USO DOS MODELOS SUAVIZAÇÃO

EXPONENCIAL, ARIMA, MLP & LSTM / Saulo Joel Oliveira Leite. — 2022.

94 f. : il. color.

Orientador(a): Prof. Dr. Roberto Célio Limão de Oliveira
Dissertação (Mestrado) - Universidade Federal do Pará,

Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2022.

1. COVID-19. 2. LSTM. 3. ARIMA. 4. MLP. 5. Suavização Exponencial. I. Título.

CDD 006.3

---

**UNIVERSIDADE FEDERAL DO PARÁ**
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**"PREDIÇÃO DE SÉRIES TEMPORAIS DA COVID-19: UMA AVALIAÇÃO DOS PREDITORES MLP, LSTM E ARIMA"**

AUTOR: **SAULO JOEL OLIVEIRA LEITE**

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM: 21/01/2022

**BANCA EXAMINADORA:**

---

**Prof. Dr. Roberto Célio Limão de Oliveira**
(Orientador – PPGEE/UFPA)

---

**Prof.ª Dr.ª Adriana Rosa Garcez Castro**
(Avaliadora Interna – PPGEE/UFPA)

---

**Prof. Dr. Lídio Mauro Lima de Campos**
(Avaliador Externo ao Programa – PPGCC/UFPA)

---

**Prof. Dr. Fábio Meneghetti Ugulino de Araújo**
(Avaliador Externo – UFRN)

**VISTO**:

---

**Prof. Dr. Carlos Tavares da Costa Júnior**
(Coordenador do PPGEE/ITEC/UFPA)

Dedico este trabalho a minha mãe, Tânia Cristina (in memoriun).

## AGRADECIMENTOS

Certamente estes parágrafos não irão atender a todas as pessoas que fizeram parte dessa importante fase de minha vida. Portanto, desde já peço desculpas àquelas que não estão presentes entre essas palavras, mas elas podem estar certas que fazem parte do meu pensamento e de minha gratidão.

Agradeço, primeiramente, a Deus.

Agradeço ao meu orientador Prof. Dr. Limão e ao Prof. Dr. Lídio, pela sabedoria na qual me guiaram nesta trajetória.

Agradeço aos membros da banca Doutor Fábio de Araújo e Doutora Adriana Castro.

Aos meus colegas de sala.

A Secretaria do Curso, pela cooperação.

Gostaria de deixar registrado também, o meu reconhecimento à minha família, pois acredito que sem o apoio deles seria muito difícil vencer esse desafio.

Agradeço, em especial, os meus avós, Aldira e Rubens, que me acolheram, vindo de outra cidade, em sua casa e me apoiaram nesta caminhada. Nada disso teria sido possível sem a ajuda deles.

Agradeço ao meu pai, Joel Leite, que apoiou a minha vinda para outra cidade começar uma vida nova, mesmo longe dele e de todos de Bragança, em busca de meus sonhos.

Agradeço grandemente a minha mãe, Tânia Cristina, que mesmo estando hoje no céu, sempre me incentivou a estudar, sempre exigiu que eu melhorasse meu desempenho escolar.

Agradeço a minha noiva, Fernanda Lobato, que esteve ao meu lado durante essa jornada. Sem dúvidas ela é mais que uma amante, é uma grande companheira e amiga.

Enfim, a todos os que por algum motivo contribuíram para a realização deste trabalho.

“Arrisque-se: se você vencer, será feliz;
se perder, será sábio.”
(Peter Kreeft)

## RESUMO

LEITE, Saulo. **Predição de séries temporais da COVID-19:** UMA AVALIAÇÃO DO USO DOS MODELOS SUAVIZAÇÃO EXPONENCIAL, ARIMA, MLP & LSTM. 2022. 94. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal do Pará. Belém, 2022.

Neste trabalho, será discutido como foram desenvolvidas implementações dos modelos preditivos ARIMA, LSTM, MLP e Suavização Exponencial para predição de séries temporais de casos confirmados e mortes por COVID-19, para a avaliação de qual dentre esses obtém o melhor resultado. A COVID-19 é a doença causada pelo coronavírus denominado SARS-CoV-2, que acarretou num grande número de infectados em nível global. Segundo a OMS, até dezembro de 2021, foram estimados mais de 305 milhões de infectados em todo mundo. Como se fez necessário o uso de dados fidedignos para a realização das predições, a base de dados utilizada para o desenvolvimento desse trabalho, é de domínio público e foi cedida pela Universidade Johns Hopkins. Os dados de séries temporais de casos confirmados e mortes do Brasil, Índia, Itália e Estados Unidos da América foram comparados e selecionados para a realização de predições. Acerca dos modelos preditivos, a rede neural *Long Short-Term Memory* é capaz de aprender longas sequências de observações para, deste modo, realizarem previsões. Além desse, a Perceptron Multicamadas (PMC ou MLP — *Multi-Layer Perceptron*) é uma rede neural com uma ou mais camadas ocultas com um número indeterminado de neurônios. Ademais, o ARIMA é um modelo autorregressivo integrado de médias móveis (*autoregressive integrated moving average*). Por fim, a Suavização exponencial (*Exponential Smoothing*) é um modelo de predição altamente preciso para suavizar dados de séries temporais. Assim sendo, após a realização dos treinos e testes de cada um dos modelos, realizou-se a avalição de desempenho com o método de raiz do erro quadrático médio (RMSE) e com base nos resultados dos modelos implementados para a predição de dados referentes aos casos confirmados e as mortes da pandemia de COVID-19, pôde-se avaliar que o modelo ARIMA obteve o melhor desempenho dentre os demais.



**Palavras-chave:** COVID-19. LSTM. ARIMA. MLP. Suavização Exponencial.

## ABSTRACT

LEITE, Saulo. **COVID-19 Time Series Prediction:** AN EVALUATION OF THE USE OF THE EXPONENTIAL SMOOTHING MODELS, ARIMA, MLP & LSTM. 2022. 93. Master's degree dissertation (Master in Electrical Engineering) – Federal University of Pará. Belém, 2022.

In this master's degree dissertation, it will be discussed how the predictive models ARIMA, LSTM, MLP and Exponential Smoothin were developed and implemented to predict time series of confirmed cases and deaths from COVID-19, to assess which among these obtains the best result. COVID-19 is a disease caused by the coronavirus called SARS-CoV-2, which has resulted in a large number of infected people globally. According to the WHO, more than 305 million people are estimated to be infected worldwide. As it was necessary to use reliable data to carry out the predictions, the database used for the development of this dissertation is in the public domain and was provided by the Johns Hopkins University. Time series data of confirmed cases and deaths from Brazil, India, Italy and the United States of America were compared and selected to make predictions. About the predict models, the Long Short-Term Memory neural network is capable of learning long sequences of observations to make predictions. Besides this, the Multi-Layer Perceptron is a neural network with one or more hidden layers with an undetermined number of neurons. In addition, the ARIMA is an autoregressive integrated moving average. Finally, Exponential Smoothing is a highly accurate prediction model for smoothing time series data. Therefore, after carrying out the training and testing of each of the models, the performance evaluation was carried out with the root-mean-square error (RMSE) method and based on the results of the implemented models for the prediction of data referring to the confirmed cases and deaths from the COVID-19 pandemic, it was possible to evaluate that the ARIMA model had the best performance among the others.



**Keywords:** COVID-19. LSTM. ARIMA. MLP. Exponential Smoothin.

## LISTA DE ILUSTRAÇÕES

## LISTA DE TABELAS

## LISTA DE ABREVIATURAS, SIGLAS E ACRÔNIMOS

### LISTA DE ABREVIATURAS

| | |
|---|---|
| COVID | Corona Vírus |

### LISTA DE SIGLAS

| | |
|---|---|
| CBIC | Congresso Brasileiro de Inteligência Computacional |
| MLP | *Multi Layer Perceptron* |
| LSTM | *Long Short Term Memory* |
| MDS | Ministério Da Saúde |
| EQM | Erro Quadrático Médio |
| ES | *Exponential Smoothing* |
| ARIMA | *Autoregressive Integrated Moving Average* |
| RMSE | *Root-Mean-Square Deviation* |
| RNA | Redes Neurais Artificiais |
| ART | *Adaptative Resonance Theory* |
| RBF | *Radial Basis Functions* |
| RNR | Rede Neural Recorrente |
| RNN | *Recurrent neural network* |
| RNS | Redes Neurais Simuladas |
| IC | Inteligência Computacional |
| IA | Inteligência Artificial |
| OMS | Organização Mundial Da Saúde |
| TI | Tecnologia da Informação |

### LISTA DE ACRÔNIMOS

| | |
|---|---|
| PMC | *Perceptron* Multicamadas |
| RNN | Redes Neurais Recorrentes |
| EUA | Estados Unidos da América |

**SUMÁRIO**

# 1 INTRODUÇÃO

A COVID-19 é uma doença infecciosa que ficou mundialmente conhecida como o novo coronavírus. Existem outros membros da família de coronavírus, como o Sars, o Mers e outros agentes infecciosos responsáveis por resfriado comuns. Essa doença alcançou o posto de uma das maiores pandemias da história por conta da taxa de mortalidade considerável, a capacidade de transmissão (inclusive entre assintomáticos) e os diferentes sintomas provocados (SAÚDE, 2021). Até dezembro de 2021, a quantidade de pessoas infectadas pelo mundo chegou a 305 milhões de pessoas, com 5,48 milhões de mortes, segundo levantamento feito pela universidade norte-americana de Johns Hopkins (ESTADÃO, 2021).

Durante a pandemia de COVID-19, a tecnologia se tornou um dos principais pilares de auxílio ao combate ao coronavírus. Atualmente, as soluções utilizadas podem perdurar por anos, auxiliando neste e em outros desafios que estão por vir. Uma pesquisa realizada pela IBM constatou que quase um terço (31%) dos profissionais de TI do mundo dizem que os negócios em que atuam estão usando inteligência artificial (IA), atualmente, e 43% que aceleraram a implementação da tecnologia devido à pandemia de COVID-19 (CIO, 2021).

Diante deste cenário, como forma de apoiar a predição do número de pessoas infectadas e mortas por COVID-19, este estudo tem como objetivo apresentar uma análise de modelos preditivos para o estudo da curva de comportamento dos casos confirmados e mortes por COVID-19. O estudo prevê as taxas de infectados e mortos, considerando a população contaminada de outros países.

Assim sendo, foram utilizados quatro modelos preditores, dentre os mais conhecidos da literatura: ARIMA, LSTM, MLP e Suavização Exponencial com base nos dados de casos confirmados e de mortos por COVID-19 dos países: Estados Unidos, Itália, Índia e Brasil, cujos dados foram disponibilizados pela Johns Hopkins University. Essa instituição tem se comprometido com a fidelidade dos dados, sendo uma das mais determinadas no auxílio do combate à pandemia (MAGGI, 2021).

## 1.1 TRABALHOS RELACIONADOS

O uso de tecnologias da informação para endereçar problemas de saúde é parte comum e já amplamente adotada. Entretanto a incorporação de ferramentas de predição, prescrição e otimização que evoluem rapidamente em complexidade ainda é bastante recente. Análises que permitem fazer predições relevantes a respeito de desfechos clínicos e em gestão de saúde são incorporadas na rotina operacional de uma pequena parte das instituições. A forma de trabalho das equipes que fazem uso dessas ferramentas também requer adaptações e deve focar no uso efetivo de seus resultados. E nesse contexto de desenvolvimento, vivemos uma pandemia histórica (AMARO JUNIOR, 2020).

Não há como negar que as questões de saúde pública são fatores relevantes no desenvolvimento da economia mundial, mudando os mercados, os fluxos de importação e exportação, bem como as atividades da indústria, comércio e turismo (REEVES *et al.*, 2019; TAMBO *et al.*, 2019). Em epidemias que atingem diferentes territórios e causam alto índice de incapacidades na população, a força de trabalho de diversos setores é fortemente afetada, o que reflete na queda do Produto Interno Bruto dos países e seus padrões econômicos (PASQUINI-DESCOMPS *et al*., 2017). Por esse motivo, o estudo dos efeitos de epidemias globais requer análises multidisciplinares, que possam considerar os impactos em cascata desses eventos.

Analisando que as epidemias podem atingir grandes parcelas da população, em grandes espaços territoriais, o uso de tecnologias e inteligência artificial fornece suporte aos gestores públicos e de saúde no entendimento das necessidades de mudanças na dinâmica social (SHAW *et al*., 2020). Assim, é possível planejar o controle dos serviços públicos e das atividades econômicas, de acordo com o diagnóstico de disseminação de doenças e seus efeitos na população afetada.

Nesse sentido, mesmo para os estudos essenciais que buscam a cura ou o tratamento para essas epidemias, a análise de projeções tem um papel importante. Esse fato se deve ao maior conhecimento sobre como as dinâmicas sociais estão associadas à disseminação de doenças; o processo de tomada de decisão para intervenções públicas agora é mais fortemente apoiado por cenários com menos incerteza e vulnerabilidade (JOSHI *et al.*, 2017; SCARABEL *et al.*, 2020).

Em decorrência disso, a partir de estudos de previsão e diagnóstico de comportamentos ao longo de anos pandêmicos, os gestores podem realizar análises

regionalizadas e adaptadas às características de seus sistemas sociais e de saúde, como comércio, turismo local, indústria, educação, logística e infraestrutura hospitalar.

Mediante a esses fatos, a partir de dezembro de 2019, havia uma necessidade mundial de gerenciar os efeitos de uma epidemia, que afetou diretamente os sistemas de saúde e a economia global, decorrente da transformação do Sars-Cov, um vírus previamente identificado em 2003 (ROTA *et al*., 2003). A mutação desse vírus, denominado Sars-Cov-2 ou Corona Vírus (COVID-19), foi prevista por Cheng *et al*. (2007) e identificados em casos iniciais registrados na China, que atingiram 82.361 casos no início de abril de 2020. Outros estudos apresentaram estimativas preliminares do comportamento da curva do COVID-19 (ZHUANG *et al.*, 2020). O contágio pelo vírus assumiu dimensões mundiais, sendo considerado uma pandemia pela Organização Mundial da Saúde (OMS).

Entre os autores mais citados em pesquisas relacionadas ao COVID-19 estão Huang *et al*. (2020), com 2.870 citações, cuja pesquisa avalia o comportamento clínico de pacientes com casos ativos em Wuhan, China. O trabalho de Guan *et al*. (2020), com 1.701 citações, que buscou caracterizar os sintomas de pacientes com COVID-19. Por fim, Chen *et al*. (2020), com 1490 citações, desenvolveu um estudo de caso epidemiológico com 99 pacientes chineses para avaliar o comportamento da doença infecciosa.

Conforme o mencionado, diversos estudos de previsão têm sido desenvolvidos, com base no uso de diferentes tecnologias, a fim de conduzir políticas que minimizem seus efeitos, como medidas de isolamento e bloqueio social. Uma das estratégias utilizadas por países em diferentes epidemias é estimar o comportamento de contaminação (ODRIOZOLA *et al.*, 2017; PÉREZ-CASTRO *et al.*, 2016). Essas medidas visam proporcionar a menor disseminação do vírus, o deslocamento de um maior número de profissionais de saúde para atendimentos especializados, bem como a infraestrutura hospitalar disponível (BENVENUTO *et al.*, 2020; JIANG *et al.*, 2020).

Assim sendo, a existência de estudos que desenvolvam tecnologias de inteligência artificial auxilia na entrega de previsões, soluções, produtos, serviços e inovações (WANG, 2019). Essas soluções podem oferecer uma redução dos impactos econômicos e de saúde da população durante os períodos de pandemias. Além disso, fornecem subsídios para a tomada de decisões públicas, com base em modelos matemáticos e estatísticos, identificando fatores de influência regionalizados para cada população afetada.

Dentro os trabalhos relacionados às predições de casos no Brasil, destaca-se a publicação da Lima *et al.* (2020), que descreve a utilização do modelo ARIMA para a predição de casos confirmados da COVID-19 em alguns estados do país, com um intervalo de confiança de 95% e projeção de 6 dias.

Além desse trabalho, há o artigo de Narváez *et al.* (2020), onde tem como objetivo explorar o melhor tipo de curva ou modelo de tendência que possa explicar o comportamento epidemiológico da infecção por COVID-19 no chile e derivar as possíveis causas que contribuem para explicar o modelo correspondente e as implicações para a saúde que se podem inferir. São utilizados vários modelos preditivos, dentre eles a Suavização Exponencial, um dos modelos de predição mais antigo, utilizado em várias áreas (GONZAGA, 2021).

Assim sendo, a utilização de redes neurais artificiais pode auxiliar na análise de problemas em diferentes contextos (MARTINS, 2012; ZHANG *et al*., 2018), considerando dados epidemiológicos em estudos COVID-19 (GHAZALY *et al.*, 2020; MOLLALO *et al*., 2020; SABA, 2020). No entanto, ainda existem poucos estudos que usaram essa técnica para prever o comportamento pandêmico (AKHATAR *et al*, 2019; KAWAGUCHI *et al*., 2020).

## 1.2 MOTIVAÇÕES

A infecção por coronavírus é uma doença causada pelo vírus SARS-COV-2, popularmente conhecido como coronavírus, é deduzido que se originou em Wuhan, na China, em dezembro de 2019. Conjectura-se que o vírus tenha uma origem zoonótica, pois os primeiros casos confirmados tinham ligações principalmente ao Mercado Atacadista de Frutos do Mar de Huanan, local onde também se vendiam animais vivos (SURVEILLANCES, 2020). Por meio de estudos, a OMS afirma que é amplamente confirmada a disseminação comunitária, de pessoa para pessoa, por gotículas respiratórias ou por contato (OMS, 2021).

Os pacientes diagnosticados com essa doença possuem um quadro clínico que varia de infecções assintomáticas a quadros de infecção respiratória grave (MDS, 2021). Segundo a Organização Mundial da Saúde (OMS), cerca de 80% dos pacientes com infecções por coronavírus podem ser assintomáticos, enquanto 20% dos casos

podem necessitar de atendimento hospitalar, sendo que, 5% desses casos, poderá ainda haver a necessidade de tratamento para insuficiência respiratória (OMS, 2021).

Devido ao aumento exponencial dos casos e da sua alta capacidade de transmissão, o coronavírus se tornou um surto emergencial de saúde pública e de interesse internacional. Em 11 de março, a doença foi caracterizada como pandemia pela Organização Mundial da Saúde (OMS, 2020), devido a presença de casos positivos em diferentes continentes (SOHRABI, 2020).

O número de casos confirmados cresceu bastante antes da criação da vacina, já ultrapassando os 273 milhões de infectados mundialmente, no segundo semestre de 2021. A taxa de letalidade reportada é de aproximadamente 6,7% e já somam mais de 5 milhões de mortes no mundo (OXFORD, 2021).

Os sintomas mais comuns de infecções por coronavírus são: febre, tosse seca e cansaço. Segundo a Organização Mundial da Saúde (OMS), a maioria dos infectados, cerca de 80%, se recuperam da doença sem precisar de tratamento hospitalar, enquanto cerca de um a cada cinco pessoas que contraem infecções por coronavírus ficam gravemente doentes, necessitando tratamento especializado.

Os pacientes que apresentam sintomas característicos da doença são submetidos a exames biológicos. Esses exames são compostos pela coleta de material respiratório, através da aspiração de vias aéreas ou indução de escarro, o qual é submetido à exames de biologia molecular, a fim de verificar a presença do RNA viral. Além deste, os testes complementares, também conhecidos como testes rápidos, realizam a coleta de sangue e verificam a presença de anticorpos ou do RNA viral (TANG, 2020).

No decorrer da pandemia de COVID19, as medidas restritivas, juntamente com a aplicação das vacinas são a melhor alternativa para conter à pandemia. Os testes de COVID-19 são caros e muitas vezes inacessíveis em determinadas regiões do mundo, tornando difícil a sua aplicação em massa para a obtenção de dados precisos da situação pandêmica, dificultando, desse modo, a tomada de decisão através de políticas públicas para conter os avanços da doença. Portanto, além da aplicação de vacinas para conter o avanço da contaminação, com predições do comportamento dos casos confirmados e mortes, medidas mais assertivas poderão ser tomadas.

## 1.3 OBJETIVOS E CONTRIBUIÇÕES

Este trabalho tem como objetivo central desenvolver uma metodologia que se baseia na implementação de modelos preditivos, sendo esses: ARIMA, LSTM, MLP e Suavização Exponencial. Deste modo, com base nos dados obtidos, poderá ser avaliado qual modelo obteve um melhor resultado. Assim sendo, a intenção primordial é que este trabalho possa vir a contribuir para pesquisas realizadas pela comunidade acadêmica.

A definição do objetivo geral resultou no entendimento acerca da problemática retratada, fazendo com que fosse possível realizar a delimitação do escopo deste estudo que se materializa através dos seguintes objetivos específicos:

- Implementar os modelos MLP, ARIMA, LSTM e Suavização Exponencial para treinamento e teste, tendo como entrada uma base de dados de casos confirmados e mortes por COVID-19;
- Implementar a medida de desempenho da Raiz do Erro Quadrático Médio (RMSE), para que esses resultados obtidos pelos quatro modelos, possam ser comparados.

A partir da definição dos objetivos deste trabalho pode-se definir como principais contribuições:

- Disponibilização de uma metodologia útil à comunidade acadêmica e aos especialistas da área de doenças infectuosas de forma a possibilitar o auxílio na predição do desenvolvimento da doença de COVID-19;
- Disponibilização do código à comunidade para a implementação;
- Auxílio no controle da propagação de pandemias e suporte na garantia de desenvolvimento saudável da população mundial.

## 1.4 ORGANIZAÇÃO DA DISSERTAÇÃO

O presente trabalho está organizado em cinco capítulos dispostos da seguinte forma:

- O Capítulo 1 descreve o contexto no qual este trabalho está inserido através de uma breve introdução, trabalhos relacionados, motivações, objetivos e contribuições;

- O Capítulo 2 introduz os principais conceitos acerca dos modelos MLP, LSTM e ARIMA e Suavização Exponencial utilizados neste trabalho;
- O Capítulo 3 aborda a metodologia empregada na tarefa do desenvolvimento e da aferição do desempenho dos quatro modelos implementados;
- O Capítulo 4 apresenta e discute os resultados obtidos com a aplicação dos modelos, a análise e comparação dos resultados;
- O Capítulo 5 discute as considerações finais e os trabalhos futuros que podem ter origem a partir deste trabalho.

# 2 FUNDAMENTAÇÃO TEÓRICA

Esta fundamentação teórica descreve, brevemente, o funcionamento dos modelos preditivos selecionados para a implementação neste trabalho. Além disso, é descrito de forma sucinta o funcionamento de técnicas e métricas de avaliação de desempenho também utilizadas no desenvolvimento da implementação dos modelos. A próxima sessão retrata o funcionamento das redes neurais artificiais.

## 2.1 REDES NEURAIS ARTIFICIAIS

As redes neurais, também conhecidas como redes neurais artificiais (RNAs) estão no núcleo dos algoritmos de aprendizagem de máquina (*machine learning*). Seu nome e estrutura são inspirados no cérebro humano, imitando a maneira como os neurônios biológicos enviam sinais uns para os outros (IBM, 2021).

É importante ressaltar que essa análise matemática tem sua inspiração biológica, pois o cérebro é constituído por unidades funcionais, denominadas neurônios, que emitem impulsos elétricos em resposta a estímulos, tal como discutido por Timoszczuk (2004). Essas unidades são densamente interconectadas, resultando em uma arquitetura altamente complexa. Inspirados nessa arquitetura complexa, vários modelos matemáticos de neurônios artificiais foram desenvolvidos, refletindo o conhecimento biológico sobre o funcionamento do cérebro humano (TAGLIARINI *et al.*, 1991).

A inspiração para a criação de redes neurais artificiais surgiu da tentativa de compreender o cérebro humano. As redes neurais artificiais buscam ser parecidas com as redes biológicas, tendo estrutura, função, técnicas de processamento de dados e métodos de cálculo muito similares. As redes neurais artificiais podem aprender com os dados e resolver problemas complexos (CAMILO; SILVA, 2019).

O cérebro humano é composto por um grande número neurônios, aproximadamente 100 bilhões. Cada neurônio biológico (Figura 1) é uma célula especializada que pode criar, receber e propagar sinais eletroquímicos (GOLDSCHMIDT; PASSOS, 2005; BASTIANI, 2017).

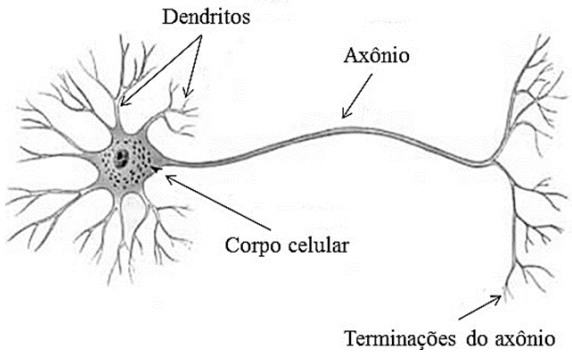


**Figura 1 - Neurônio Biológico**
**Fonte: icmc.usp.br**

Segundo Guyon (1991), o neurônio artificial é uma estrutura lógico-matemática que procura simular a forma, comportamento e funções de um neurônio biológico. Assim, os dendritos são substituídos por entradas, cujas ligações com o corpo celular artificial são realizadas através de elementos chamados de peso (simulando as sinapses neuronais). Os estímulos são captados pelas entradas e processados pela função de soma e o bias. O limiar de disparo do neurônio biológico tem como analogia a função de ativação no neurônio artificial (CHUA L.O.; YANG L., 1988). A Figura 2 exibe uma representação do neurônio artificial.

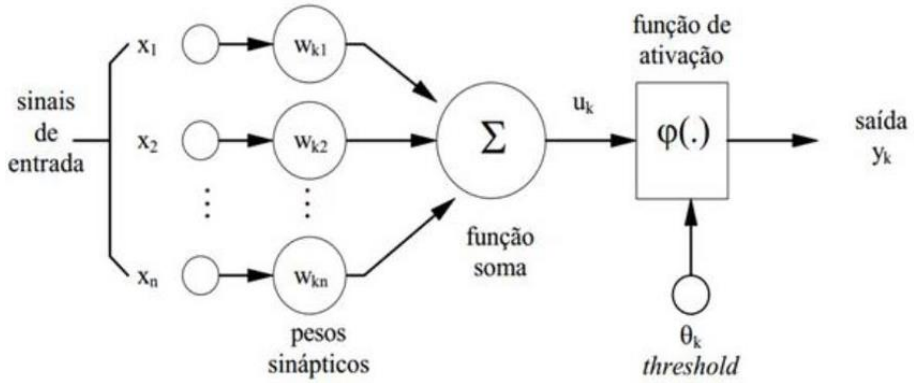


**Figura 2 - Modelo de um Neurônio Artificial**
**Fonte: (HAYKIN, 2001)**

## 2.2 REDES NEURAIS PERCEPTRON MULTI-CAMADAS (MLP)

As redes *Perceptron* Multicamadas (*MultiLayer Perceptron*), são caracterizadas pela presença de múltiplas camadas de unidades básicas de processamento do tipo *Perceptron*. O incremento de mais camadas neurais implica

em um aumento na capacidade de processamento não linear e generalização da rede neural. Assim, foram superadas as limitações do modelo *Perceptron* proposto por Minsky e Papert, em 1969 (GUYON, 1991).

A rede MLP é composta minimamente por uma camada de entrada e uma camada de neurônios de saída, sendo que entre essas camadas podem conter inúmeras camadas intermediárias, também denominadas de camadas escondidas. Cada camada pode conter inúmeros neurônios *Perceptrons*, sendo que as saídas dos neurônios de cada camada são as entradas dos neurônios da camada subsequente, desde a camada de entrada, até a camada de saída. Neste tipo de rede não há retroalimentações, pois a propagação do processamento neural é unidirecional (SILVA, 2015; BASTIANI, 2017).

As redes MLP possuem treinamento supervisionado, sendo que o algoritmo de treinamento visa ajustar os pesos da rede a partir de um conjunto conhecido de vetores de entrada para obter da rede um conjunto de saídas desejadas (BINOTI, 2010). A Figura 3 exemplifica uma rede *Perceptron* Multicamadas.

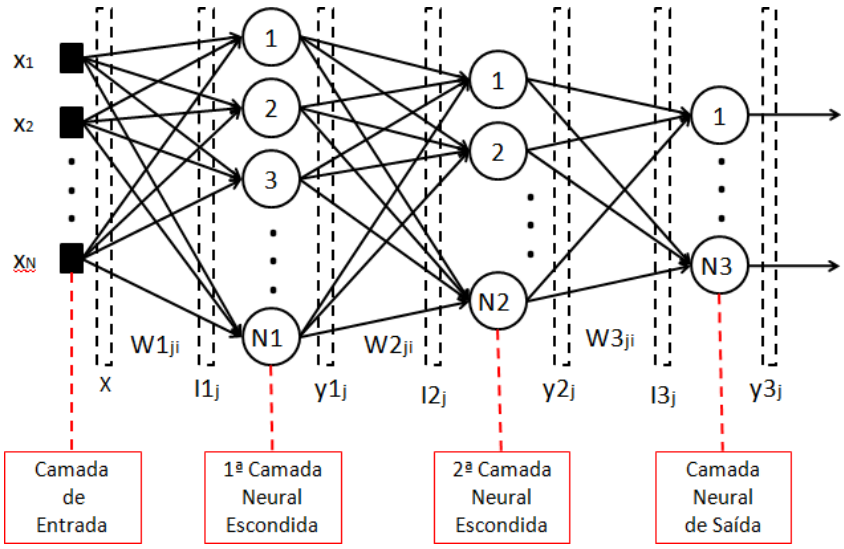


**Figura 3 - Perceptron Multicamadas**
**Fonte: (SILVA, 2002)**

Na Figura 3 tem-se a seguinte simbologia:

- **N:** Número de pontos do vetor de entrada **X**;
- **X:** Vetor de entrada da rede: $X = [-1, x1, x, \dots, x_N]^T$;
- **Wlji**: Valor do peso sináptico conectado ao *j-ésimo* neurônio da camada l ao *i-ésimo* neurônio da camada (l-1); ressalta-se que nesta convenção o valor

do peso $\mathbf{W_{j0l}}$, para todos valores de **l** e **j**, teremos o valor do bias (ou limiar) de cada neurônio;

- $\mathbf{I_{lj}}$**:** Valor da entrada ponderada do j-ésimo neurônio da camada **l**.

Silva (2015) ressalta que as redes MLP com entradas atrasadas, pertencem a arquitetura *feedfoward* de múltiplas camadas. A predição é realizada a partir de um instante t, com base nos valores anteriores da série, de acordo com a Equação 1 abaixo.

$$x(t) = f(x(t\text{-}1), x(t\text{-}2), \ldots, x(t\text{-}np)) \quad (1)$$

Onde $np$ é a quantidade de medidas passadas necessárias para estimação do valor. A Figura 4 apresenta a topologia de rede. Verifica-se por meio da figura que a rede possui atrasos temporais na camada de entrada. Esta linha de atrasos corresponde a uma memória, de forma a garantir que amostras anteriores sejam sempre consideradas pela rede na estimação do valor futuro (BASTINI, 2017).

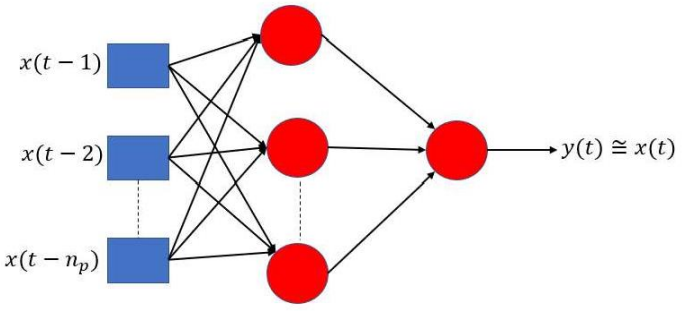


**Figura 4 - Topologia Para a MLP com Entradas Atrasadas no Tempo**
**Fonte: dee.ufc.br**

Durante o processo de treinamento a rede deve minimizar o erro entre a saída real e a saída estimada, sendo este erro dado pela expressão: $E = x(t)\text{-}y(t)$ , para o intervalo $(n_p + 1) \leq t \leq N$ , onde N é a quantidade total de amostras da série temporal. (SILVA, 2015). Assim sendo, a rede MLP utilizada neste trabalho tem apenas x(t-1) na sua entrada.

## 2.3 REDES NEURAIS RECORRENTES

Segundo Braga (2019) a computação convencional não realiza de forma satisfatória tarefas de reconhecimento de padrão dinâmico, caso das séries temporais. Uma série temporal corresponde a uma coleção de observações feitas sequencialmente ao longo do tempo (BROWLEE, 2020).

Assim sendo, as séries temporais necessitam de estruturas que sejam capazes de representar o tempo e apresentar memória. Dentre estas estruturas, tem-se as Redes Neurais Recorrentes (*recurrent neural network* - RNN). As RNNs são redes em que as saídas dos neurônios são realimentadas como sinais de entrada para outros neurônios. Desse modo, apresentam laços de repetição (*loops*) que permitem que a informação persista ao longo da rede. Uma rede recorrente pode ser vista como uma cópia dela mesma, cada uma passando mensagem para seu sucessor (HAYKIN, 2005). A Figura 5 ilustra uma rede desdobrada.

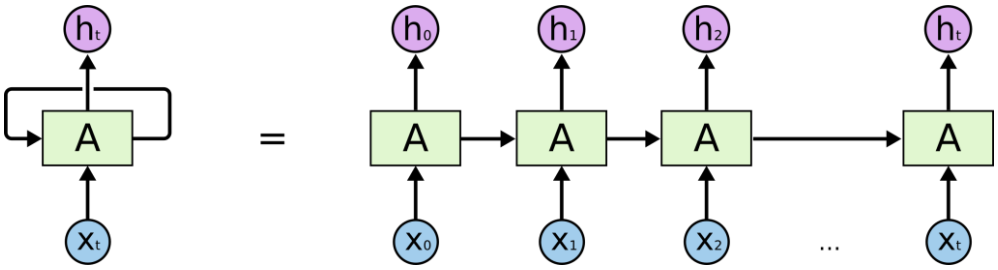


**Figura 5 - Rede recorrente desdobrada**
**Fonte: dataio.ir**

Na Figura 5 temos a estrutura de uma Rede Neural *Long Short-Term Memory* que possui uma cadeia que contém quatro redes neurais e diferentes blocos de memória chamados células. As Redes LSTM possuem só uma entrada, as setas em *loop* (laço de repetição) indicam a natureza recursiva da célula. Isso permite que as informações dos intervalos anteriores sejam armazenadas na célula LSTM. O estado da célula anterior esquece, multiplica-se com a porta do esquecimento e adiciona novas informações através da saída das portas de entrada (DAS, 2021).

Nos últimos anos, as RNNs têm sido aplicadas em diversos problemas: modelagem de idiomas, tradução, séries temporais, entre outros. Em séries temporais tem-se como objetivo prever o próximo valor de acordo com valores anteriores. Portanto, a entrada para a RNN, em cada intervalo de tempo, é o valor atual, bem

como o vetor de estado com os dados do período anterior, que é a memória. Em todos estes problemas a questão da sequência e da memória são relevantes. Sendo que as redes *Long Short Term Memory* (LSTM) foram as que mais tiveram êxito dentre as redes recorrentes, porque lidam melhor com as dependências de longos prazos (BRAGA, 2019).

### 2.3.1 Rede Neural Long Short-Term Memory (LSTM)

A rede neural *Long Short-Term Memory*, ou rede LSTM, é uma rede neural recorrente que é treinada usando *Backpropagation Through Time* e supera o problema do desaparecimento do gradiente. Esse desaparecimento gradiente ocorre porque redes neurais convencionais não “armazenam informações” (DAS, 2021). Dessa forma, a rede LSTM pode ser usada para criar grandes redes recorrentes que, por sua vez, podem ser usadas para resolver problemas de sequência difíceis no aprendizado de máquina (FILIPA, 2020).

Diferente das redes neurais convencionais, as redes neurais LSTM têm blocos de memória que são conectados por meio de camadas. Um bloco possui componentes que o tornam mais inteligente do que um neurônio clássico e uma memória para sequências recentes. Um bloco contém portas que gerenciam o estado e a saída do bloco. Um bloco opera em uma sequência de entrada e cada porta dentro de um bloco usa as unidades de ativação sigmoide para controlar se elas são disparadas ou não, tornando a mudança de estado e adição de informações que fluem através do bloco condicional (FILIPA, 2020).

A LSTM tem a capacidade de remover ou adicionar informações ao estado da célula, cuidadosamente reguladas por estruturas chamadas portas. Vale destacar que, quanto mais blocos de células LSTM a rede neural possuir, maior a sua possibilidade de “aprender” com o tempo (JÚNIOR, 2019).

Para entender o modelo de forma matemática, a Equação 2 a seguir resumem tudo o que foi discutido acima.

$$h_t = \emptyset(b_h + xW_x + h_{t-1}W_h) \quad (2)$$

Para um melhor entendimento da Equação 2, os seus componentes serão descritos.

- $h_t$: O vetor de estado oculto, também conhecido como vetor de saída;
- $\emptyset$: A função de ativação;
- $b_h$ e $W_h$: Os pesos das conexões de entrada;
- $h_{t-1}$: O estado oculto anterior do vetor de saída.

Nota-se que a diferença fundamental entre as equações acima e as das redes neurais clássicas é que são acionadas informações do estado oculto do período anterior. Além disso, observa-se que os parâmetros que fazem a transição da informação entre os estados ocultos, de diferentes períodos, são sempre os mesmos. Isso mostra que redes neurais recorrentes compartilham parâmetros através do tempo. Na prática, define-se a recorrência acima como um laço de repetição no código (BROWLEE, 2020).

No exemplo a seguir, as equações da RNN foram escritas de forma desdobrada no tempo. Desse modo, é considerada uma RNN com uma camada oculta, processando quatro períodos e realizando uma previsão de uma variável contínua (problema de regressão), apenas após observar os 4 períodos (GERS, 2001).

$$H_0 = \emptyset(b_h + xW_x + XW_x) \quad (3)$$

$$H_1 = \emptyset(b_h + xW_x + XW_x + H_0W_h) \quad (4)$$

$$H_2 = \emptyset(b_h + xW_x + XW_x + H_1W_h) \quad (5)$$

$$H_3 = \emptyset(b_h + xW_x + XW_x + H_0W_h) \quad (6)$$

Vale ressaltar que, não é necessário escrever manualmente as equações, já que os frameworks de *Deep Learning*, como o TensorFlow, realizam as operações matemáticas automaticamente, com bases nos dados recebidos (ABADI *et al.*, 2015). No caso da pesquisa desta dissertação, são utilizados dados de séries temporais de casos confirmados e mortes por COVID19, sendo apenas uma entrada, que se trata da quantidade acumulativa dos dados da pandemia.

## 2.4 MODELO SUAVIZAÇÃO EXPONENCIAL (EXPONENTIAL SMOOTHING)

A suavização exponencial foi proposta no final dos anos 1950 e motivou alguns dos métodos de previsão de maior sucesso. As previsões produzidas usando métodos de suavização exponencial são médias ponderadas de observações anteriores, com os pesos decaindo exponencialmente à medida que as observações envelhecem. Em outras palavras, quanto mais recente a observação, maior o peso associado (BROWN, 1959; HOLT, 1957; WINTERS, 1960).

A seguir serão apresentados breves resumos dos mais importantes métodos de suavização exponencial e sua aplicação na previsão de séries temporais com suas características. Isso ajuda a desenvolver uma intuição de como esses métodos funcionam.

### *Algoritmo de suavização exponencial simples*

É o mais simples de todos os algoritmos e é o menos usado. Esse algoritmo só tem um componente não observável, que é o nível. O nível não precisa ser globalmente fixo e pode evoluir ao longo do tempo. Esse tipo de modelo é utilizado quando não existe um padrão claro de sazonalidade ou tendência na série. (FALBEL, 2019).

### *Algoritmo de Suavização Holt*

O algoritmo de suavização *Holt* difere do algoritmo de suavização simples, pois é inserido um outro componente não observável: a tendência. A tendência é o movimento de crescimento ou de decrescimento da série temporal e que ocorrem de forma sustentada. A tendência pode ser global (estar em toda série) ou local (em uma parte da série) (GOMES, 2019).

### *Algoritmo de Suavização Holt-Winters*

Uma evolução do modelo linear de Holt foi criado por Holt e Winter para possibilitar a modelagem de séries temporais por suavização exponencial que também possuam um componente sazonal. O algoritmo de suavização *Holt-Winters*

possui uma componente não observável a mais que o algoritmo de *Holt*: A Sazonalidade. Sazonalidade são comportamentos que ocorrem na série temporal de forma repetitiva e com uma periodicidade inferior a um ano. Esse algoritmo pode ser aditivo ou multiplicativo. A diferença entre o aditivo e o multiplicativo é que onde há diferença é substituída pela razão entre os componentes (GONZAGA, 2021).

*Algoritmo de Suavização Pegels*

Pegels (1969) propôs uma taxonomia de métodos de suavização exponencial, onde cada método tem uma componente de tendência e um componente sazonal. Este foi posteriormente analisado por Gardner (1985). O algoritmo de *Pegels* é pouco utilizado. Ele se diferencia do *Holt* pelo fato de adicionar a tendência de forma multiplicativa e isso pode ser estendido para o caso sazonal (HYNDMAN, 2002).

## 2.5 MODELO AUTO-REGRESSIVO INTEGRADO DE MÉDIAS MÓVEIS (ARIMA)

O Método Autorregressivo Integrado de Médias Móveis (ARIMA) foi elaborado por Box e Jenkins em 1970 para prever séries temporais. Valenzuela *et al*. (2008) afirmam que para aplicação desse é necessária a suposição de que as variáveis tenham uma relação de autodependência linear.

Khashei e Bijari (2011) informam que o método ARIMA é um dos mais consagrados modelos lineares, para previsão de séries temporais. Este é largamente adotado em modelos híbridos com o objetivo de elevar a capacidade de predição.

Séries temporais podem ser descritas por modelos paramétricos ou não paramétricos. Esses modelos são processos estocásticos, ou seja, regidos por leis probabilísticas. Na classe de modelos paramétricos, uma metodologia muito utilizada é a proposta por Box e Jenkins (1976), que consiste em ajustar modelos autorregressivos integrados de médias móveis, ARIMA (**p,d,q**), a uma dada série temporal, onde:

- **p** é a ordem do modelo autorregressivo;
- **d** é o grau de diferenciação e;
- **q** é a ordem do modelo de média móvel.

A classe dos modelos ARIMA é capaz de descrever séries estacionárias e não estacionárias do tipo homogênea. Séries temporais estacionárias são aquelas que se desenvolvem no tempo em torno de uma média constante. O tipo não estacionário homogêneo, flutua por certo tempo em torno de um nível, e depois passa a flutuar ao redor de um novo nível e assim por diante (MORETTIN; TOLOI, 2006).

Quando as séries apresentam esse tipo de não-estacionaridade, são tomadas diferenças sucessivas da série original até a série tornar-se estacionária. O número de diferenças, que são efetuadas para alcançar a estacionaridade, é a ordem d do modelo ARIMA (p,d,q). Assim, quando a série atinge esta condição o modelo é reduzido à forma autorregressiva integrada de médias móveis, ARIMA (p,q). Onde p e q são as ordens dos processos autorregressivos e médias móveis respectivamente (JENKINS, 1976).

Se uma das ordens p ou q forem nulas, o modelo torna-se autorregressivo de ordem p, AR(p), ou médias móveis de ordem q, MA(q) respectivamente. Esses modelos mais simples, apresentados nas pesquisas de Yule (1963) e Wold (1938), fundamentaram o trabalho de Box & Jenkis na construção dos modelos ARIMA.

De forma geral, os modelos ajustados por meio da metodologia Box & Jenkis são parcimoniosos, e as previsões obtidas são bastante precisas comparadas com os demais métodos de previsão (MORETTIN; TOLOI, 2006). Um fator limitante no procedimento é a identificação do modelo, pois depende de razoável experiência do analista (CHAVES, 1991).

Por fim, é de suma importância relatar que diferente dos modelos MLP e LSTM o modelo ARIMA não precisa de hiperparâmetros na sua implementação (SPRINGML, 2021). Para o melhor entendimento desses, a sessão a seguir descreve os principais que foram utilizados nesse trabalho.

## 2.6 HIPERPARÂMETROS DAS REDES NEURAIS E MÉTODO DE AVALIAÇÃO DE DESEMPENHO

Como mencionado, alguns modelos preditivos contam com hiperparâmetros, que são parâmetros ajustáveis que permitem controlar o processo de treinamento do modelo. Além desses, para avalição dos resultados do modelo na etapa de teste, faz-

se necessário o uso de uma métrica de avalição de desempenho. Todas essas informações encontram-se a seguir.

### 2.6.1 Hiperparâmetros

O otimizador utilizado no código deste trabalho foi o *Adam*. A sua otimização é um método estocástico de descida gradiente, baseado na estimativa adaptativa, de momentos de primeira e segunda ordem (KERAS, 2021). Esse método foi escolhido, pois é computacionalmente eficiente, tem pouca necessidade de memória, invariante para redimensionamento diagonal de gradientes e é adequado para problemas que são grandes em termos de dados/parâmetros (KINGMA, 2021).

Além do otimizador, outro método de hiperparâmetro importante utilizado nas redes neurais foi o *Dropout*. Essa é uma técnica de regularização para reduzir o sobre-ajuste em redes neurais artificiais, evitando co-adaptações complexas nos dados de treinamento. É uma maneira eficiente de realizar a média do modelo com redes neurais (KERAS, 2021).

O hiperparâmetro de taxa de aprendizado, também conhecido como "*learing rate*", tem grande influência durante o processo de treinamento da uma rede neural. Tendo em vista que uma taxa de aprendizado muito baixa torna o aprendizado da rede muito lento, ao passo que uma taxa de aprendizado muito alta provoca oscilações no treinamento e impede a convergência do processo de aprendizado (KERAS, 2021).

### 2.6.2 Método de avaliação de desempenho

Segundo Randolfo (2020), após treinar um modelo de predição, o procedimento mais comum é testá-lo. Assim, é analisado se o modelo é capaz de generalizar bem para novos dados. Desse modo, se o modelo possui um bom desempenho para predizer os dados de treino, mas não o tem ao prever os dados de teste, há um problema de *overfitting*. No entanto, o *underfitting* é quando o modelo não conseguiu aprender suficiente sobre os dados (BRANCO, 2020).

Neste trabalho, o método de avaliação de desempenho dos modelos utilizado foi de raiz do erro quadrático médio, também conhecido "*Root-Mean-Square Deviation*" RMSE. Esse método representa a raiz quadrada do MSE (erro quadrático

médio), que é o erro ao quadrado médio das previsões do modelo. O RMSE mede a diferença entre os valores previstos pelo modelo e os valores observados. O RMSE é como o "desvio padrão dos erros". Quanto maior esse número, pior o modelo. O seu uso é destinado a compreender um "erro de previsão" (AZANK, 2020).

No próximo capítulo será descrito como foi realizado o desenvolvimento, com base nas informações discutidas nos capítulos anteriores.

# 3 DESENVOLVIMENTO

O desenvolvimento está dividido em seções, estas descrevem como realizou-se a aquisição e preparação do conjunto de dados, para que esses pudessem ser usados nas predições. Além disso, é descrito como implementou-se os modelos ARIMA, MLP, LSTM e Suavização Exponencial para as predições de séries temporais de casos confirmados e mortes por COVID-19. Finalmente, após seguir essas etapas, é discutido como foi feita a predição e a comparação de seus respectivos resultados obtidos.

## 3.1 AQUISIÇÃO E PREPARAÇÃO DE DADOS

A primeira etapa é realizar a aquisição dos dados. Os dados utilizados são disponibilizados pela Universidade de Johns Hopkins. Essa instituição tem se dedicado em disponibilizar dados relacionados a COVID-19, sendo umas das primeiras a iniciar a distribuição destes (MAGGI, 2021). Como esses dados recebidos são de diversos países, o primeiro passo é tratá-los e extrair somente os dados que são necessários para este trabalho. Sendo esses os dados do Brasil, Índia, Itália e Estados Unidos da América.

Os dados obtidos são em forma de série temporal. No caso deste trabalho, as séries temporais são das quantidades diárias de casos confirmados e mortes por COVID-19 no Brasil, Índia, Itália e EUA ao decorrer do período de janeiro de 2020 a dezembro de 2021. Na Figura 6 é possível observar, de forma gráfica, os dados obtidos de casos confirmados por COVID-19 desses países. Na Figura 7 é mostrado o número de mortos na pandemia nestes mesmos países.

Os dados da Tabela 1 foram retirados da base de dados da Universidade de Jhons Hopkins, na data de 31/12/2021. Analisa-se que os Estados Unidos da América é o país com a maior quantidade de casos confirmados e de mortes por COVID19. Na análise, a Índia ocupa o segundo lugar, com o número de caso confirmados abaixo dos EUA. No entanto, o Brasil é o segundo país com maior quantidade de mortes entre os quatro, atrás dos EUA. A Itália fica em última nos dois gráficos por ser um país consideravelmente menor, em quantidade de habitantes.

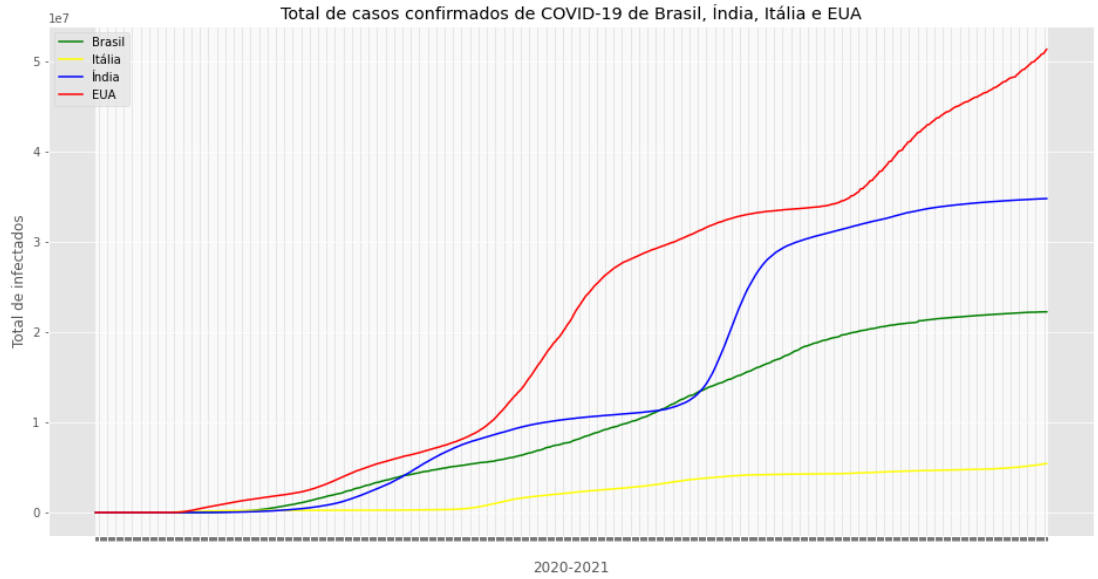


**Figura 6 - Casos confirmados de COVID-19, no Brasil, Índia, Itália e EUA, de 2020 e 2021**
**Fonte: Autor**

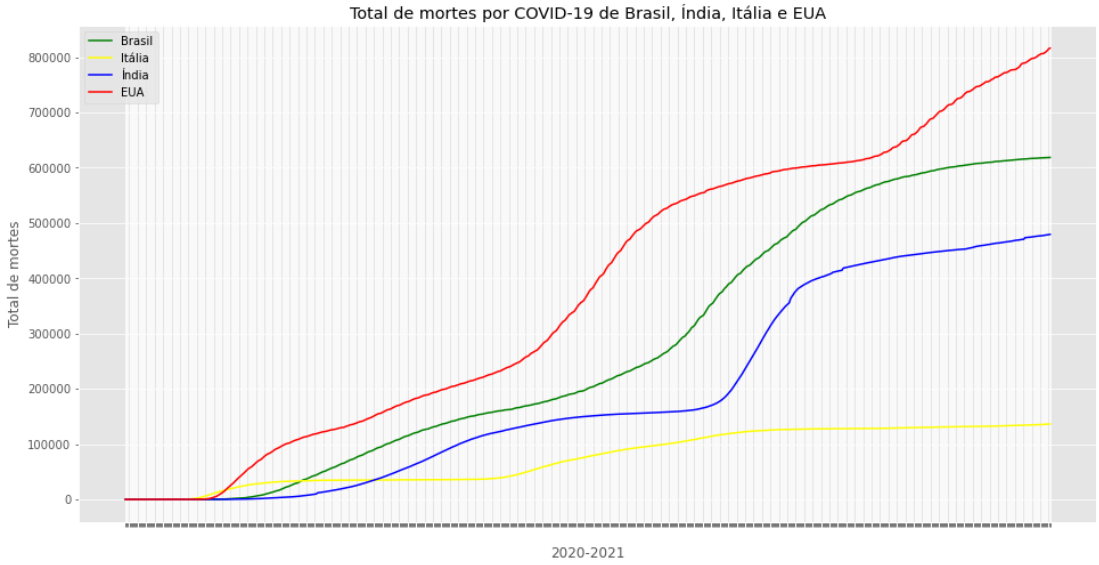


**Figura 7 – Mortes por COVID-19, no Brasil, Índia, Itália e EUA, de 2020 e 2021**
**Fonte: Autor**

**Tabela 1 – Total de casos confirmados e mortes por COVID19**

| Países | Casos confirmados | Mortes |
|---|---|---|
| Brasil | 22.291.839 | 619.334 |
| Índia | 34.861.579 | 481.486 |
| Itália | 6.125.683 | 137.402 |
| Estados Unidos | 54.743.982 | 825.536 |

**Fonte: Base de dados da Universidade Johns Hopkins**

Foi desenvolvida uma função para converter uma única coluna de dados em um conjunto de dados de duas colunas: a primeira coluna contendo a contagem de dias, desde janeiro de 2020 até dezembro de 2021 e a segunda coluna contendo a contagem de casos confirmados/mortos por COVID19 dos próximos meses, a ser prevista (BROWLEE, 2020). Assim sendo, seguiu-se para a próxima etapa que é a implementação dos modelos preditores.

## 3.2 DESENVOLVENDO OS MODELOS PREDITORES ARIMA, LSTM, MLP E SUAVIZAÇÃO EXPONENCIAL

Para desenvolver os modelos preditivos ARIMA, LSTM, MLP e Suavização Exponencial, foi necessário avaliar o funcionamento de cada um deles e como eles são implementados. As sessões a seguir descrevem como foi realizada a implementação desses modelos, começando pelo método de predição.

### 3.2.1 Método de Predição

O método de predição foi dividido em duas fases. Na primeira fase, realizaram-se o treinamento dos modelos preditivos. Na segunda fase ocorreu a fase de teste com o uso de dados que não foram utilizados na etapa de treinamento, onde realizou-se a predição dos dados de cada um dos modelos. Por fim, as predições são avaliadas seguindo a mesma métrica de análise de desempenho, sendo essa a raiz do erro quadrático médio (RMSE), encerrando com a comparação dos resultados dos modelos.

Inicialmente, as predições foram realizadas para os casos confirmados de COVID-19 no Brasil, Índia, Itália e EUA, posteriormente para as mortes, nos mesmos países. Esses quatro foram selecionados para que pudesse haver uma comparação de desempenho dos modelos em diferentes países.

Com dados de séries temporais, a sequência de valores é importante. Depois de modelar os dados e estimar a habilidade do modelo no conjunto de dados de treinamento, é necessário ter uma ideia da habilidade do modelo. O método utilizado

foi dividir o conjunto de dados ordenado em conjuntos de dados de treinamento e teste (BROWLEE, 2020).

Como mencionado anteriormente, os dados de entrada foram obtidos por meio da base de dados da Universidade Johns Hopkins, que conta com dados de muitos países. Após limpar e tratar os dados de cada um dos países separadamente, eles ficaram com duas colunas, sendo a primeira coluna as datas de todos os dias, de janeiro de 2020 até dezembro de 2021, dos quatro países avaliados. Além dessa, a segunda coluna contém a contagem acumulativa de casos confirmados e mortes, sendo essas as variáveis de entrada de cada país, para os quatro modelos preditivos.

Ao desenvolver a implementação de um modelo de predição, uma boa prática é realizar a separação dos dados em conjuntos de dados de treinamento com 67% do total de dados adquiridos, sendo dados utilizados para treinar o modelo e deixando os 33% restantes para testar o modelo (BROWLEE, 2020). A ação descrita foi realizada em todos os quatro modelos deste trabalho.

Assim sendo, as sessões a seguir descrevem como foram realizadas as implementações dos modelos utilizados para a realização das predições.

### 3.2.2 Implementação das Redes Neurais LSTM e MLP

Como mencionado na fundamentação teórica, os dados utilizados são séries temporais de casos confirmados e mortes por COVID19, sendo esses as respectivas datas e os dados acumulativos. Ou seja, a rede MLP tem apenas x(t-1) na sua entrada e a rede LSTM também possui só uma entrada. Em ambas as redes a entrada é a quantidade acumulativa de mortes ou de casos confirmados.

A função de ativação de todos os neurônios da rede neural LSTM foi a “linear”, com o número de células LSTM que variam de 200 a 600. Contudo, para a configuração da função de ativação da rede MLP, foi utilizada a função de ativação “Unidade Linear Retificada” (relu) para todos os neurônios. Nesta mesma rede foram testadas combinações de números de neurônios por camada (três camadas) de: 8, 16, 32 e 64.

Portanto, a seguir, será discutido acerca da técnica *Grid Search* utilizada nos modelos LSMT e MLP.

#### 3.2.2.1Grid Search nas redes neurais LSTM e MLP

Para decidir quais foram os melhores valores de hiperparâmetros testados nos modelos das Redes Neurais LSTM e MLP, foi realizada uma *Grid Search* (grade de experimentos), modificando os valores dos hiperparâmetros, mencionados anteriormente. Os parâmetros são a taxa de aprendizagem, valor do *dropout*, número de células e número de neurônios por camadas das redes neurais utilizadas. A Tabela 2, a seguir, mostra quais valores foram experimentados.

**Tabela 2 – Grid Search**

| Taxa de aprendizagem | Valor do dropout | Número de células da Rede Neural LSTM | Número de neurônios por camada da Rede Neural MLP |
|---|---|---|---|
| 0.01 | 0.01 | 200 | 8 |
| 0.05 | 0.05 | 300 | 16 |
| 0.1 | 0.1 | 400 | 32 |
| 0.2 | 0.2 | 600 | 64 |

**Fonte: Autor**

Assim sendo, todos os valores da tabela de *Grid Search* foram testados, para que pudesse ser avaliado qual combinação teve o melhor desempenho. São 64 possibilidades e para garantir que todas essas fossem utilizadas, foram desenvolvidos laços de repetição onde todos os valores da Tabela 2 fossem utilizados nas predições realizadas, para os dados de casos confirmados e mortes por COVID-19. Como a construção dos modelos ARIMA e de Suavização Exponencial é diferente do LSTM e MLP, essa tabela não se aplica a eles.

Por fim, para a avaliação de qual foi a melhor combinação de valores dos hiperparâmetros, foi implementada no código uma condição para escolher a predição que obteve o menor RMSE. Após 200 épocas, por meio da condição mencionada, os dados foram armazenados para que pudesse ser avaliado qual foi a melhor configuração, no final da implementação do modelo. Deste modo, escolheu-se a combinação que alcançou os melhores dados de predição.

Por fim, a avaliação de resultados é aplicada de modo que dos modelos LSTM MLP possuem condições que analisam se o desempenho da predição avaliada foi superior ao anterior. Caso essa condição seja satisfeita, é realizado o armazenamento de todos os dados de resultados, hiperparâmetros, RMSE, entre outros que

culminaram naquele melhor resultado. Desse modo, foram analisados quais os melhores resultados e quais seus hiperparâmetros.

### 3.2.3 Implementação do modelo ARIMA

Retomando o que foi discutido na sessão 2.5, os estágios da metodologia Box e Jenkins compreendem: a identificação, a estimação e a verificação da adequação do modelo ARIMA com base nos dados que serão preditos (JENKINS, 1976). Ao realizar essas etapas, iniciou-se com a identificação e estimação da melhor sequência de parâmetros (p,d,q). O pacote “pmdarima” foi utilizado para o processo de “*auto-ARIMA”*, que com base nos dados, busca identificar os parâmetros mais adequados para o modelo (JUNIOR LIRA, 2017). Por fim, realizou-se a verificação das predições obtidas, que serão melhor discutidas na análise de resultados.

O código de implementação do modelo ARIMA foi criado usando uma biblioteca importada do “Statsmodels”. Esse é um módulo Python que fornece classes e funções para a estimativa de muitos modelos diferentes, bem como para a realização de testes e exploração de dados estatísticos (PERKTOLD *et al*., 2019). Desse modo, a predição dos dados de treino e teste foram realizadas de forma simples, com base nos melhores parâmetros (p,d,q) avaliados anteriormente.

Vale ressaltar que no modelo ARIMA não serão utilizados hiperparâmetros, pois esse modelo requer os parâmetros (p,d,q), já expostos na fundamentação teórica. Deve-se enfatizar que os modelos LSTM e MLP não exigem a configuração dos parâmetros (p,d,q). Todavia, eles utilizam hiperparâmetros que permitem mais ajustes na predição do modelo.

### 3.2.4 Implementação do modelo de Suavização Exponencial

Para que fosse realizada uma análise minuciosa do desempenho do modelo de Suavização Exponencial (*Exponential Smothing*), neste trabalho foram implementados os cinco principais métodos deste modelo. Os métodos são: Algoritmo de suavização Simples, Holt, Holt-Winters (aditivo e multiplicativo) e Pegels.

O modelo de Suavização Exponencial é de fácil implementação e não conta com hiperparâmetros. Por conta de sua simplicidade, foi possível implementar

facilmente os 5 métodos e obter os resultados de predição para os casos confirmados e mortes por COVID-19 dos países avaliados neste trabalho. Os resultados serão abordados posteriormente.

Entretanto, na implementação do modelo de Suavização Exponencial com a biblioteca “*ExponentialSmoothing*”, há a limitação de não possibilitar a obtenção dos resultados do desempenho do modelo na etapa de treino, somente na de teste. A documentação da biblioteca não explica o motivo disso, só instrui como realizar a etapa de treino e teste, mas há somente a instrução da visualização dos resultados de teste (PERKTOLD *et al.*, 2019).

Portanto, com estes códigos desenvolvidos para a implementação dos modelos discutidos ao longo desse desenvolvimento é possível gerar predições usando os modelos para o conjunto de dados para estimar o desempenho desses por meio do cálculo da raiz quadrada do erro médio quadrático (RMSE), onde são selecionados os melhores desempenhos dos modelos. Os resultados obtidos dos modelos com esses dados serão melhor discutidos na análise dos resultados, a seguir.

# ANÁLISE DE RESULTADOS

Como foram avaliadas informações de dois conjuntos de dados, sendo estes casos confirmados e mortes por COVID-19, de quatro modelos preditivos, a análise de resultados está dividida em quatro tópicos, sendo esses os resultados obtidos com a predição com base nos dados. Desse modo, foram selecionados os 32 melhores resultados, sendo dividido metade para casos confirmados e a outra metade para mortes, dos quatro países analisados.

No entanto, como são muitos gráficos de resultados de predições, selecionou-se os gráficos do Brasil para exposição na análise de resultados. Os demais gráficos com as descrições estão no “Anexo A”, no final deste trabalho. Essa decisão foi necessária para que a dissertação não ficasse com muitas páginas de imagens no capítulo de análise de resultados. Porém, os dados mais importantes, sendo esses RMSE, hiperparâmetros e configurações, que obtiveram melhor desempenho, encontram-se nesta análise. Portanto, a seguir serão descritos os detalhes dos resultados obtidos.

## 3.3 RESULTADOS DOS MODELOS DE REDES NEURAIS LSTM

Para uma visualização mais analítica, temos os gráficos de predições de treino e de teste dos países avaliados. Em cada uma das figuras das predições do modelo LSTM, a linha vermelha representa os dados reais e a linha azul a predição. Na Figura 8, com dados do Brasil, é possível observar que houve um grande aumento de casos, a partir do 300º dia. Aparentemente, os dados preditos ficaram bem correspondentes com os dados reais, mostrando que o modelo teve um bom desempenho no treino e teste. No entanto, o que poderá comprovar a eficiência dos modelos analisados será o resultado do RMSE.

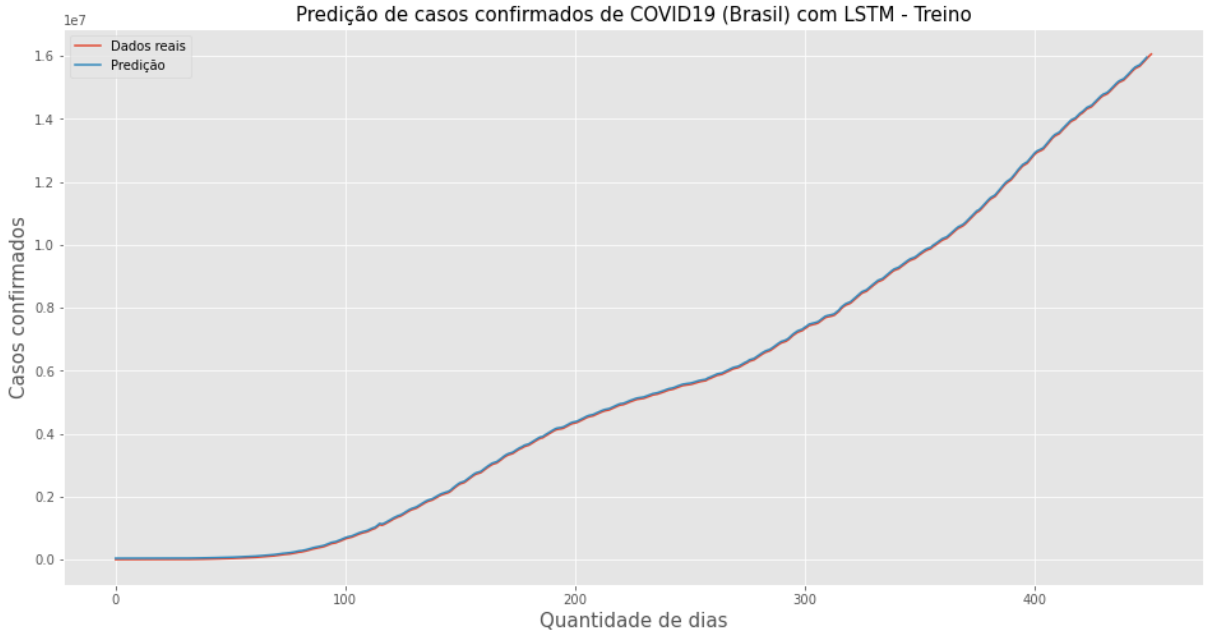


**Figura 8 – Resultado das predições de casos confirmados do Brasil com LSTM - Treino**
**Fonte: Autor**

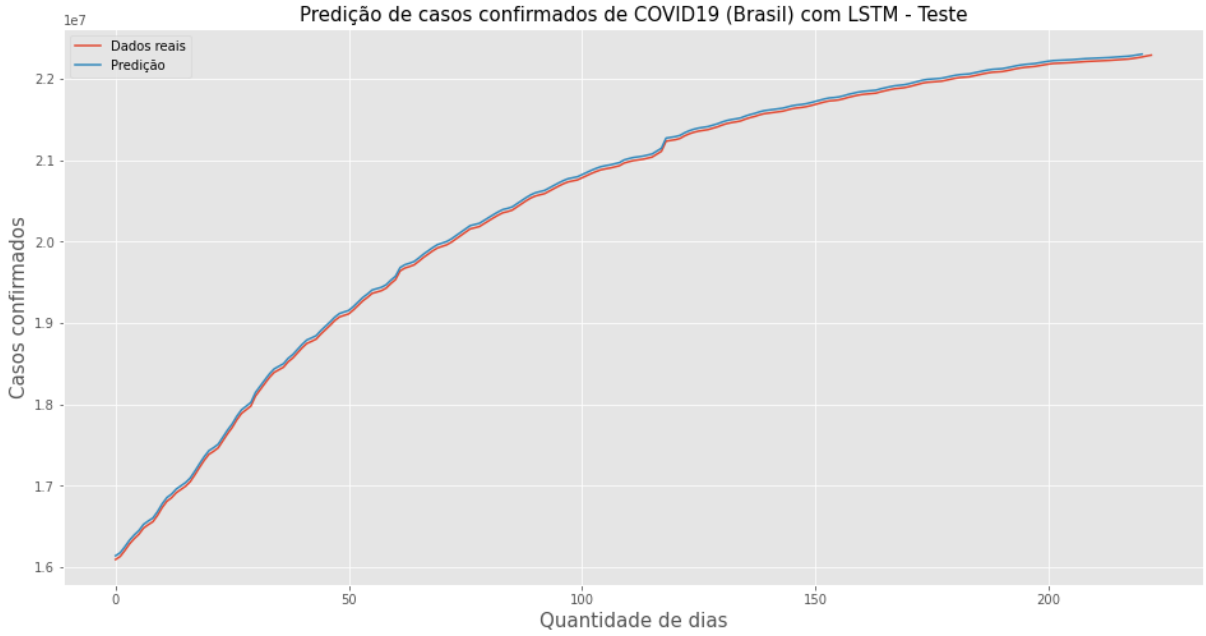


**Figura 9 – Resultado das predições de casos confirmados do Brasil com LSTM - Teste**
**Fonte: Autor**

Ao visualizarmos os gráficos, é notório que cada gráfico tem suas características diferentes dos demais, isso ocorre nas duas etapas de cada país. Nas Figuras 10 e 11, as predições dos dados de mortes no Brasil parecem próximas dos dados reais. Entretanto, vale ressaltar, que a comprovação de bons resultados de um

modelo se dá por meio da avalição de desempenho, que no caso deste trabalho, é a raiz do erro quadrático médio (RMSE).

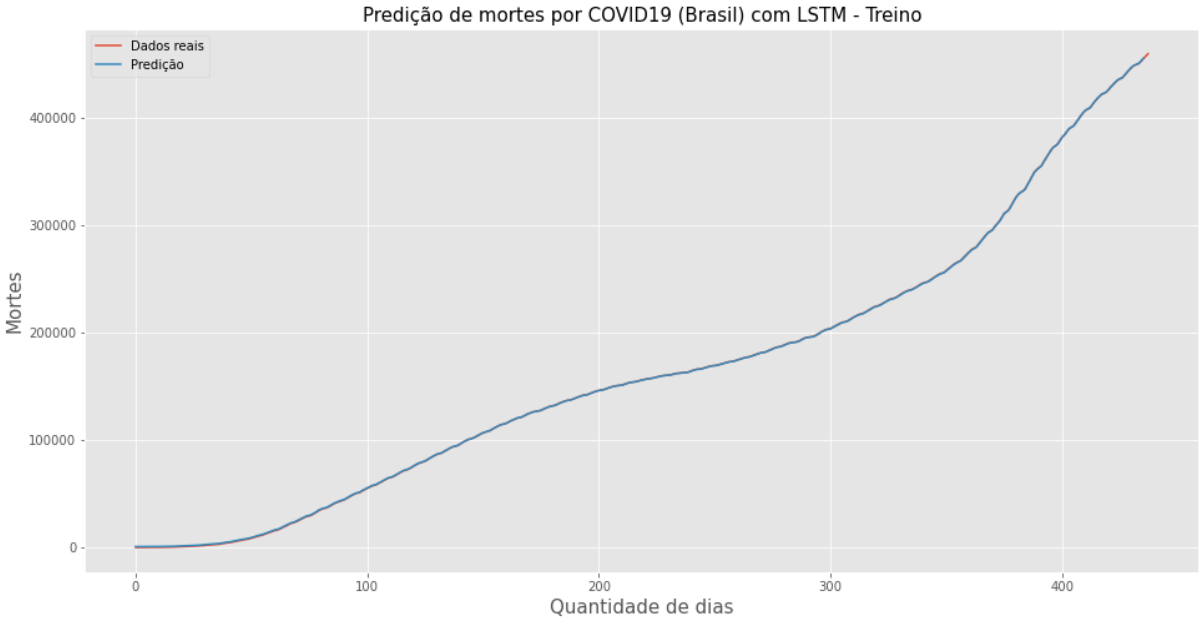


**Figura 10 – Resultado das predições de treino de mortes do Brasil com LSTM**
**Fonte: Autor**

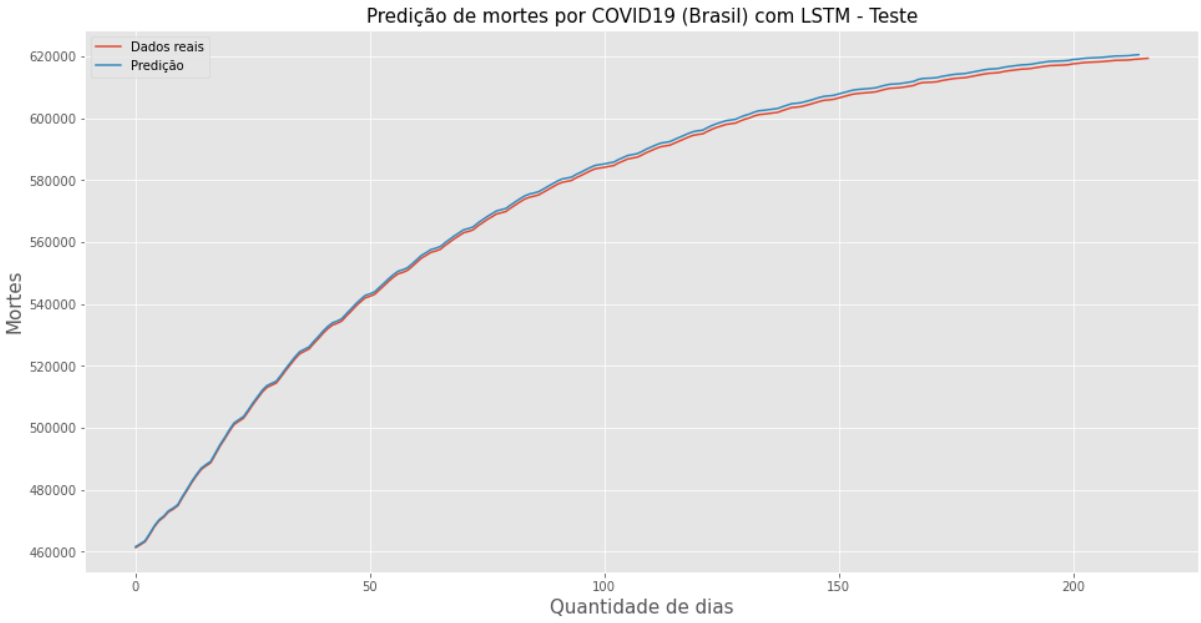


**Figura 11 – Resultado das predições de teste de mortes do Brasil com LSTM**
**Fonte: Autor**

Conforme demonstrado nos gráficos de predições de casos confirmados e mortes do Brasil e dos demais países presentes no Anexo A, é possível observar que as predições realizadas pelos modelos estão parecidas dos dados reais obtidos. As Tabelas 3 e 4 descrevem os erros médios quadráticos de cada um dos modelos, na

etapa de Treino e de Teste. É possível observar que o modelo LSTM teve um desempenho de treino próximo com o de teste. No entanto, os resultados na etapa de teste foram melhores.

| **RMSE de casos confirmados do modelo LSTM** | | | | |
|---|---|---|---|---|
| Scores | ***Brasil*** | ***Índia*** | ***Itália*** | ***EUA*** |
| Treino | 24.079,53 casos | 33.381,28 casos | 7.358,22 casos | 82.928,35 casos |
| Teste | 25.283,47 casos | 14.0352,03 casos | 4.557,34 casos | 68.993,10 casos |

**Tabela 3 – Resultados obtidos de casos confirmados de COVID19 - LSTM**
**Fonte: Autor**

| **RMSE de mortes do modelo LSTM** | | | | |
|---|---|---|---|---|
| Scores | ***Brasil*** | ***Índia*** | ***Itália*** | ***EUA*** |
| Treino | 1.375,84 mortes | 729,42 mortes | 297,45 mortes | 1.383,28 mortes |
| Teste | 986,82 mortes | 9.482,91 mortes | 53,18 mortes | 826,46 mortes |

**Tabela 4 – Resultados obtidos de mortes por COVID19 - LSTM**
**Fonte: Autor**

Os valores de raiz do erro médio quadrático médio parecem ser significativos, porém são milhões de casos e milhares de mortes, como é exposto na Tabela 1, sendo estes erros aceitáveis perto do valor total dos dados (AZANK, 2020). Assim sendo, há uma taxa de assertividade considerável. Contudo, para uma melhor visualização e compreensão, as Tabelas 5 e 6 expõem os valores dos hiperparâmetros do modelo.

| **Hiperparâmetros** | ***Brasil*** | ***Índia*** | ***Itália*** | ***EUA*** |
|---|---|---|---|---|
| Taxa de aprendizagem | 0.2 | 0.01 | 0.1 | 0.2 |
| Valor de Dropout | 0.1 | 0.05 | 0.05 | 0.2 |
| Número de Células LSTM | 600 | 600 | 600 | 400 |

**Tabela 5 – Hiperparâmetros do modelo LSTM de casos confirmados de COVID19**
**Fonte: Autor**

| **Hiperparâmetros** | ***Brasil*** | ***Índia*** | ***Itália*** | ***EUA*** |
|---|---|---|---|---|
| Taxa de aprendizagem | 0.2 | 0.05 | 0.2 | 0.01 |
| Valor de Dropout | 0.05 | 0.05 | 0.1 | 0.1 |
| Número de Células LSTM | 400 | 600 | 400 | 200 |

**Tabela 6 – Hiperparâmetros do modelo LSTM de mortes por COVID19**
**Fonte: Autor**

Pode-se observar que cada série tem as suas características próprias, o que exige um conjunto de hiperparâmetros diferentes para cada uma delas. Todos os países tiveram a combinação de hiperparâmetros desiguais, comprovando assim, que é de extrema eficácia a utilização de uma *Grid Search* para obtenção do melhor resultado.

## 3.4 RESULTADOS DOS MODELOS DE REDES NEURAIS MLP

A demonstração dos resultados obtidos pelo modelo de redes neurais MLP segue a mesma lógica de exposição do modelo LSTM. Serão expostos os gráficos de predições de treino, teste e posteriormente, as tabelas contendo os resultados e hiperparâmetros do modelo.

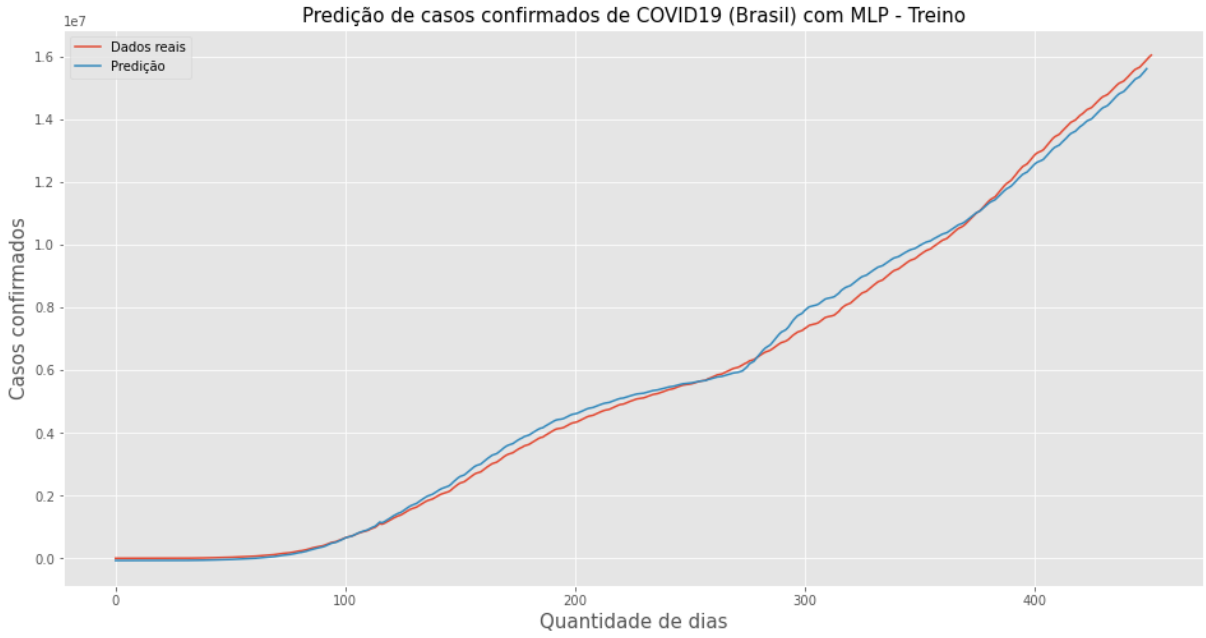


**Figura 12 – Resultado da predição de treino de casos confirmados do Brasil com MLP**
**Fonte: Autor**

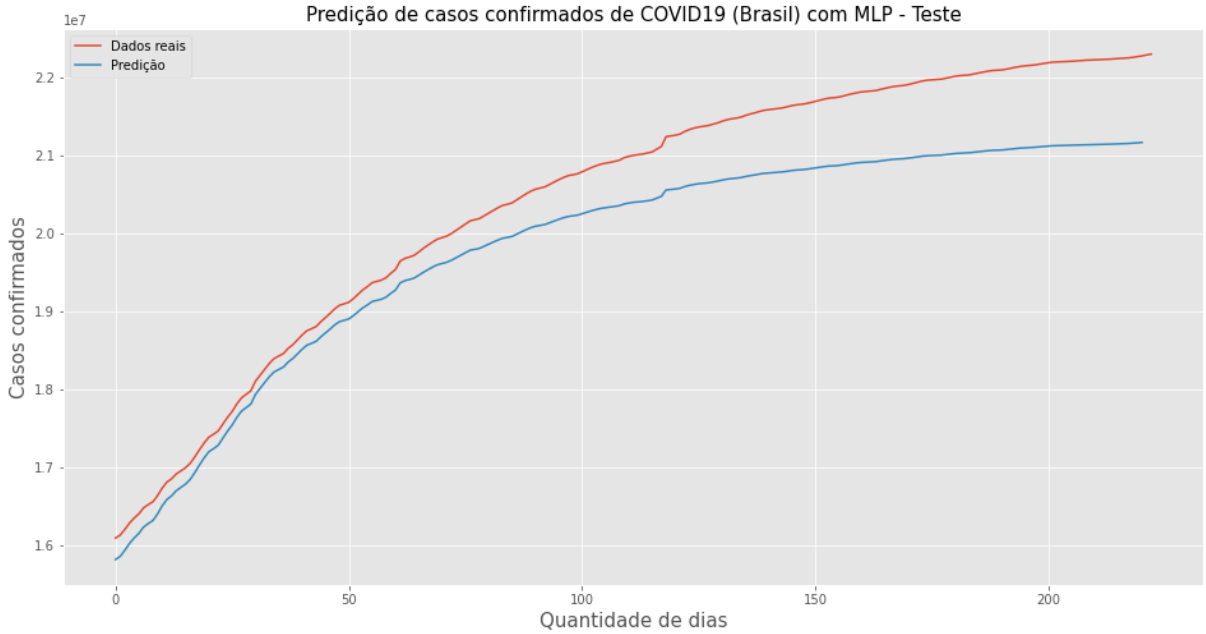


**Figura 13 – Resultado da predição de teste de casos confirmados do Brasil com MLP**
**Fonte: Autor**

Comparado com o modelo LSTM, as predições do modelo MLP visualmente parecem com um desempenho inferior nos gráficos, conforme é visto nas Figuras de 12 a 15 e demais no Anexo A. Entretanto, os resultados da avaliação de desempenho mostrarão de forma mais precisa a comparação entre os modelos.

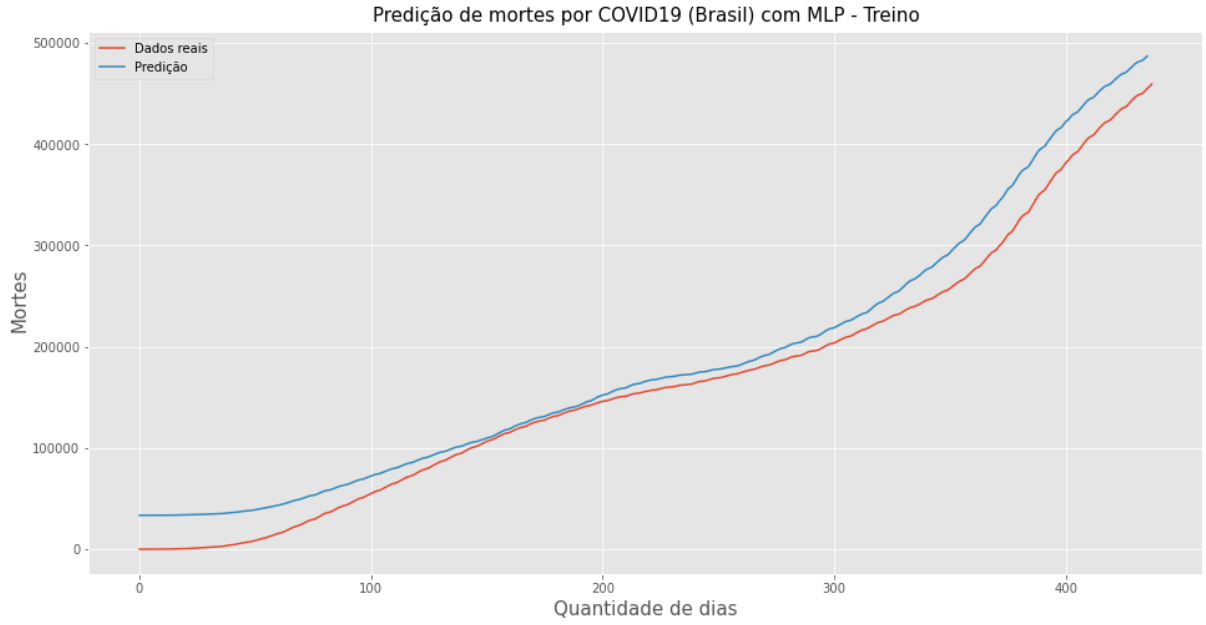


**Figura 14 – Resultado da predição de treino de mortes do Brasil com MLP**
**Fonte: Autor**

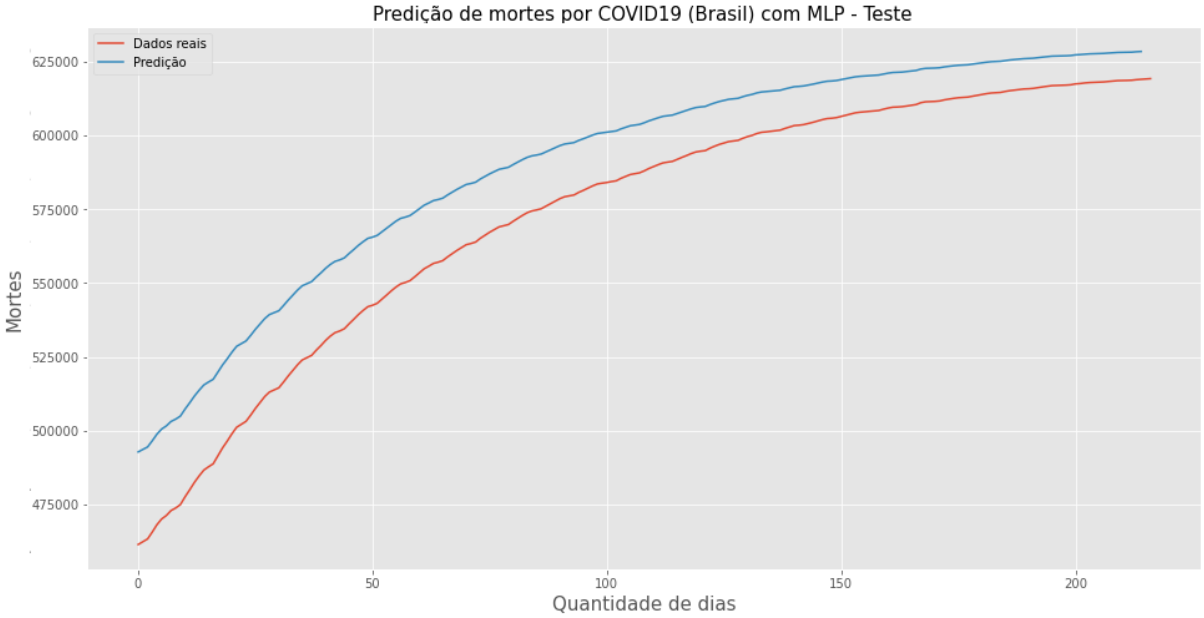


**Figura 15 – Resultado da predição de teste de mortes do Brasil com MLP**
**Fonte: Autor**

Mediante aos resultados expostos, tem-se os gráficos de predições de casos confirmados e mortes do Brasil (demais países no Anexo A). Em todos os gráficos é notório que o modelo MLP, mesmo com a técnica de *Grid Search*, não obteve resultados de predições próximos dos reais.

Do mesmo modo como os resultados dos modelos de LSTM, as Tabelas 7 e 8 descrevem os erros médios quadráticos de cada um dos modelos, na etapa de Treino e de Teste.

| **RMSE de casos confirmados do modelo MLP** | | | | |
|---|---|---|---|---|
| Scores | ***Brasil*** | ***Índia*** | ***Itália*** | ***EUA*** |
| Treino | 956.955,99 casos | 567.785,85 casos | 64.537,91 casos | 280.513,35 casos |
| Teste | 728.409,38 casos | 778.756,96 casos | 123.301,96 casos | 4.690.290,78 casos |

**Tabela 7 – Resultados obtidos de casos confirmados de COVID19 - MLP**
**Fonte: Autor**

| **RMSE de mortes do modelo MLP** | | | | |
|---|---|---|---|---|
| Scores | ***Brasil*** | ***Índia*** | ***Itália*** | ***EUA*** |
| Treino | 18.340,55 mortes | 8.527,79 mortes | 3.080,88 mortes | 21.627,85 mortes |
| Teste | 29.593,56 mortes | 42.296,30 mortes | 2.003,11 mortes | 43.273,58 mortes |

**Tabela 8 – Resultados obtidos de mortes por COVID19 - LSTM**
**Fonte: Autor**

Abaixo, as Tabelas 9 e 10 mostram os hiperparâmetros dos melhores resultados obtidos pelos modelos de Redes Neurais MLP. Diferente dos hiperparâmetros do modelo LSTM, o modelo MLP conta com o número de neurônios por camada. Apesar de alguns parecerem semelhantes, não houve combinação de hiperparâmetros iguais em nenhum dos países.

| **Hiperparâmetros** | ***Brasil*** | ***Índia*** | ***Itália*** | ***EUA*** |
|---|---|---|---|---|
| Taxa de aprendizagem | 0.1 | 0.05 | 0.01 | 0.01 |
| Valor de Dropout | 0.2 | 0.1 | 0.01 | 0.01 |
| Número de neurônios por camada | 8 | 16 | 32 | 64 |

**Tabela 9 – Hiperparâmetros do modelo MLP de casos confirmados de COVID19**
**Fonte: Autor**

| **Hiperparâmetros** | ***Brasil*** | ***Índia*** | ***Itália*** | ***EUA*** |
|---|---|---|---|---|
| Taxa de aprendizagem | 0.05 | 0.2 | 0.05 | 0.01 |
| Valor de Dropout | 0.2 | 0.2 | 0.01 | 0.05 |
| Nº de neurônios por camada | 8 | 8 | 32 | 64 |

**Tabela 10 – Hiperparâmetros do modelo MLP de mortes por COVID19**
**Fonte: Autor**

Em comparação com os resultados obtidos anteriormente, é evidente que os valores de RMSE são bem maiores dos obtidos pela rede neural LSTM. Desse modo, mostrando o modelo MLP pouco eficaz para a predição dos dados da COVID-19.

## 3.5 RESULTADOS DOS MODELOS ARIMA

Como mencionado no desenvolvimento, a arquitetura do modelo ARIMA é diferente dos modelos LSTM e MLP. Por conta disso, não há uma tabela de hiperparâmetros nos resultados, pois esses não são utilizados na construção do modelo. Porém, os parâmetros (p,d,q) foram acrescentados nas tabelas de resultados.

Assim sendo, abaixo estão os gráficos para a visualização dos resultados dos treinos e testes. Nas figuras dos gráficos, a linha vermelha são os dados preditos e a

linha azul os dados reais. Além desses, os resultados obtidos nos testes da predição, juntamente com os parâmetros estão nas Tabelas 11 e 12.

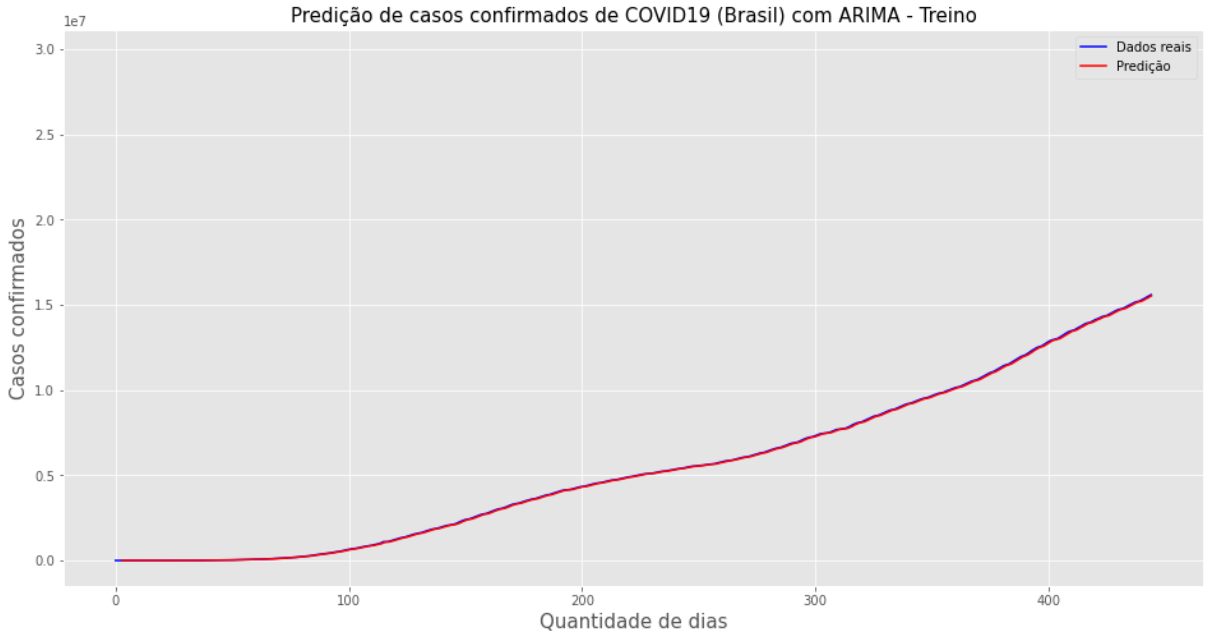


**Figura 16 – Resultado da predição de treino de casos confirmados do Brasil com ARIMA**
**Fonte: Autor**

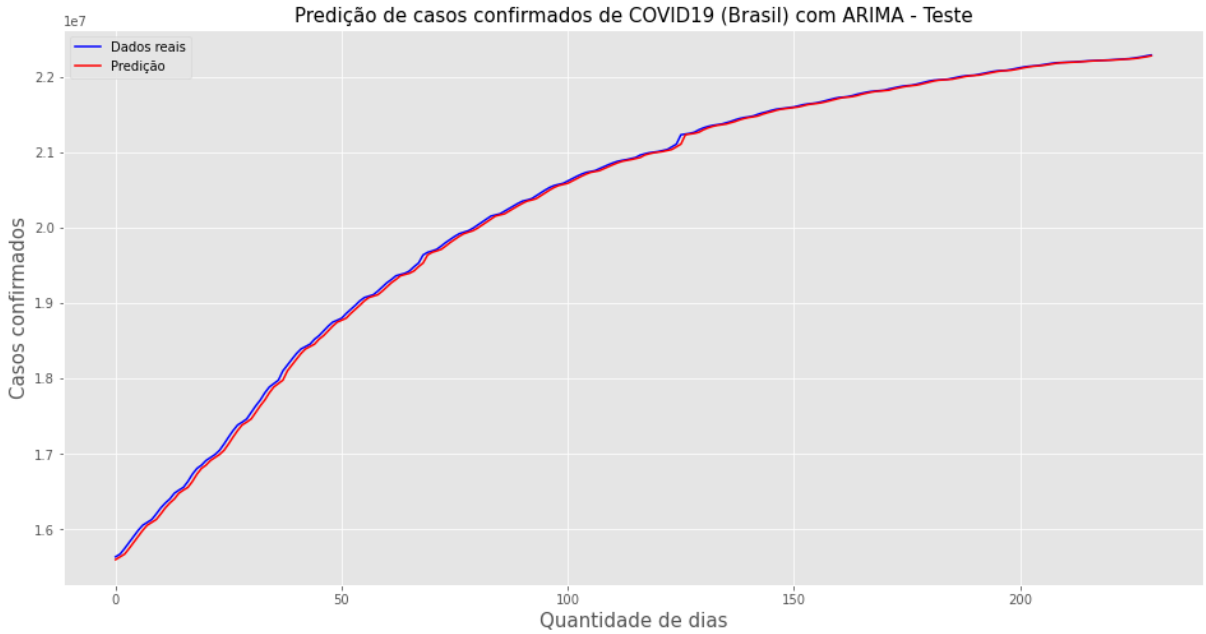


**Figura 17 – Resultado da predição de teste de casos confirmados do Brasil com ARIMA**
**Fonte: Autor**

Ao analisar os gráficos, nota-se que aparentemente os resultados de predições são similares aos dados originais. Dentre os modelos analisados neste

trabalho, os gráficos do modelo (demais no ANEXO A) ARIMA foram o que apresentaram melhor aproximação dos dados reais.

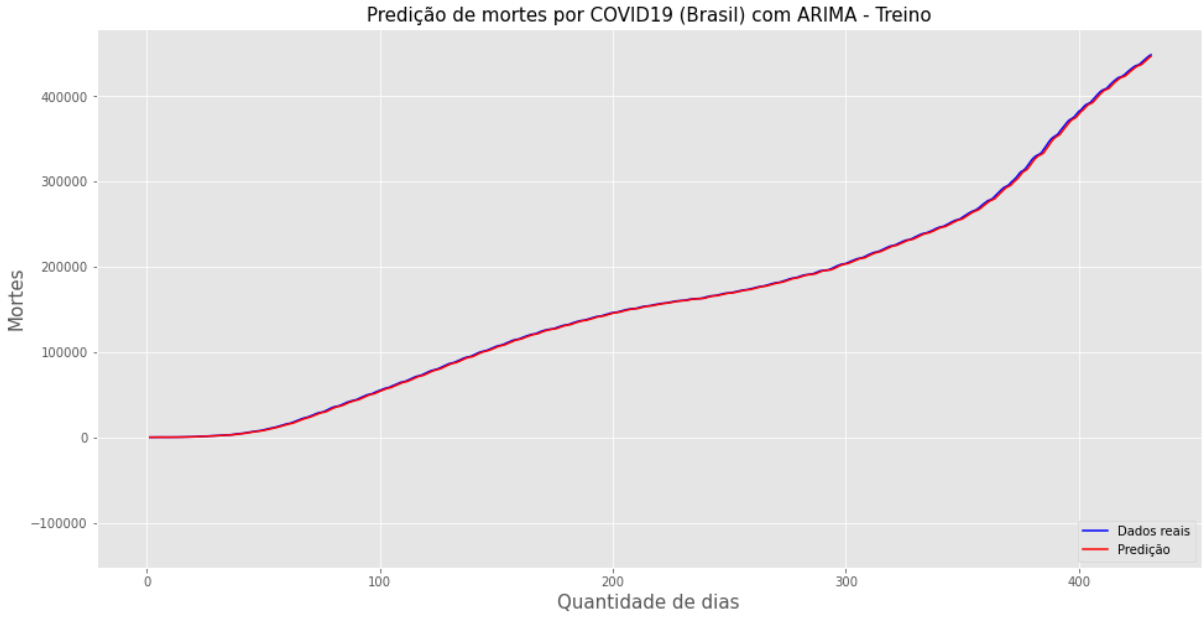


**Figura 18 – Resultado da predição de treino de mortes do Brasil com ARIMA**
**Fonte: Autor**

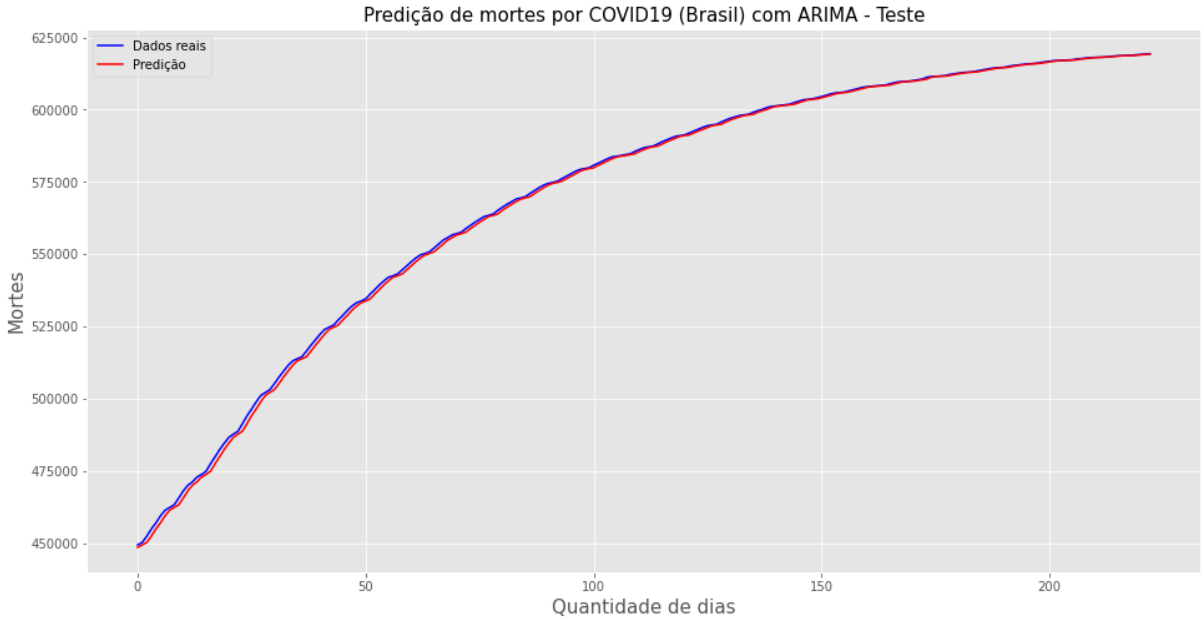


**Figura 19 – Resultado da predição de teste de mortes do Brasil com ARIMA**
**Fonte: Autor**

Os RMSE obtidos pelo modelo ARIMA foram os menores em comparação com os dos outros modelos. Alguns dos fatores que podem ter influenciado no melhor resultado são que o modelo é simples de implementar, sem ajuste de hiperparâmetro,

mais fácil de lidar com dados multivariáveis e de rápido execução (VASCONSELLOS, 2018).

| **RMSE e parâmetros de casos confirmados do modelo ARIMA** | | | | |
|---|---|---|---|---|
| Resultados | ***Brasil*** | ***Índia*** | ***Itália*** | ***EUA*** |
| RMSE de Treino | 20.846,28 casos | 11.751,93 casos | 1.698,52 casos | 43.582,58 casos |
| RMSE de Teste | 21.333,27 casos | 13.221,05 casos | 1.863,12 casos | 46.495,07 casos |
| Parâmetros (p,d,q) | (2,1,2) | (3,1,0) | (1,1,1) | (1,1,3) |

**Tabela 11 – Resultados obtidos de casos confirmados de COVID19 - ARIMA**
**Fonte: Autor**

| **RMSE e parâmetros de mortes do modelo ARIMA** | | | | |
|---|---|---|---|---|
| Resultados | ***Brasil*** | ***Índia*** | ***Itália*** | ***EUA*** |
| RMSE de Treino | 593,62 mortes | 536,27 mortes | 47,52 mortes | 559,58 mortes |
| RMSE de Teste | 646,42 mortes | 551,93 mortes | 51,52 mortes | 577,58 mortes |
| Parâmetros (p,d,q) | (3,1,0) | (1,1,2) | (1,1,2) | (1,1,2) |

**Tabela 12 – Resultados obtidos de mortes por COVID19 - ARIMA**
**Fonte: Autor**

Chatterjee (2020) comprova que ao comparar o desempenho de outros modelos algorítmicos disponíveis para séries temporais sobre epidemias, verificou-se que esses métodos de aprendizado de máquina foram todos superados por métodos clássicos simples, onde o modelo ARIMA teve o melhor desempenho geral para esse tipo de série temporal (séries sobre epidemias). Concluindo, desse modo, que os resultados obtidos neste trabalho não são diferentes de outras pesquisas realizadas outrora.

## 3.6 RESULTADOS DOS MODELOS DE SUAVIZAÇÃO EXPONENCIAL

Distintos dos outros resultados expostos anteriormente, nestes gráficos das figuras a seguir, temos a linha ciano que mostra os dados reais e as demais cores, os métodos de cada um dos cinco métodos utilizados na predição do modelo Suavização Exponencial. Sendo o Algoritmo de suavização Simples (Método 1), Holt (Método 2),

Holt-Winters aditivo (Método 3), Holt-Winter multiplicativo (Método 4) e Pegels (Método 5).

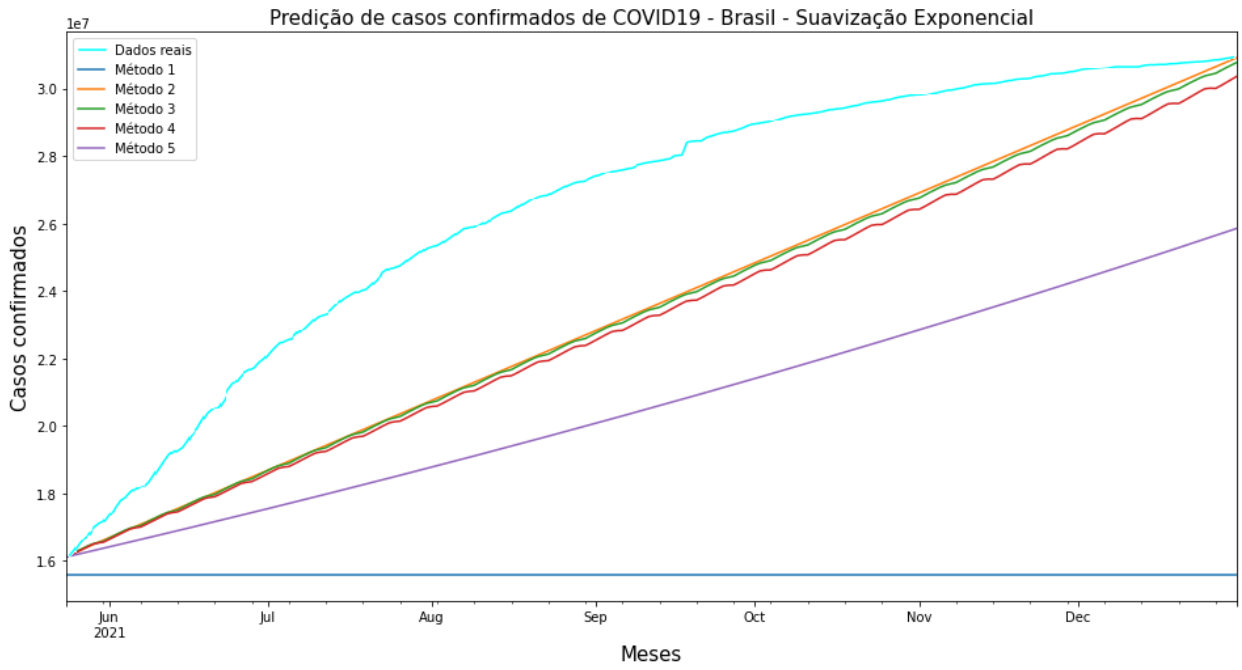


**Figura 20 – Resultado das predições de casos confirmados do Brasil com Suavização Exponencial**
**Fonte: Autor**

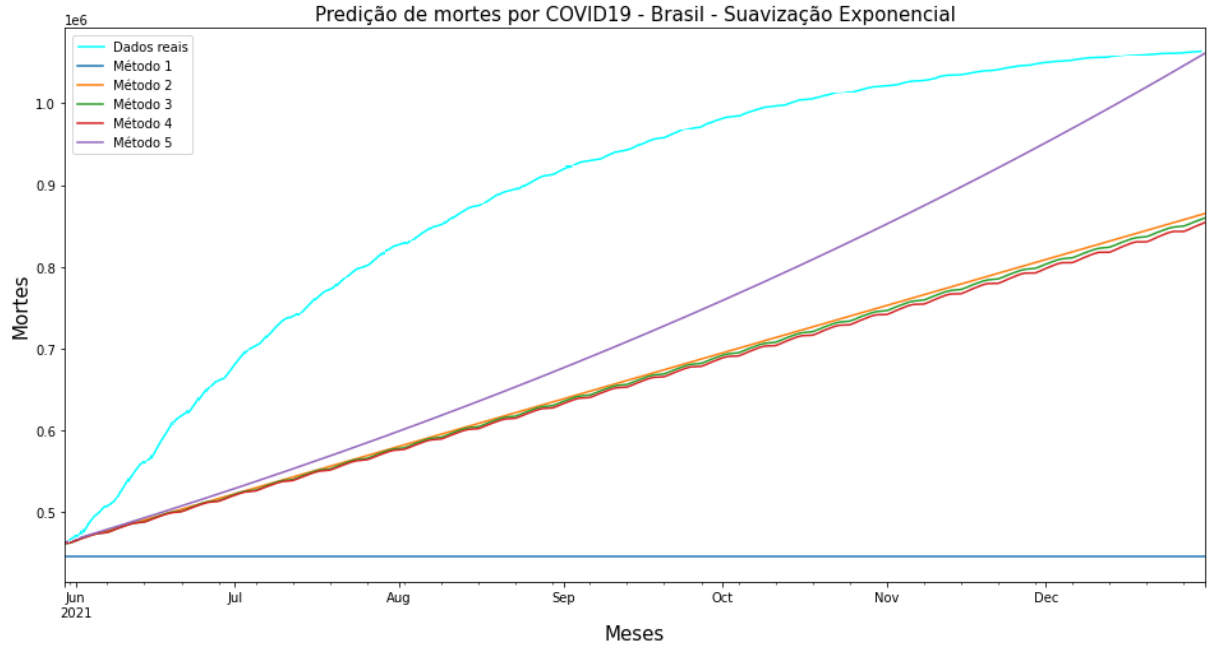


**Figura 21 – Resultado das predições de mortes do Brasil com Suavização Exponencial**
**Fonte: Autor**

Como mencionado no desenvolvimento, só é permitido avaliar o desempenho do modelo de Suavização Exponencial na etapa de treino. Nas Figuras 20 e 21,

observa-se que cada método de suavização teve um desempenho diferente de lidar com os dados reais. Isso ocorre, porque cada método tem suas particularidades, como mencionado na fundamentação teórica.

Nas Tabelas 13 e 14, reúnem-se os resultados de RMSE, na etapa na etapa de teste, de casos confirmados e mortes por COVID19, do modelo de Suavização Exponencial, de todos os cinco métodos utilizados na predição dos dados. Como são muitos resultados, para uma melhor análise, os valores em vermelho são os menores RMSE de cada um dos países analisados.

| **RMSE de casos confirmados dos modelos Suavização Exponencial** | | | | |
|---|---|---|---|---|
| **Algoritmos de predição exponencial** | ***Brasil*** | ***Índia*** | ***Itália*** | ***EUA*** |
| Suavização Simples | 5.880.239,90 casos | 1.6833.914,10 casos | 797.103,86 casos | 6.159.386,40 casos |
| Suavização de Holt | 1.628.307,20 casos | 6.223.606,10 casos | 396.897,45 casos | 2.728.253,96 casos |
| Suavização de Holt-Winters aditivo | 1.581.421,80 casos | 6.546.642,90 casos | 181.105,57 casos | 3.032.557,42 casos |
| Suavização de Holt-Winters multiplicativo | 1.593.886,01 casos | 5.641.418,50 casos | 207.596,76 casos | 2.545.500,11 casos |
| Suavização Pegels aditivo | 4.939.944,50 casos | 34.106.846,40 casos | 319.838,01 casos | 3.602.089,00 casos |

**Tabela 13 – Resultados obtidos de casos confirmados de COVID19 – Suavização exponencial**
**Fonte: Autor**

| **RMSE de mortes dos modelos Suavização Exponencial** | | | | |
|---|---|---|---|---|
| **Algoritmos de predição exponencial** | ***Brasil*** | ***Índia*** | ***Itália*** | ***EUA*** |
| Suavização Simples | 176.682,28 mortes | 219.817,45 mortes | 13.013,19 mortes | 66.938,78 mortes |
| Suavização de Holt | 113.766,00 mortes | 53.232,66 mortes | 23.800,31 mortes | 1.5087,67 mortes |
| Suavização de Holt-Winters aditivo | 117.798,57 mortes | 57.778,78 mortes | 22.529,16 mortes | 31.064,54 mortes |
| Suavização de Holt-Winters multiplicativo | 133.004,34 mortes | 61.272,79 mortes | 28.860,81 mortes | 22.057,97 mortes |
| Suavização Pegels aditivo | 374.998,51 mortes | 170.448,87 mortes | 35.354,41 mortes | 30.945,12 mortes |

**Tabela 14 – Resultados obtidos de mortes por COVID19 - Suavização exponencial**
**Fonte: Autor**

Os RMSE obtidos pelo modelo de Suavização Exponencial foram os maiores dentre todos os modelos testados no desenvolvimento. Portanto, com base nestes

dados, é visível que os modelos que apresentaram os menores resultados são o ARIMA e o LSTM. Essa comparação será melhor descrita na sessão a seguir.

## 3.7 COMPARANDO RESULTADOS DOS MODELOS

Para uma melhor compreensão dos dados, temos três tabelas. A Tabela 15 expõe os valores RMSE de teste de todos os modelos dos quatro países avaliados. No caso do modelo de Suavização Exponencial (SE), foram selecionados os menores valores de RMSE de cada modelo e país. Nas tabelas, os dados marcados em vermelho representam os menores valores.

**Tabela 15 – Comparação dos resultados de testes obtidos dos modelos de predição de teste de casos confirmados de COVID19**

| Países | RMSE ARIMA | RMSE LSTM | RMSE MLP | RMSE SE |
|---|---|---|---|---|
| Brasil | 21.333,27 | 25.283,71 | 728.409,38 | 1.581.421,80 |
| Índia | 13.221,05 | 140.352,03 | 778.756,96 | 5.641.418,50 |
| Itália | 1.863,12 | 4.557,34 | 123.301,96 | 181.105,57 |
| Estados Unidos | 46.495,07 | 68.993,10 | 4.690.290,78 | 2.545.500,11 |

**Fonte: Autor**

É possível notar que o modelo ARIMA obteve uma boa predição de teste em todos os países. No entanto, o modelo LSTM obteve um melhor desempenho comparado com os modelos MLP e de Suavização Exponencial. Por fim, o modelo MLP demonstrou um bom desempenho melhor que o de Suavização Exponencial.

**Tabela 16 – Comparação dos resultados de testes obtidos dos modelos de predição de mortes por COVID19**

| Países | RMSE ARIMA | RMSE LSTM | RMSE MLP | RMSE SE |
|---|---|---|---|---|
| Brasil | 646,42 | 986,77 | 29.593,56 | 113766.00 |
| Índia | 551,93 | 9482,46 | 42.296,30 | 53232.66 |
| Itália | 51,52 | 53,18 | 2.003,11 | 13013.19 |
| Estados Unidos | 577,58 | 826,46 | 43.273,58 | 15087.67 |

**Fonte: Autor**

A Tabela 16 segue a mesma premissa da Tabela 15, porém mostra os valores obtidos pelos modelos por meio da predição de mortes da doença. É notório que o

modelo ARIMA alcançou os menores erros em comparação aos demais modelos. Contudo, o modelo LSTM teve um desempenho bem próximo na predição de mortes dos países Itália e EUA. O MLP não teve um desempenho tão bom, porém o modelo de suavização foi inferior a todos os outros.

A Tabela 17 demonstra as médias dos resultados de RMSE, dos quatro países analisados, de cada modelo, em casos confirmados e mortes. Em primeiro lugar, o modelo ARIMA obteve as menores médias em todos os casos. Em segundo lugar, temos o modelo LSTM. Em terceiro, o modelo MLP e em último, o modelo de Suavização Exponencial. Esses resultados comprovam a eficácia do modelo ARIMA comparado com os outros modelos.

**Tabela 17 – Média de RMSE por modelo**

| Médias | ARIMA | LSTM | MLP | SE |
|---|---|---|---|---|
| Média de RMSE de casos confirmados | 20.728,13 | 59.796,54 | 1.580.189,77 | 2.487.361,50 |
| Média de RMSE de mortes | 456,87 | 2.651,21 | 29.291,63 | 48.774,88 |

**Fonte: Autor**

## 3.8 PUBLICAÇÃO DO TRABALHO NO CBIC 2021

Além dos resultados expostos anteriormente, destaca-se a publicação e apresentação de um artigo no “XV Congresso Brasileiro de Inteligência Computacional (CBIC) 2021”, onde o assunto da dissertação está, em parte, descrito na publicação (LEITE, 2021). Assim sendo, a aprovação do artigo no CBIC reforça a importância dos estudos realizados no desenvolvimento desse trabalho. O link do evento, com o artigo encontra-se no Anexo B, no final desta dissertação.

Abaixo, a Figura 22 mostra a apresentação do trabalho realizada pelo autor. Assim como o CBIC 2021, a apresentação ocorreu de forma online, em decorrência do afastamento social por conta da pandemia de COVID19.


**Figura 22 – Apresentação no CBIC 2021**
**Fonte: Autor**

# 4 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Resgatando o objetivo do presente trabalho, destaca-se que a principal contribuição que esta dissertação pretende trazer consiste na apresentação e na comparação de desempenhos de modelos preditores, com foco nas predições de séries temporais de casos confirmados e mortes por COVID19. Esse voltado para o auxílio à comunidade acadêmica, no combate à pandemia.

Pôde-se observar que os modelos conseguiram atingir o objetivo final de predizer com o mínimo de erro possível os dados da pandemia nos países: Brasil, Itália, Índia e EUA. Os valores de RMSE mostram que o melhor modelo de predição foi o ARIMA. Desse modo, sendo esse fortemente recomendado para a realização de predições futuras. Além disso, os maiores RMSE foram os do modelo de suavização Exponencial, não sendo recomendado a sua utilização.

No entanto, o modelo LSTM também obteve resultados significativos. Com isso, podendo ser utilizado como modelo de predição auxiliar, para realização de comparações de desempenho em alguns casos com relação ao ARIMA. Todavia, também pode ser utilizado junto com o ARIMA, formando um comitê de preditores.

Entretanto, é importante ressaltar que os modelos podem ser mais uma ferramenta que pode ser utilizada para auxiliar no combate da COVID19. Assim, conseguindo maximizar o bem-estar da sociedade que vem sofrendo em decorrência da pandemia.

A implementação desenvolvida neste trabalho pode servir como “molde” para outros países. O que se faz necessário é a aquisição e tratamento dos dados, para que possam ser aplicados nos modelos preditores mencionados nesta dissertação. Os métodos de adequação utilizaram as melhores configurações de modelos preditores, com base nos dados recebidos, independente do país. O link do código do trabalho encontra-se no Anexo B e está disponível no site “github.com” para possíveis reutilizações.

Portanto, como sugestões para trabalhos futuros, fica a possibilidade de avaliar o desempenho na predição de dados de outros países em torno do mundo, como países da Europa e do oriente, por exemplo. Desse modo, visando aprimorar o resultado e atingir maior assertividade nas predições, trazendo desta forma soluções tecnológicas cada vez melhores para a sociedade em todo o globo.

# REFERÊNCIAS

ABADI, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2015. Disponível em: <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45166.pdf>. Acesso em: 22 dez. 2021.

AKHATAR, M.; KRAEMER, M.U.G; GARDNER, L, M. Um modelo de rede neural dinâmica para prever o risco de Zika em tempo real. BMC Medicine. BMC Medicine, v. 17, n. 1, pág. 171, 2019.

AZANK, F. Como avaliar seu modelo de regressão. 2020. Disponível em: <https://medium.com/turing-talks/como-avaliar-seu-modelo-de-regress%C3%A3o-c2c8d73dab96>. Acesso em 09 jul. 2021.

BASTIANI, M. Aplicação de controle estatístico de processos e algoritmos de mineração de dados na gestão de frangos de corte. [Dissertação]. Medianeira: Universidade Tecnológica Federal do Paraná, 2017.

BENVENUTO, D.; GIOVANETTI, M.; VASSALLO, L.; ANGELETTI, S.; CICCOZZI, M. Aplicação do modelo ARIMA no conjunto de dados epidêmicos COVID-2019. Dados resumidos, v. 29, p. 105340, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S2352340920302341>. Acesso em 19 jun. 2021.

BINOTI, M. L. M. S. Rede neurais artificiais para prognose da produção de povoamentos não desbastados de eucalipto. [Dissertação]. Viçosa: Universidade Federal de Viçosa; 2010.

BOX, G. E. P.; JENKINS, G. Time Series Analysis Forecasting and Control. New Jersey: Holen Day, 1976.

BOX, G. E. P.; JENKINS, G.; REINSEL, G. C. Time Series Analysis. New Jersey: Prentice Hall, 1994.

BOX, G. E. P.; PIERCE, D. A. Distribuition of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the american statistical association, n. 64, p. 1509–1526, 1970.

BRANCO, H. Overfitting e underfitting em Machine Learning. 2020. Disponível em: <https://abracd.org/overfitting-e-underfitting-em-machine-learning/>. Acesso em 28 mai. 2021.

BROWN, R. G. (1959). Statistical forecasting for inventory control. Disponível em: <http://documents.irevues.inist.fr/handle/2042/28540>. Acesso em 14 mai. 2021.

BROWNLEE, J. Time Series Forecasting with the Long Short-Term Memory Network in Python. 2020. Disponível em: <https://machinelearningmastery.com/time-series-forecasting-long-short-term-memory-network-python>. Acesso em 05 mai. 2021.

BROWNLEE, J. Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras. 2020. Disponível em: <https://machinelearningmastery.com/time-series-forecasting-long-short-term-memory-network-python>. Acesso em 05 mai. 2021.

CAMILO, C. O.; SILVA, J. C. Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas. Disponível em: <http://www.inf.ufg.br/sites /default/files/ uploads/relatorios-tecnicos/RT-INF_001-09.pdf >. Acesso em: 21 jun. 2021.

CHATTERJEE, S. NARIMA/SARIMA vs LSTM with Ensemble learning Insights for Time Series Data. Data Science Central, 2020. Disponível em: <https://www.datasciencecentral.com/profiles/blogs/arima-sarima-vs-lstm-with-ensemble-learning-insights-for-time-ser>. Acesso em: 20 set. 2021.

CHAVES, A. N. Bootstrap em Séries Temporais. Tese (Doutorado) — PUC, Rio de Janeiro, 1991.

CHEN, N.; ZHOU, M.; DONG, X.; et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet, 2020.

CHENG, VCC; LAU, SKP; WOO, PCY; YUEN, KY Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection. Clinical Microbiology Reviews, v. 20, n. 4, pág. 660–694, 2007. Disponível em: <https://cmr.asm.org/content/20/4/660>. Acesso em 18 jun. 2021.

CHUA L.O.; YANG L. (1988). Cellular neural networks: theory. IEEE Transactions on Circuits and Systems, v. 35, n°. 10, p. 1257-1272.

CIO. Pandemia fez 43% dos negócios globais acelerarem uso de IA. 2021. Disponível em: <https://cio.com.br/noticias/pandemia-fez-43-dos-negocios-globais-acelerarem-uso-de-ia>. Acesso em: 22 abr. 2021.

DAS (DATA SCIENCE ACADEMY). Capítulo 51 – Arquitetura de Redes Neurais Long Short Term Memory (LSTM). 2021. Disponível em: <https://www.deeplearningbook.com.br/arquitetura-de-redes-neurais-long-short-term-memory/>. Acesso em: 22 dez. 2021.

ESTADÃO. Tudo sobre a Universidade Jhons Hopkins. 2021. Disponível em: < https://tudo-sobre.estadao.com.br/universidade-johns-hopkins>. Acesso em: 22 dez. 2021.

FACURE, Matheus. Disponível em: <https://matheusfacure.github.io//>. Acesso em: 12 nov. 2021.

FALBEL, D. Suavização exponencial simples com R. 2019. Disponível em: <https://blog.curso-r.com/posts/2019-02-10-ses/>. Acesso em: 17 dez. 2021.

FILIPA, S. Compreendendo LSTM Networks. 2021. Disponível em: <https://www.cetax.com.br/blog/compreendendo-lstm-networks>

FILIPATTI F.; FRANCESCHINI G.; TASSONI C.; VAS P. (2000). Recent developments of induction motor drives fault diagnosis using AI techniques. IEEE Transactions on Industrial Eletronics, v. 47, n°. 5, p. 994-1004.

G1. Brasil registra 2.399 mortes por Covid em 24 horas. 2021. Disponível em: <https://g1.globo.com/jornal-nacional/noticia/2021/05/26/brasil-registra-2399-mortes-por-covid-em-24-horas.ghtml/>. Acesso em 07 mai. 2021.

GERS, F; SCHMIDHUBER, J. "Redes recorrentes LSTM aprendem linguagens simples livres de contexto e sensíveis ao contexto". IEEE Transactions on Neural Networks. 12(6): 1333–1340. Disponível em: <ftp://ftp.idsia.ch/pub/juergen/L-IEEE.pdf>. Acesso em 03 mai. 2021.

GHAZALY, NM; ABDEL-FATTAH, MA; ABD EL-AZIZ, AA Novel coronavirus forecasting model using nonlinear autoregressive artificial neural network. International Journal of Advanced Science and Technology, 2020.

GOLDSCHMIDT, R.; PASSOS E. Data mining: um guia prático, conceitos, técnicas, ferramentas, orientações e aplicações. São Paulo: Elsevier; 2005.

GOMES, S, T. O QUE É O ALGORITMO HOLT-WINTERS E COMO FUNCIONA. 2019. Disponível em: <https://www.opservices.com.br/holt-winters/>. Acesso em: 17 dez. 2021.

GONZAGA, S. Holt-winter aditivo e multiplicativo. 2021. Disponível em: <http://sillasgonzaga.com/material/curso_series_temporais/suavizacao.html#holt-winter-aditivo-e-multiplicativo/>. Acesso em: 17 dez. 2021.

GUAN, W.; NI, Z.; HU, Y.; et al. Clinical characteristics of coronavirus disease 2019 in China. New England Journal of Medicine, 2020

GUAN, W.; NI, Z.; HU, Y.; et al. Clinical characteristics of coronavirus disease 2019 in China. New England Journal of Medicine, 2020.

GUYON I. (1991). Neural networks and applications tutorial. Physics Reports, v. 17 n°. 3, p. 215-259.

HAYKIN, S. Neural networks: a comprehensive foundation. New Delhi: Pearson Prentice Hall, 2005.

HOLT, C. E. (1957). Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA.

HUANG, C.; WANG, Y.; LI, X.; et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet, 2020.

HYNDMAN, R.J. A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods. International Journal of Forecasting, Volume 18, Issue 3, July-September 2002, Pages 439-454.

JIANG, X.; RAYNER, S.; LUO, M. Does SARS-CoV-2 has a longer incubation period than SARS and MERS? Journal of Medical Virology, v. 92, n. 5, pág. 476–478, 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.25708>. Acesso em 20 jun. 2021.

JOSHI, R.; JOHN, O.; JHA, V. The Potential Impact of Public Health Interventions in Preventing Kidney Disease. Seminars in Nephrology, v. 37, n. 3, pág. 234–244, 2017. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0270929517300049>. Acesso em: 15 jun. 2021

JUNIOR, A, LIRA. Séries temporais e componentes: aplicando ARIMA para forecast em dados do COVID-19. 2017. Disponível em: <https://imasters.com.br/data/series-temporais-e-componentes-aplicando-arima-para-forecast-em-dados-do-covid-19>. Acesso em 22 jul. 2021.

JUNIOR, E, AMARO. COVID-19: Desafios para modelagem e análise de dados. 2020. Disponível em: <http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/773>. Acesso em 29 abr. 2021.

JÚNIOR, J. Redes Neurais Recorrentes — LSTM. 2019. Disponível em: <https://medium.com/@web2ajax/redes-neurais-recorrentes-lstm-b90b720dc3f6>. Acesso em 29 abr. 2021.

KAWAGUCHI, M.; NUKAGA, T.; SEKINE, S.; et al. Mechanism-based integrated assay systems for the prediction of drug-induced liver injury. Toxicology and Applied PHARMACOLOGY, V. 394, p. 114958, 2020. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0041008X2030082X>. Acesso em: 29 jun. 2021.

KERAS. Adam. 2021. Disponível em: <https://keras.io/api/optimizers/adam/>. Acesso em 10 mai. 2021.

KERAS. Dropout layer. 2021. Disponível em: <https://keras.io/api/layers/regularization_layers/dropout/>. Acesso em 10 mai. 2021.

KINGMA, D. “Adam: A Method for Stochastic Optimization”, Conrell University, vol. 1, Jan. 2017. Disponível em: <https://arxiv.org/abs/1412.6980>. Acesso em 13 mai. 2021.

LEITE, S. J. O. ; OLIVEIRA, R. C. L. ; CAMPOS, L. M. L. . Predição de séries temporais da COVID19: uma avaliação de redes neurais com células LSTM. XV CONGRESSO BRASILEIRO DE INTELIGÊNCIA COMPUTACIONAL - CBIC 2021, Joinville, SC, 2021. v. 1. p. 1

LIMA, C, L.; SILVA, A, C, G.; SILVA, C, C.; FILHO, A, G, S.; SANTOS, W, P. Sistema Inteligente para o monitoramento e predição da COVID-19 em tempo real. 2020. Disponível em: < http://revistas.poli.br/index.php/anais/article/view/1577>. Acesso em: 22 dez. 2021.

MAGGI, L. Universidade Johns Hopkins: o primeiro centro de pesquisa dos EUA. Publicado em: 2021. Disponível em: <https://www.estudarfora.org.br/universidade-johns-hopkins-o-primeiro-centro-de-pesquisa-dos-eua/>. Acesso em 22 ago. 2021.

MARTINS, VLM; WERNER, L. Forecast combination in industrial series: A comparison between individual forecasts and its combinations with and without correlated errors. Expert Systems with Applications, 2012.

MCCULLOCH W. S.; PITTS W. H. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, v. 5, p. 115-133.

MINISTÉRIO DA SAÚDE. Painel Coronavírus. Disponível em: <https://covid.saude.gov.br/>. Acesso em: 15 jun. 2021.

MOLLALO, A.; RIVERA, KM; VAHEDI, B. Artificial neural network modeling of novel coronavirus (COVID-19) incidence rates across the continental United States. International Journal of Environmental Research and Public Health, 2020.

MORETTIN, P.; TOLOI, C. M. C. Análise de Séries Temporais. São Paulo: Blucher, 2006.

NARVÁEZ, V, D.; ROLDÁN, D, S, M.; NÚÑEZ, A, C.; ROLDÁN, P, S, M.; VELOSO, G, R. Qual é a curva que melhor explica o crescimento de casos confirmados da COVID-19 no Chile? 2020. Disponível em: <https://www.scielo.br/j/rlae/a/S33463KTxp6yymPpW3Swhzw/?lang=pt>. Acesso em: 22 dez. 2021.

NIAZKAR, H.R., NIAZKAR, M. Application of artificial neural networks to predict the COVID-19 outbreak. glob health res policy, Global Health Research and Policy, Vol. 1, nov 2020. Disponível: <https://doi.org/10.1186/s41256-020-00175-y>. Acesso em 13 mai. 2021.

ODRIOZOLA, EP; QUINTANA, AM; GONZÁLEZ, V.; et al. Towards leprosy elimination by 2020: Forecasts of epidemiological indicators of leprosy in corrientes,

a province of northeastern Argentina that is a pioneer in leprosy elimination. Memorias do Instituto Oswaldo Cruz, 2017.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. QA on coronavírus. Disponível em: <https://www.who.int/emergencies/diseases/ novelcoronavirus-2019/question-and-answers-hub/q-adetail/q- a-coronaviruses/>. Acesso em: 19 jun. 2021.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. WHO Coronavirus Disease (COVID-19) Dashboard. Disponível em: <https:// covid19.who.int/>. Acesso em: 19 jun 2021.

OXFORD. Our World in Data. 2021. Disponível em: <https://ourworldindata.org/covid-deaths/>. Acesso em: 30 nov. 2021.

PASQUINI-DESCOMPS, H.; BRENDER, N.; MARADAN, D. Value for Money in H1N1 Influenza: A Systematic Review of the Cost-Effectiveness of Pandemic Interventions. Value in Health, 2017.

PÉREZ-CASTRO, R.; CASTELLANOS, JE; OLANO, VA; et al. Detection of all four dengue serotypes in Aedes aegypti female mosquitoes collected in a rural area in Colombia. Memorias do Instituto Oswaldo Cruz, 2016.

PERKTOLD, J.; SEABOLD, S.; TAYLOR, J. Statistical models, hypothesis tests, and data exploration. 2019. Disponível em: <https://www.statsmodels.org/stable/index.html>. Acesso em 27 ago. 2021.

PERKTOLD, J.; SEABOLD, S.; TAYLOR, J. Time Series analysis statsmodels.tsa.arima.model.ARIMA. 2019. Disponível em: <https://www.statsmodels.org/dev/generated/statsmodels.tsa.arima.model.ARIMA.html>. Acesso em 22 jul. 2021.

PICCININI G. (2004). The first computational theory of mind and brain: a close look at McCulloch and Pitts's "Logical calculus of ideas immanent in nervous activity" Kluwer Academic Publisher, v. 141, p. 175-215.

REEVES, P.; EDMUNDS, K.; SEARLES, A.; WIGGERS, J. Economic evaluations of public health implementation-interventions: a systematic review and guideline for practice. Public Health, 2019. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30877961/>. Acesso em: 15 jun. 2021.

ROTA, P, A.; OBERSTE, M, S.; MONROE, S, S.; et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science, 2003.

SABA, AI; ELSHEIKH, AH Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. Process Safety and Environmental Protection, 2020.

SCARABEL, F.; PELLIS, L.; BRAGAZZI, NL; WU, J. Canada needs to rapidly escalate public health interventions for its COVID-19 mitigation strategies. Infectious Disease Modelling, 2020.

SELLI, M. F. (2007). Identificação de padrões de escoamento horizontal bifásico gás-líquido através de distribuição tempo-freqüência e redes neurais. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos. 2007.

SELLI, M. F.; SELEGHIM P. J. (2007). On-line identification of horizontal two-phase flow regimes through Gabor transform and neural network processing. Heat Transfer Engineering, v. 28, n°. 6, p. 541-548.

SHAW, R.; KIM, Y.; HUA, J. Governance, technology and citizen behavior in pandemic: Lessons from COVID-19 in East Asia. Progress in Disaster Science, 2020.

SOHRABI, C. “World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)”. International Journal of Surgery, 2020.

SPRINGML. Time Series Forecasting – ARIMA vs LSTM. Disponível em: <https://www.springml.com/blog/time-series-forecasting-arima-vs-lstm/>. Acesso em: 15 jun. 2021.

SURVEILLANCES, V. “The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID- 19) - China, 2020”. China CDC Weekly 2.8(2020), pp. 113–122.

TAGLIARINI G. A.; CHRIST J. F.; PAGE E. W. (1991). Optimization using neural network. IEEE Transactions on Computers, v. 40, n°. 12, p. 1347-1358.

TANG, Y. “Laboratory diagnosis of COVID-19: current issues and challenges”. Journal of clinical microbiology 58.6, 2020.

TEES R.C. (2002). Review of the organization of behavior: a neuropsychological theory by Donald O. Hebb (1949). Canadian Psychological Association v. 44. 2002.

TIMOSZCZUK, A. P. (2004). Reconhecimento automático do locutor com redes neurais pulsadas. Tese (Doutorado) – Escola Politécnica, Universidade de São Paulo, São Paulo. 2004.

WANG AND MENGJUAN LIU, G. Dynamic Trust Model Based on Service Recommendation in Big Data. Computers, Materials & Continua, v. 58, n. 3, pág. 845–857, 2019. Disponível em: <http://www.techscience.com/cmc/v58n3/23034>. Acesso em 23 jun. 2021.

WANG AND MENGJUAN LIU, G. Dynamic Trust Model Based on Service Recommendation in Big Data. Computers, Materials & Continua, v. 58, n. 3, pág. 845–857, 2019. Disponível em: <http://www.techscience.com/cmc/v58n3/23034>. Acesso em 24 jun. 2021.

WINTERS, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6(3), 324–342.

WOLD, H. A study in the analyses of stationary time series. Stockhoml: Almqvist e Wiksell, 1938.

YULE, G. Why do we sometimes get nonsense-correlations between times series? A study in sampling and the nature time series. SIAM Journal, n. 2, p. 229–239, 1963.

ZANG, G.P. Avoiding pitfalls in neural network research. IEEE Transactions on Systems, Man and Cybernetics, Part C, p. 3-16, 2007.MENEZES JUNIOR, J. M. P. Contribuições ao problema de predição recursiva de séries temporais univariadas usando redes neurais recorrentes. Tese (doutorado) - Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-graduação em Engenharia de Teleinformática, Fortaleza, 2012.

ZHUANG, Z.; ZHAO, S.; LIN, Q.; et al. Preliminary estimation of the novel coronavirus disease (COVID-19) cases in Iran: A modelling analysis based on overseas cases and air travel data. International Journal of Infectious Diseases, 2020.

**ANEXO A -** Imagens e descrições de gráficos de resultados de predições dos modelos deste trabalho

As figuras presentes neste ANEXO A seguem a mesma premissa das imagens mostradas anteriormente, na análise de resultados. Primeiro, são mostradas as predições de treino e teste de casos confirmados e posteriormente, as de mortes dos países Índia, Itália e EUA. As figuras estão divididas por modelos, começando pelos resultados do modelo LSTM da Índia.

**Modelo LSTM:**

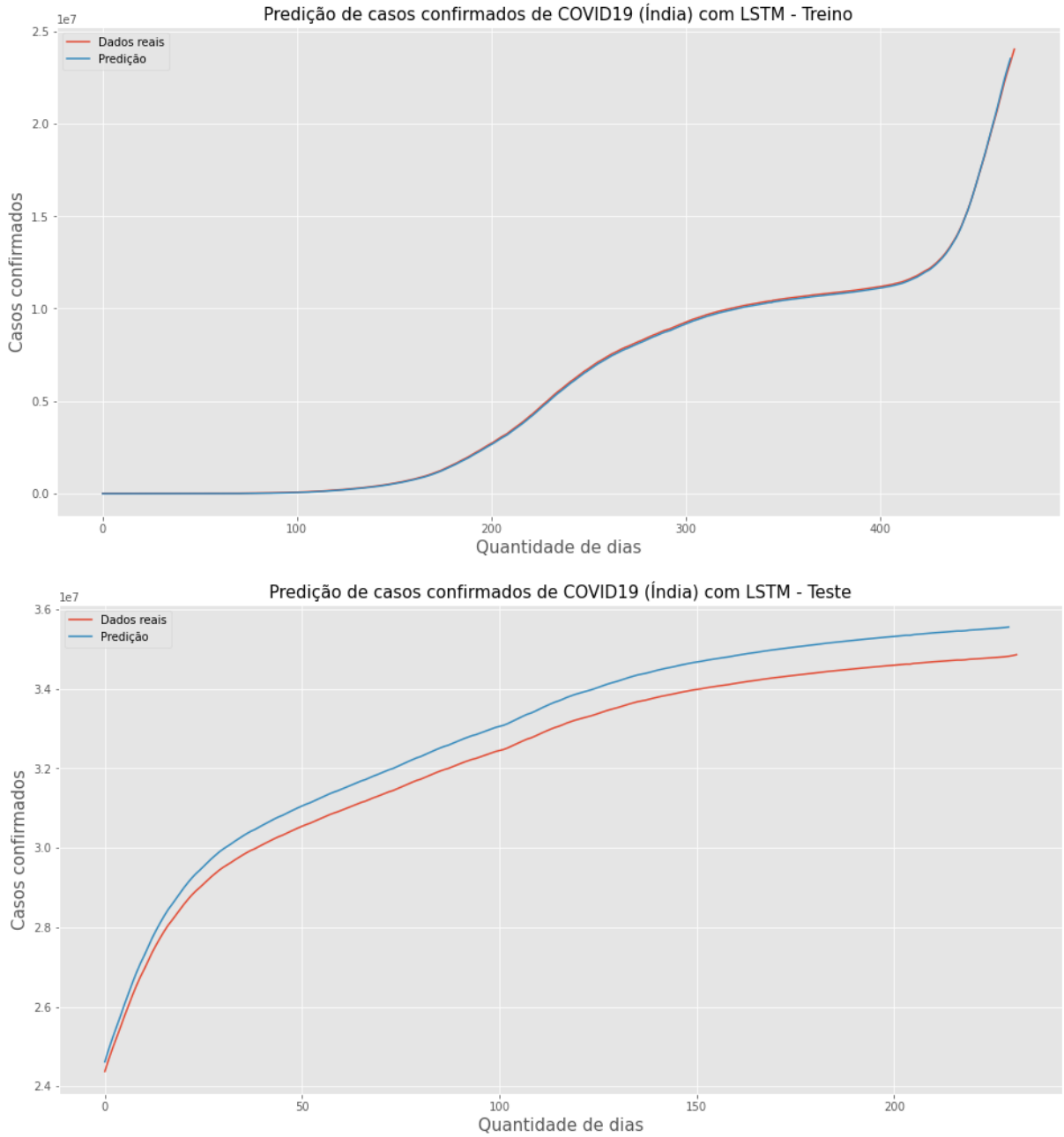


**Figura 23 – Resultado das predições de casos confirmados da Índia com LSTM**
**Fonte: Autor**

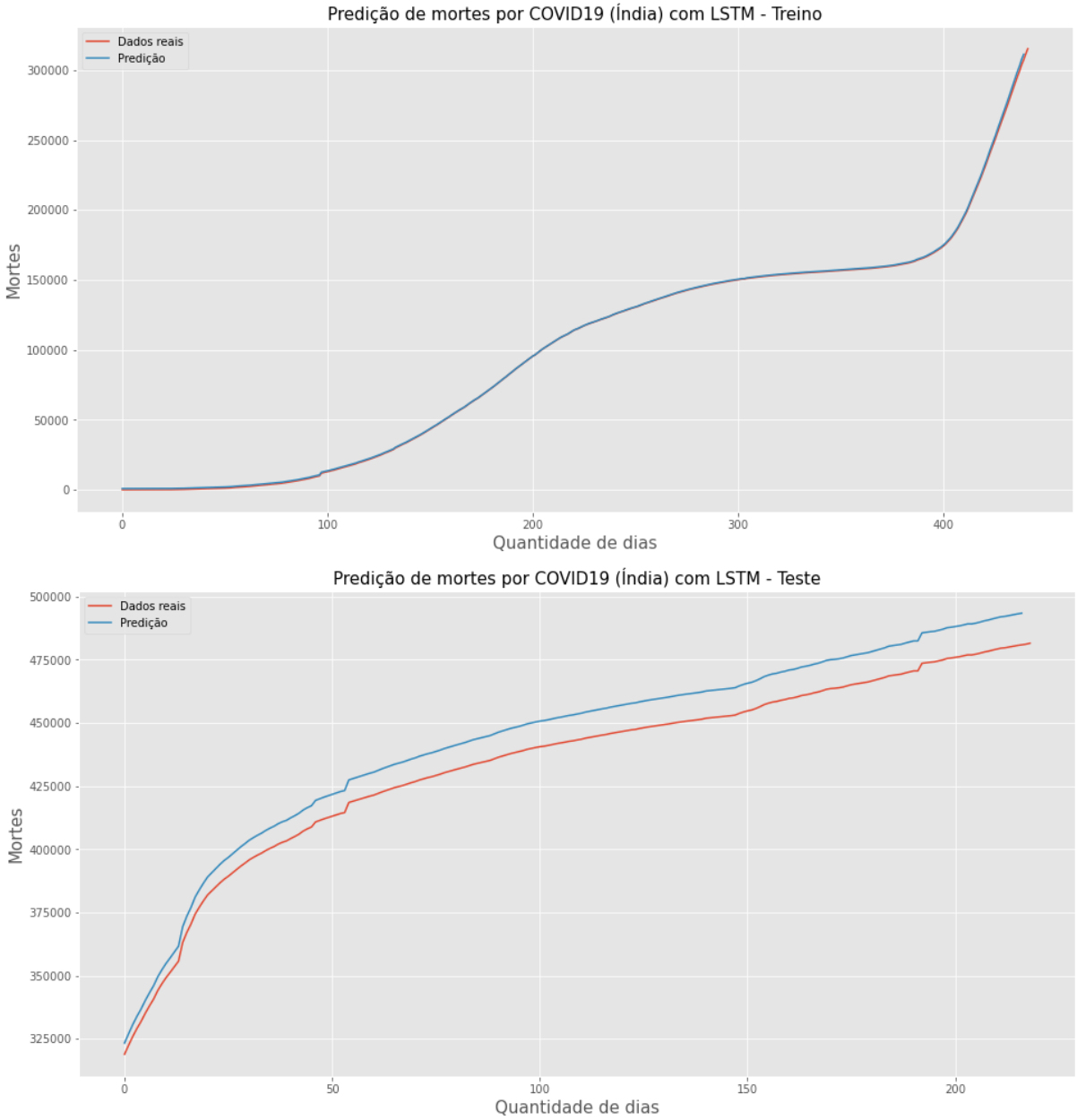


**Figura 24 – Resultado das predições de mortes da Índia com LSTM**
**Fonte: Autor**

Abaixo, as Figuras 25 e 26 expõe os resultados das predições com bases nos dados de casos confirmados e mortes na Itália. Os dados são consideravelmente menores em comparação com os demais países avaliados. Isso se dá por conta de a quantidade de habitantes do país ser menor que aos demais avaliados.

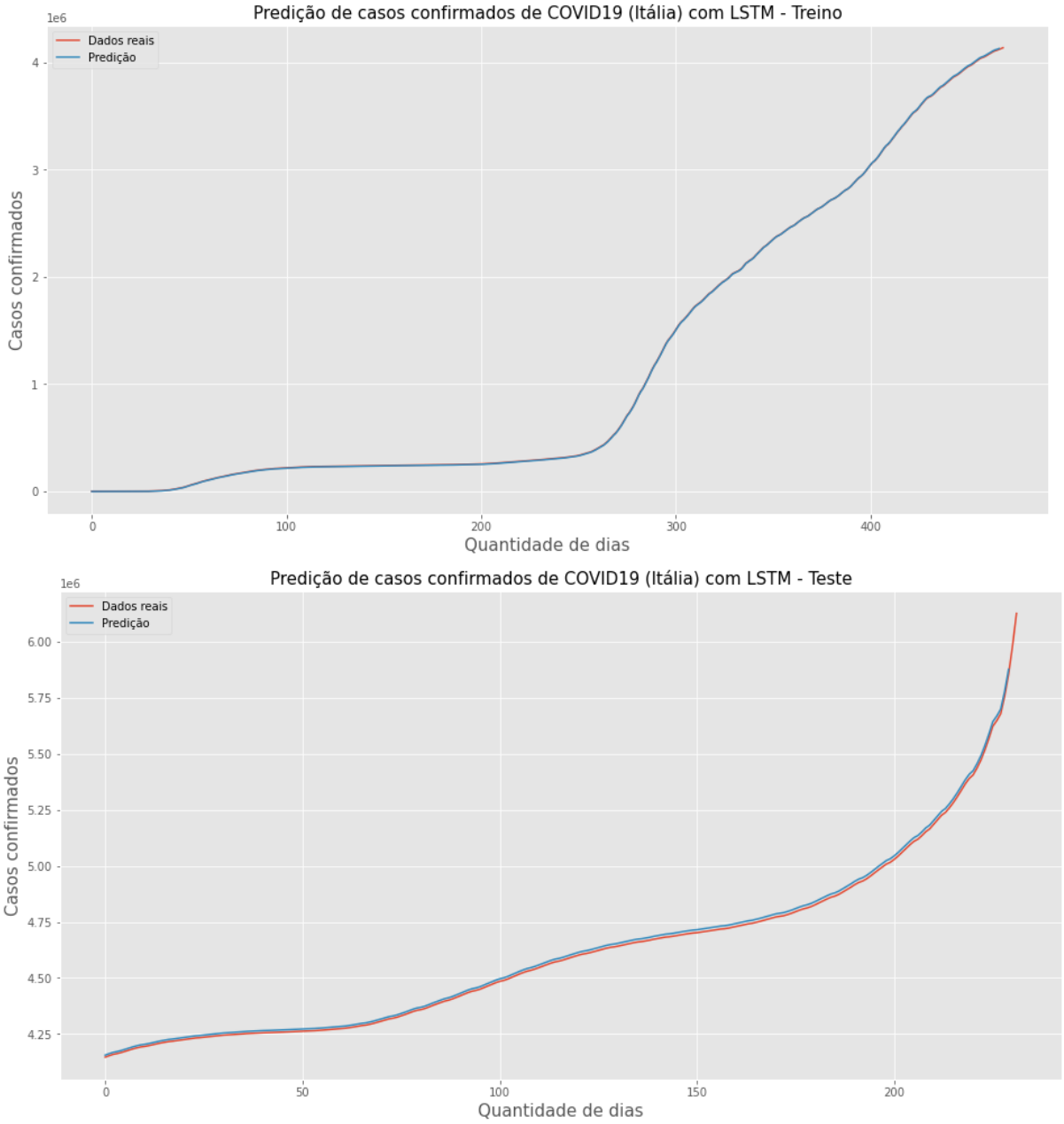


**Figura 25 – Resultado da predição de casos confirmados da Itália com LSTM**
**Fonte: Autor**

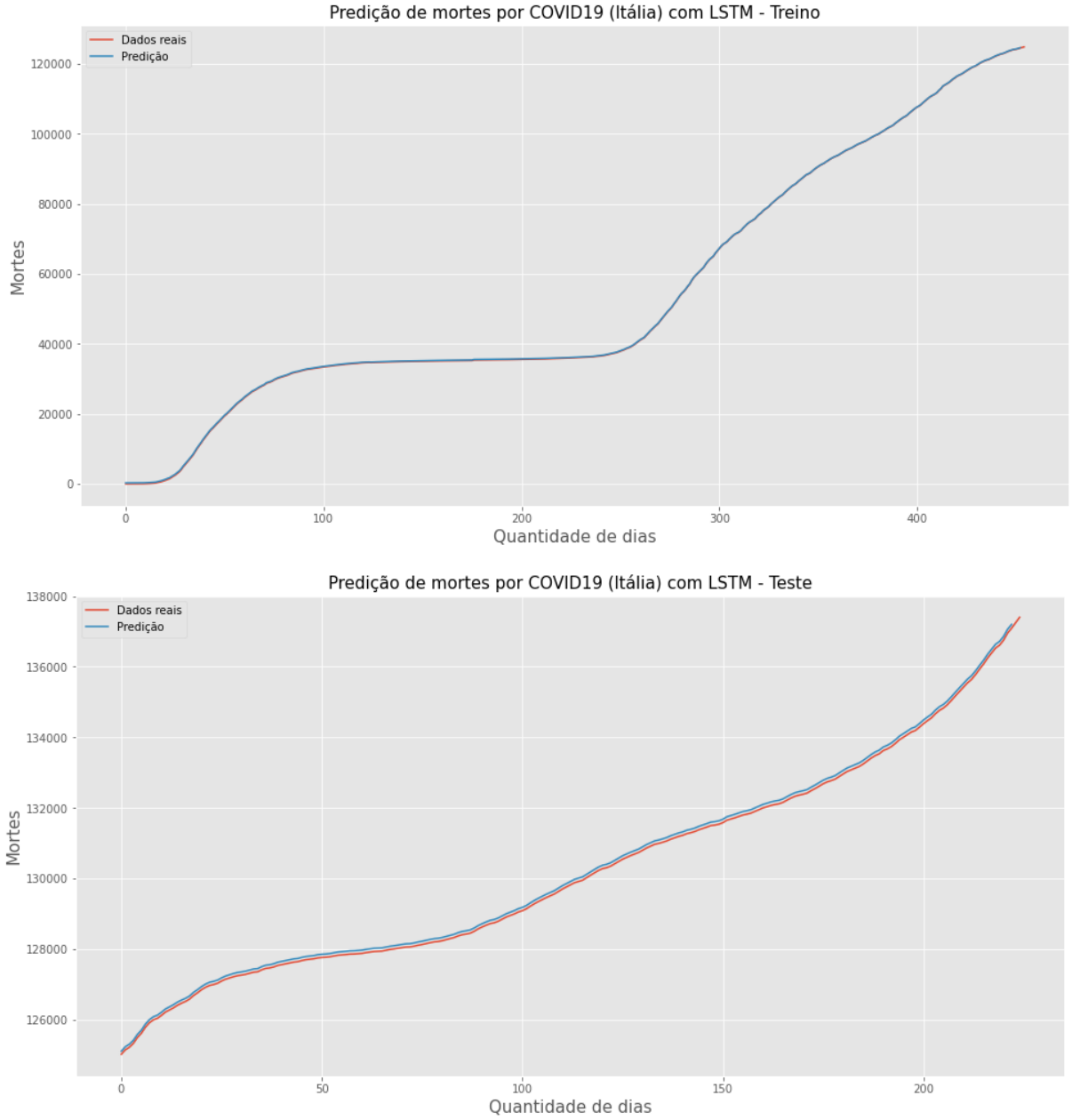


**Figura 26 – Resultado da predição de mortes da Itália com LSTM**
**Fonte: Autor**

A seguir, por fim, os gráficos das Figura 27 e 28 mostram os resultados da predição com base nos dados do país dos Estados Unidos da América.

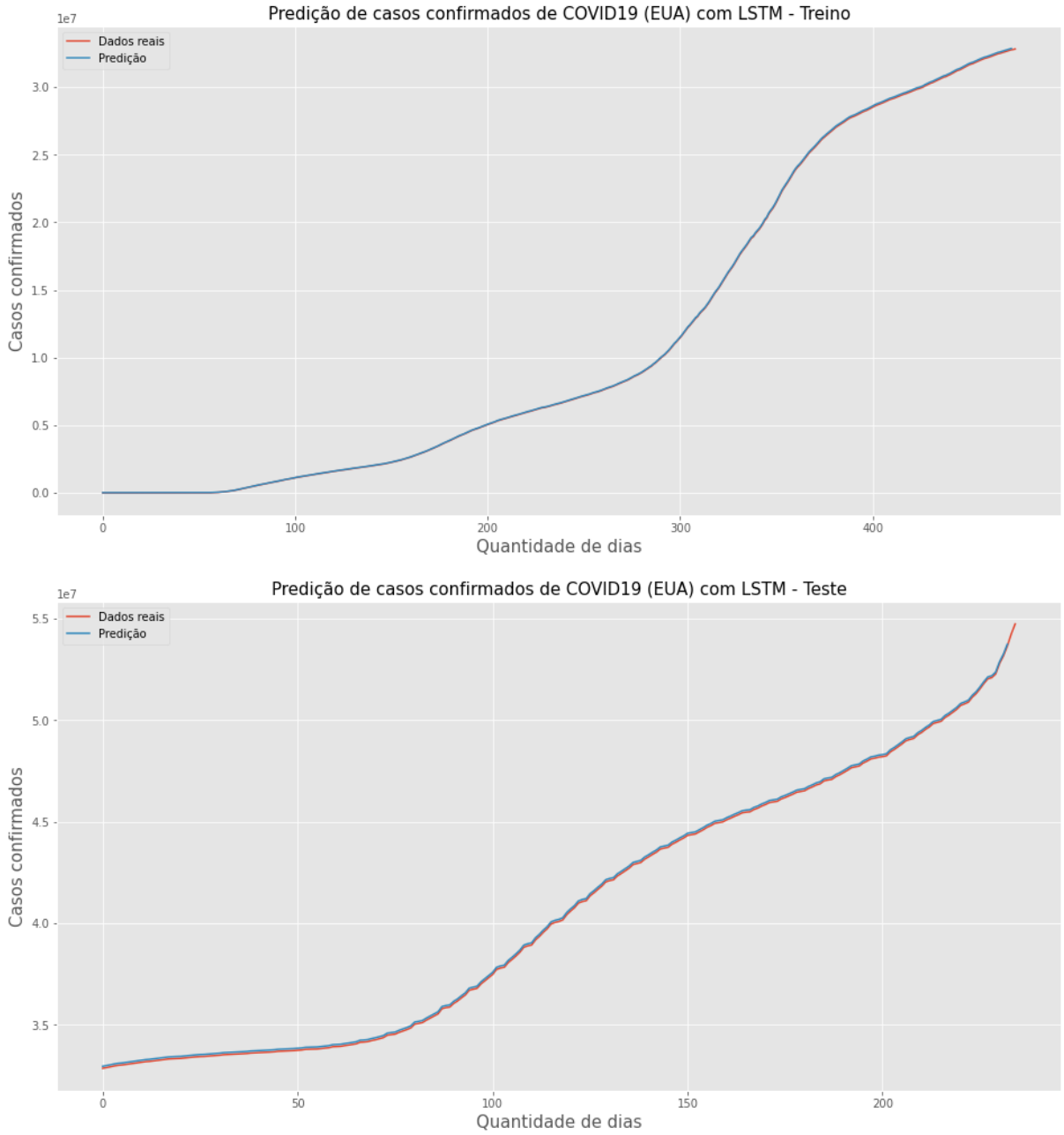


**Figura 27 – Resultado da predição de casos confirmados dos EUA com LSTM**
**Fonte: Autor**

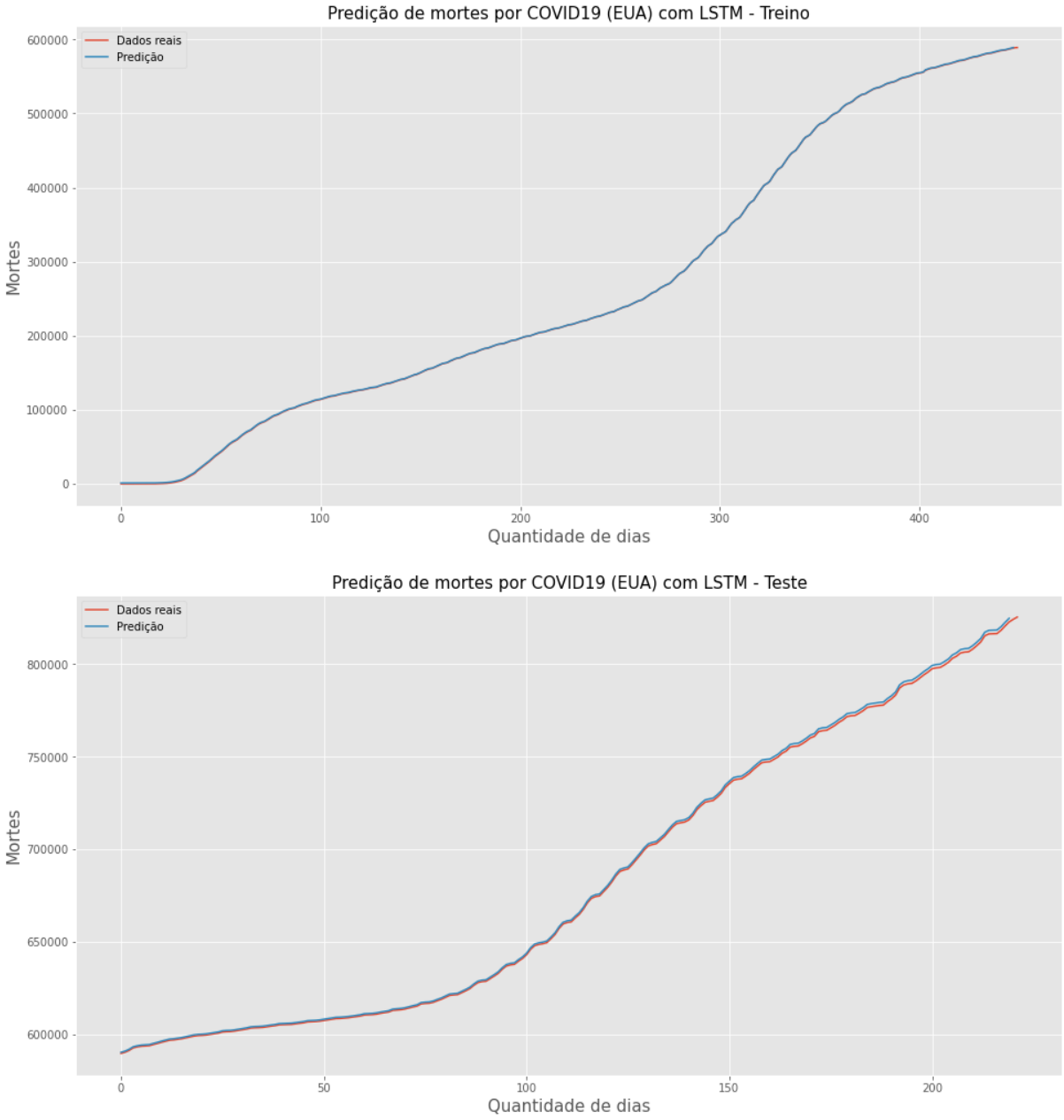


**Figura 28 – Resultado da predição de mortes do EUA com LSTM**
**Fonte: Autor**

**Modelo MLP:**

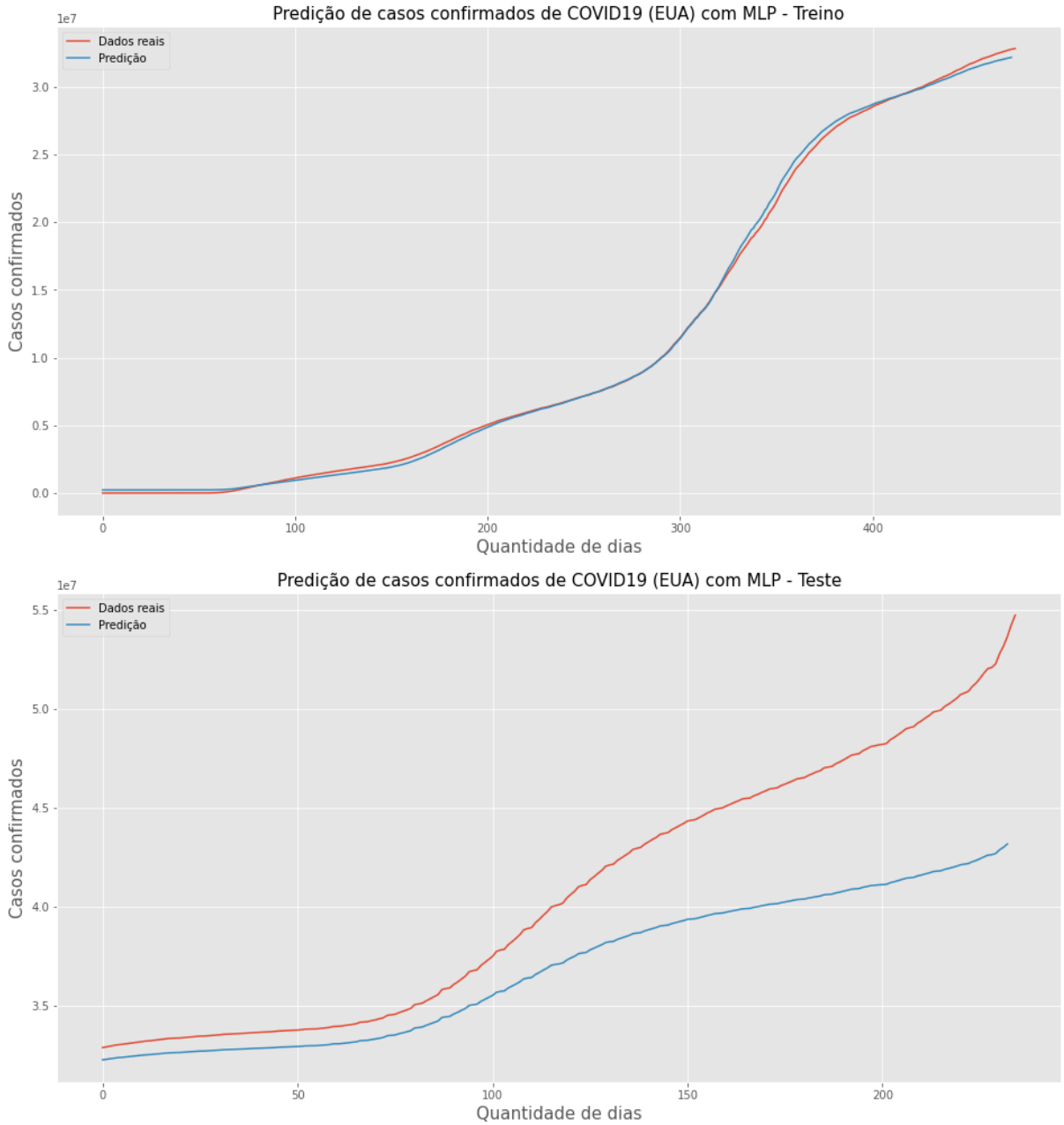


**Figura 29 – Resultado da predição de casos confirmados da Índia com MLP**
**Fonte: Autor**

Como mostram as Figuras 29 e 30, nas predições dos dados da Índia, é notório que o modelo está indo bem nos treinos. Porém, nos testes, visualmente, não houve um bom desempenho. Os resultados de RMSE serão discutidos posteriormente.

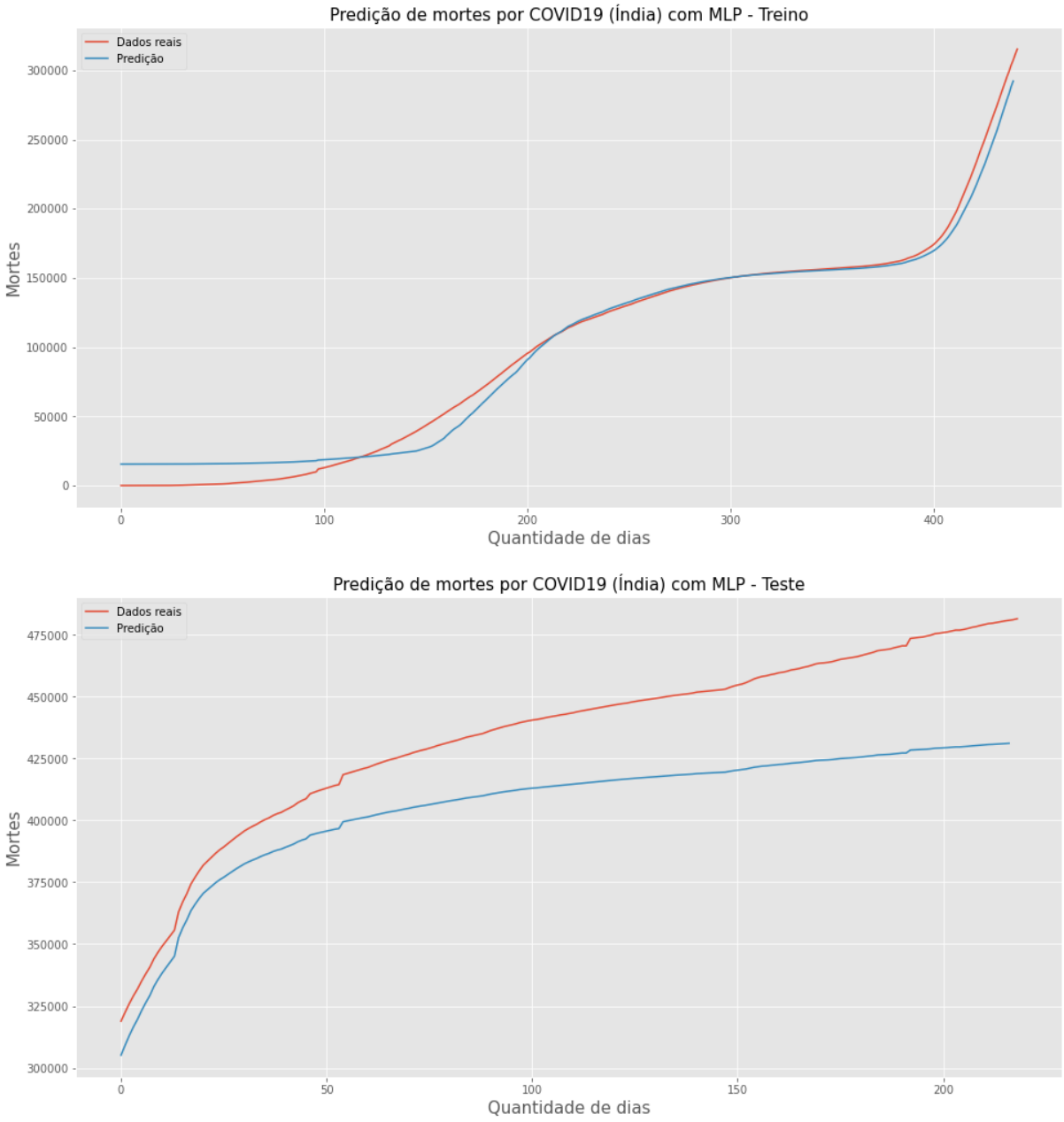


**Figura 30 – Resultado da predição de mortes da Índia com MLP**
**Fonte: Autor**

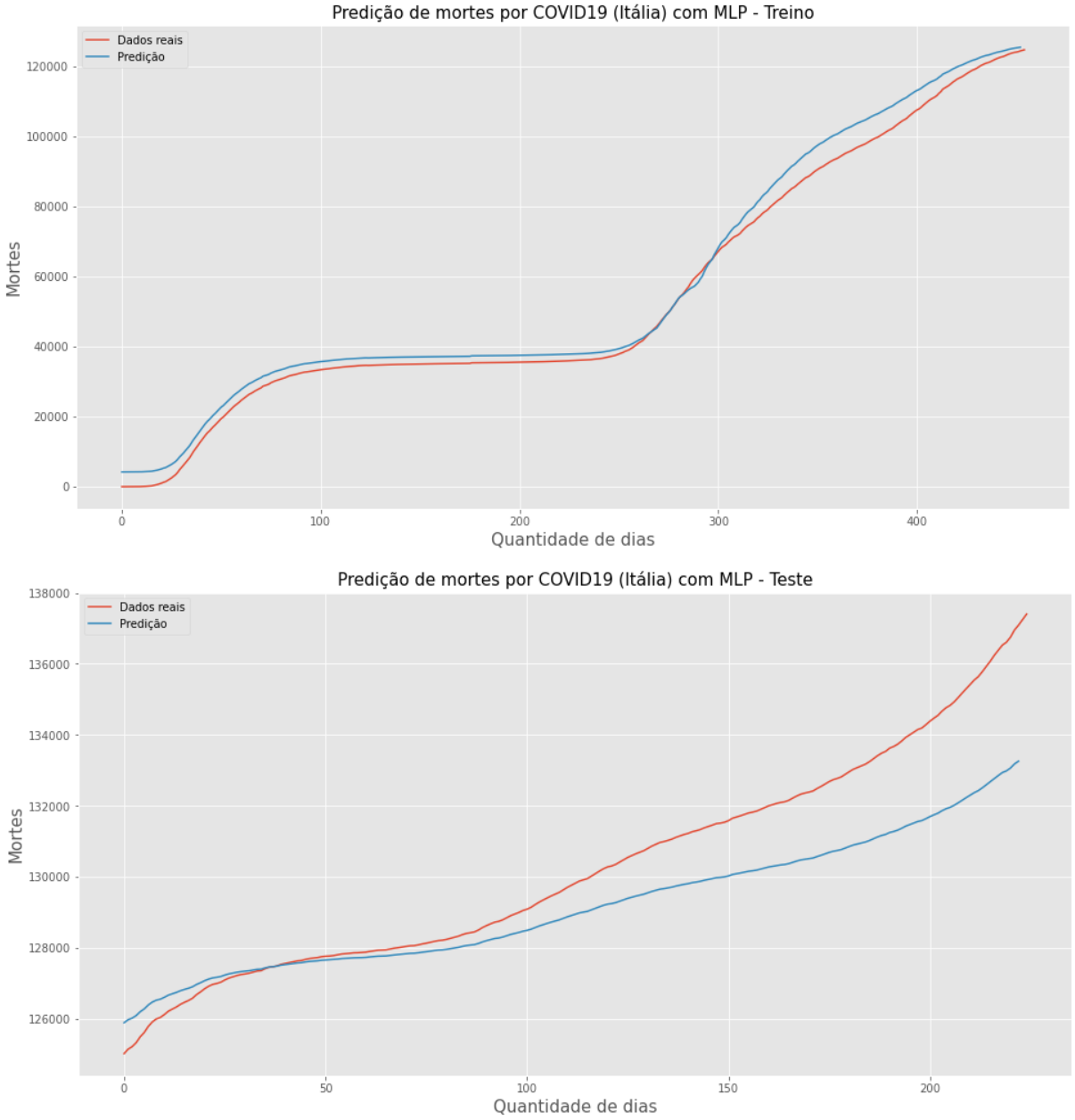


**Figura 31 – Resultado da predição de casos confirmados da Itália com MLP**
**Fonte: Autor**

As Figuras 31 e 32 mostram que, assim como na Índia, nas etapas de treinos, as predições dos dados da Itália do modelo estão indo bem. No entanto, nos testes, aparentemente, não está tendo um desempenho tão bom.

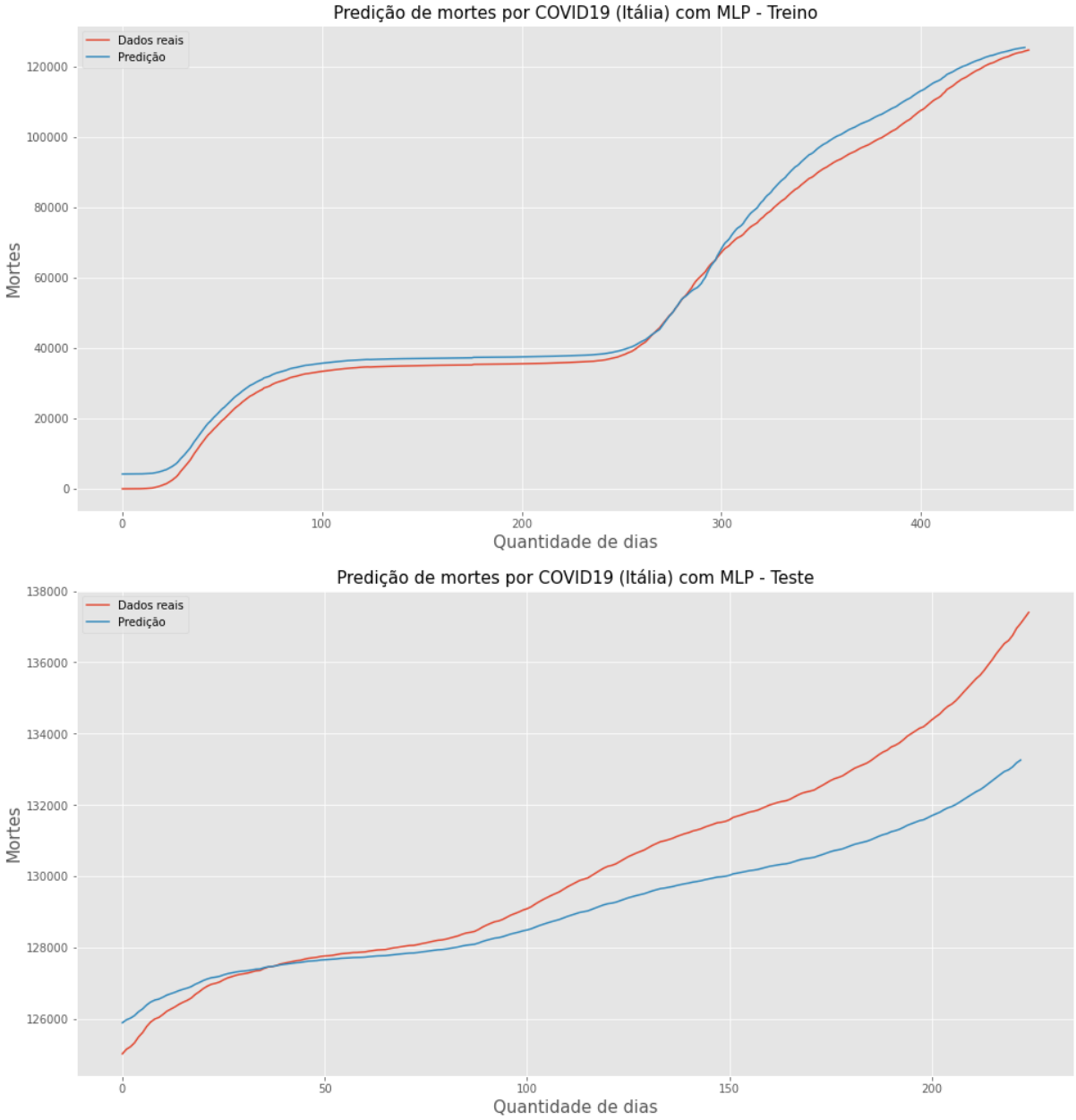


**Figura 32 – Resultado da predição de mortes da Itália com MLP**
**Fonte: Autor**

Pode-se notar que a as Figuras 33 e 34, nos resultados das predições dos Estados Unidos da América, obtiveram um desempenho parecido com o das demais predições dos outros países.

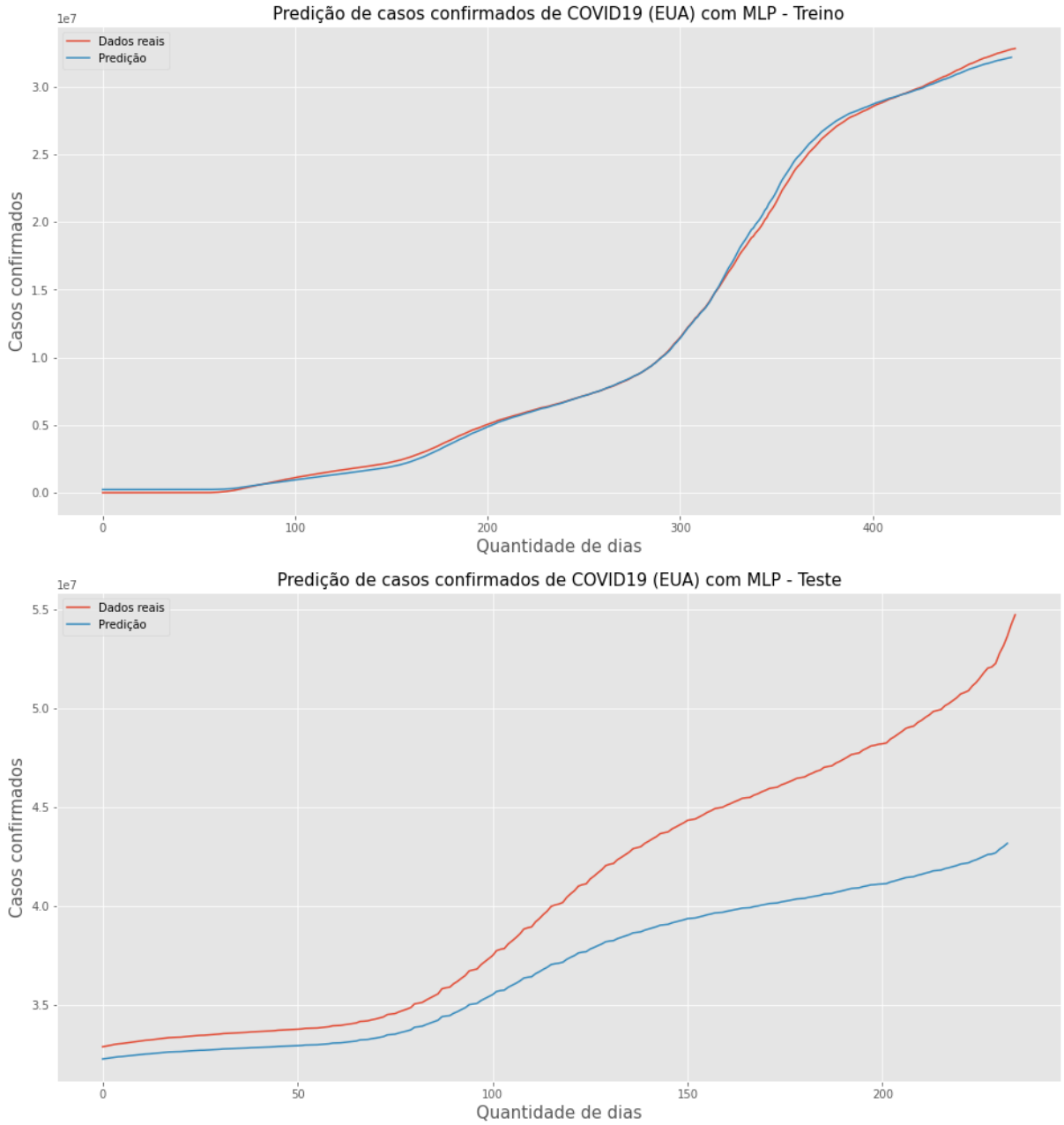


**Figura 33 – Resultado da predição de casos confirmados dos EUA com MLP**
**Fonte: Autor**

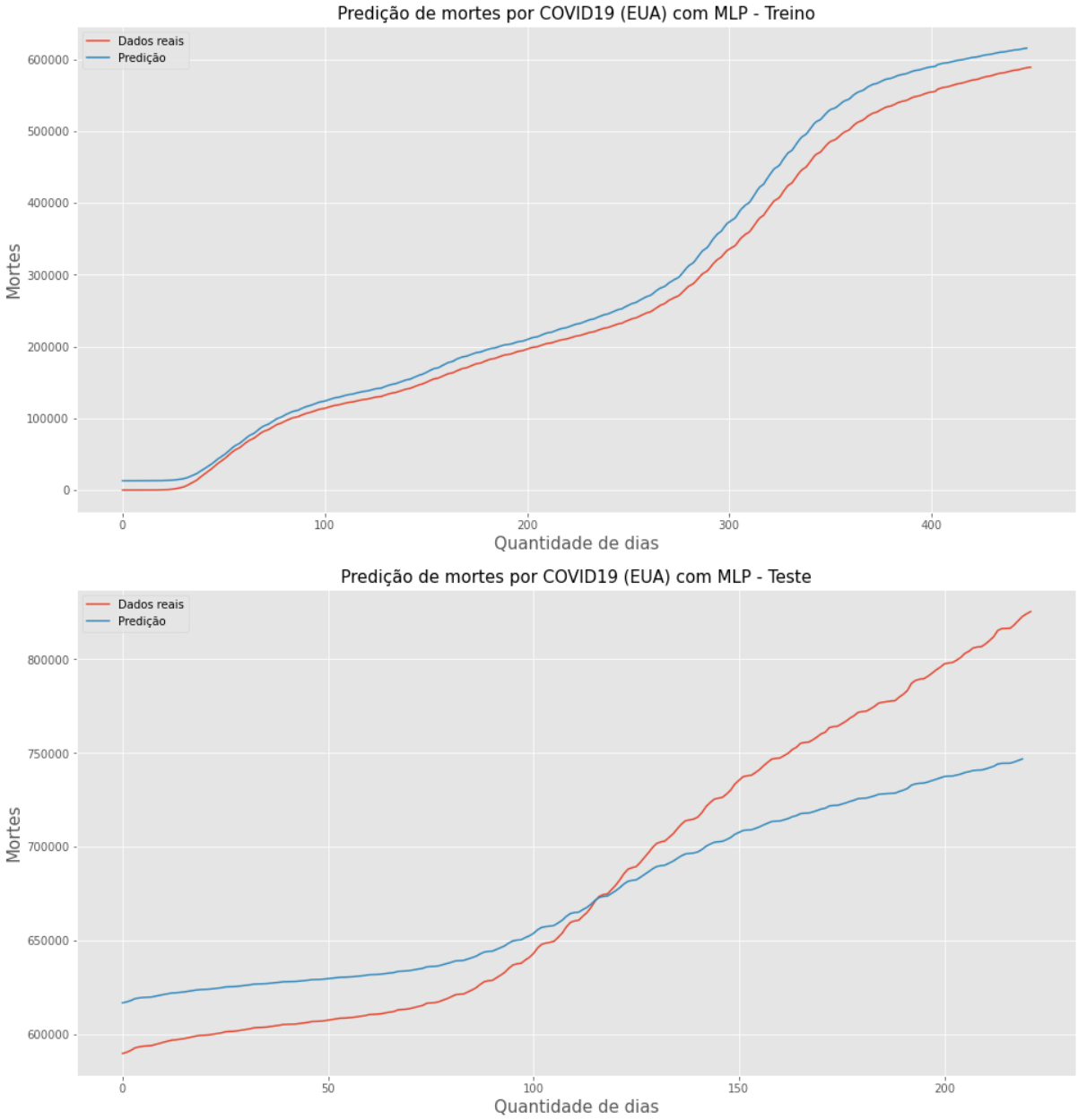


**Figura 34 – Resultado da predição de mortes do EUA com MLP**
**Fonte: Autor**

**Modelo ARIMA**:

O modelo ARIMA foi o modelo que obteve os melhores resultados com o menor erro em relação a todos os outros três modelos. Seus gráficos de resultados continuam com a mesma lógica de exibição dos demais. Abaixo encontram-se todos os gráficos que não foram adicionados na análise de resultados.

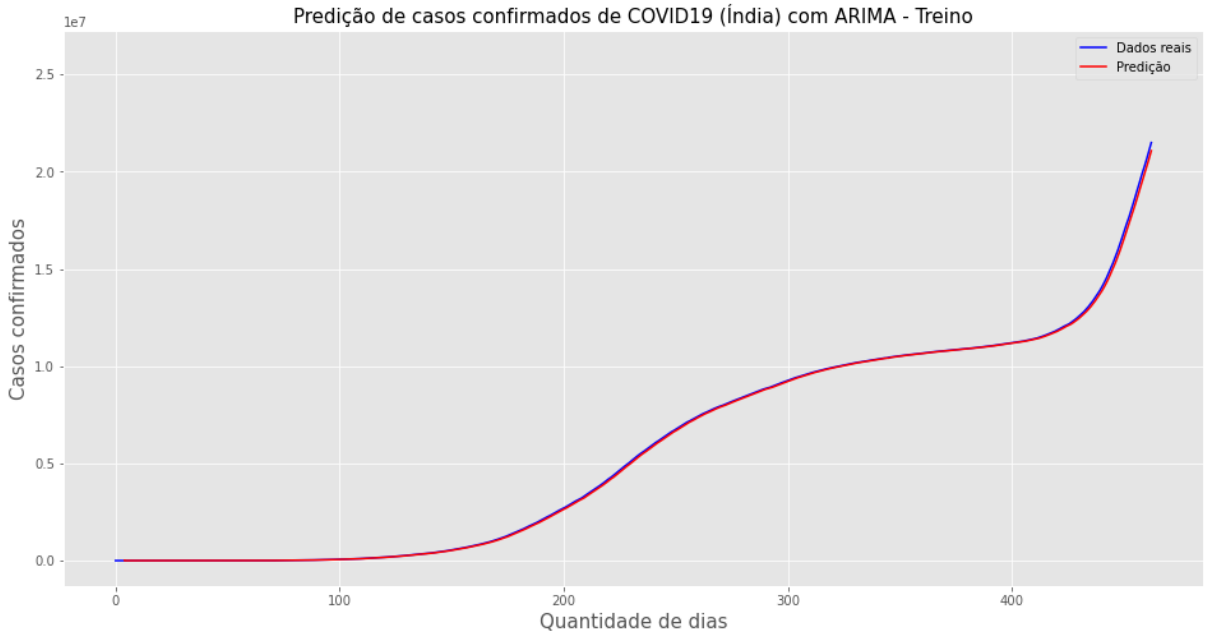


**Figura 35 – Resultado da predição casos confirmados da Índia com ARIMA - Treino**
**Fonte: Autor**

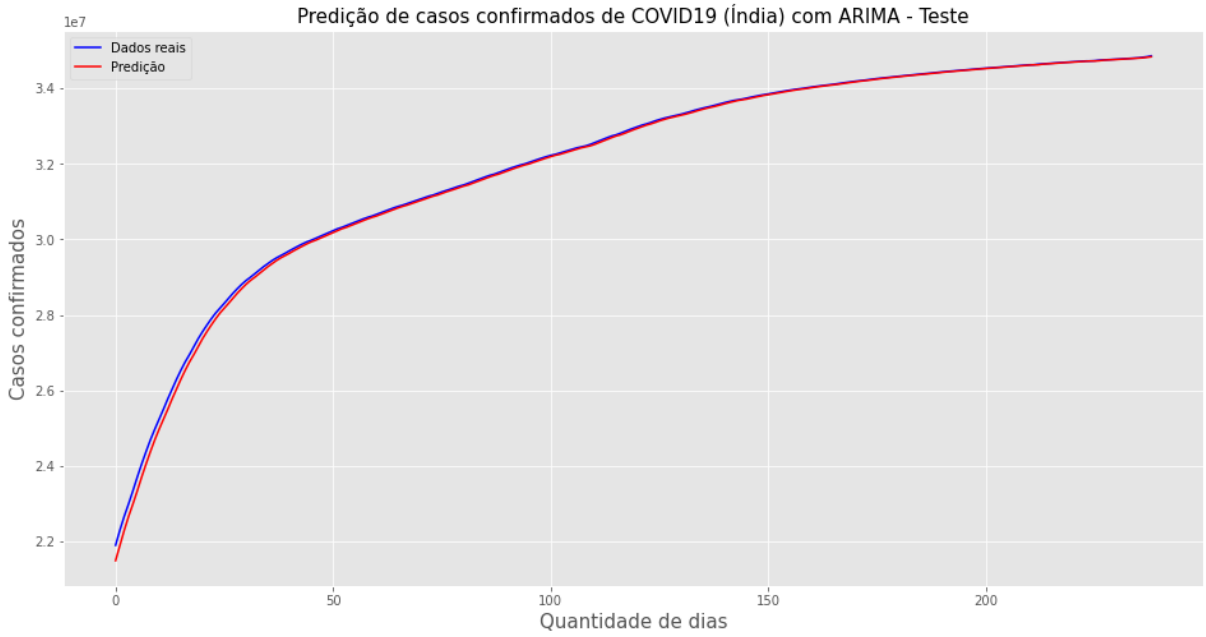


**Figura 36 – Resultado da predição casos confirmados da Índia com ARIMA - Teste**
**Fonte: Autor**

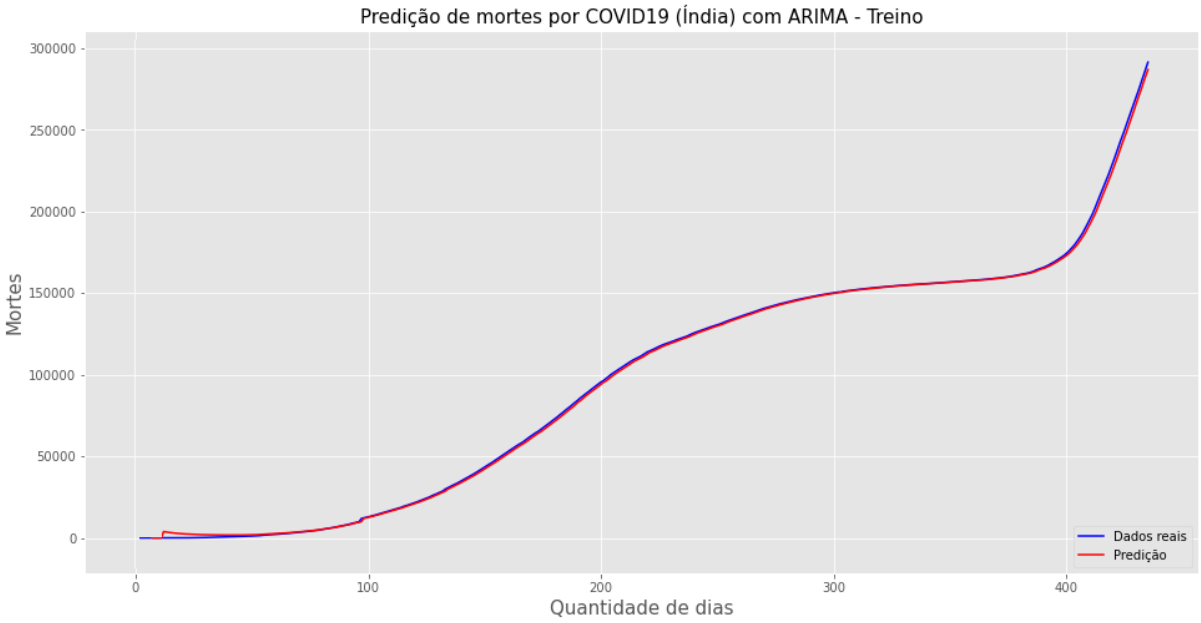


**Figura 37 – Resultado da predição de mortes da Índia com ARIMA - Treino**
**Fonte: Autor**

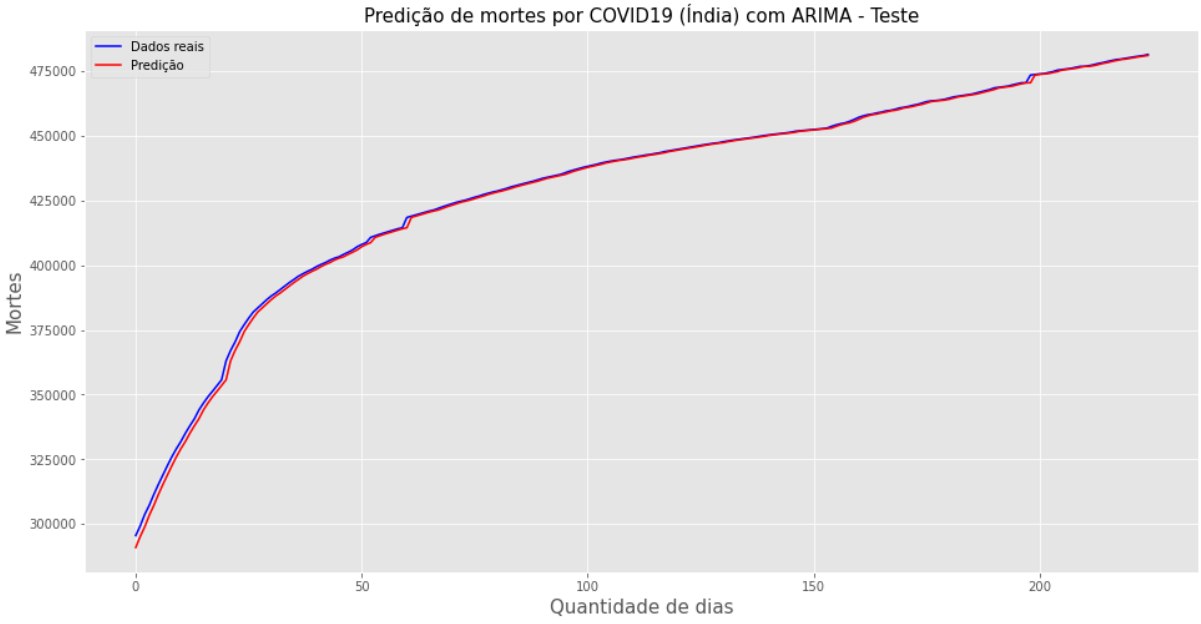


**Figura 38 – Resultado da predição de mortes da Índia com ARIMA - Teste**
**Fonte: Autor**

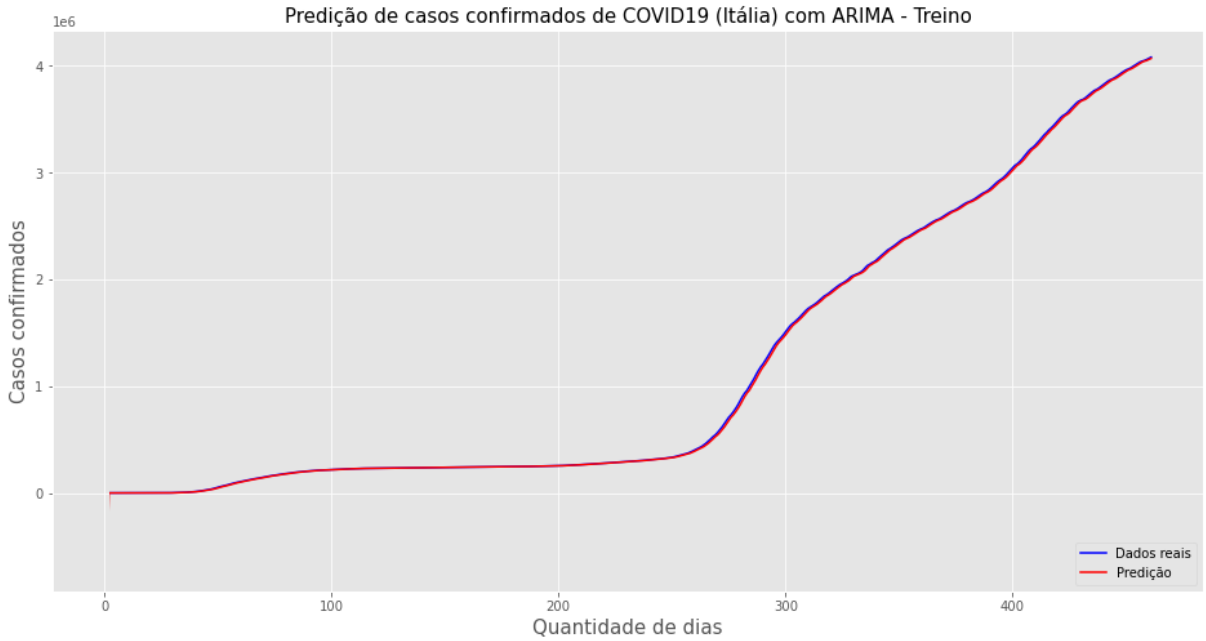


**Figura 39 – Resultado da predição de casos confirmados da Itália com ARIMA - Treino**
**Fonte: Autor**

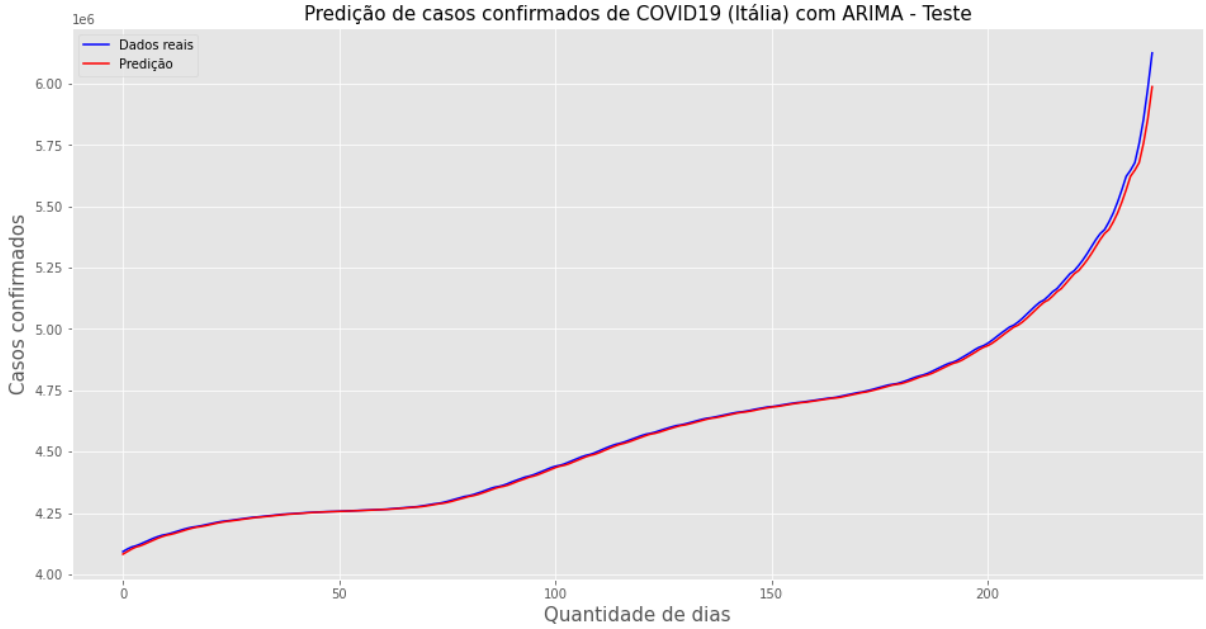


**Figura 40 – Resultado da predição de casos confirmados da Itália com ARIMA - Teste**
**Fonte: Autor**

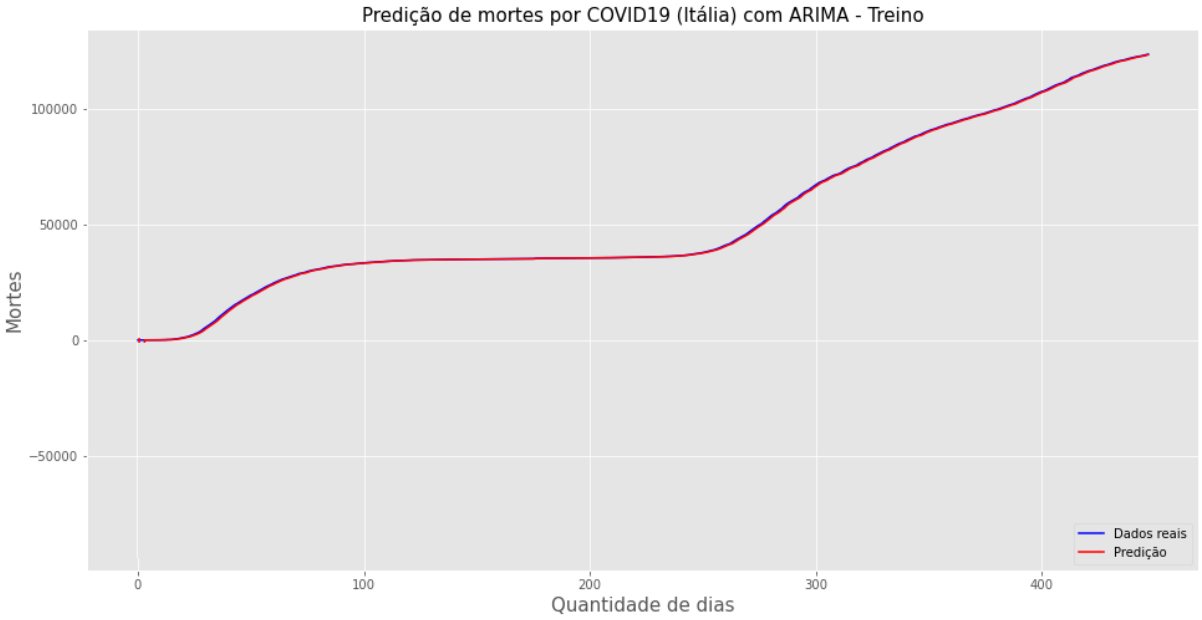


**Figura 41 – Resultado da predição de mortes da Itália com ARIMA - Treino**
**Fonte: Autor**

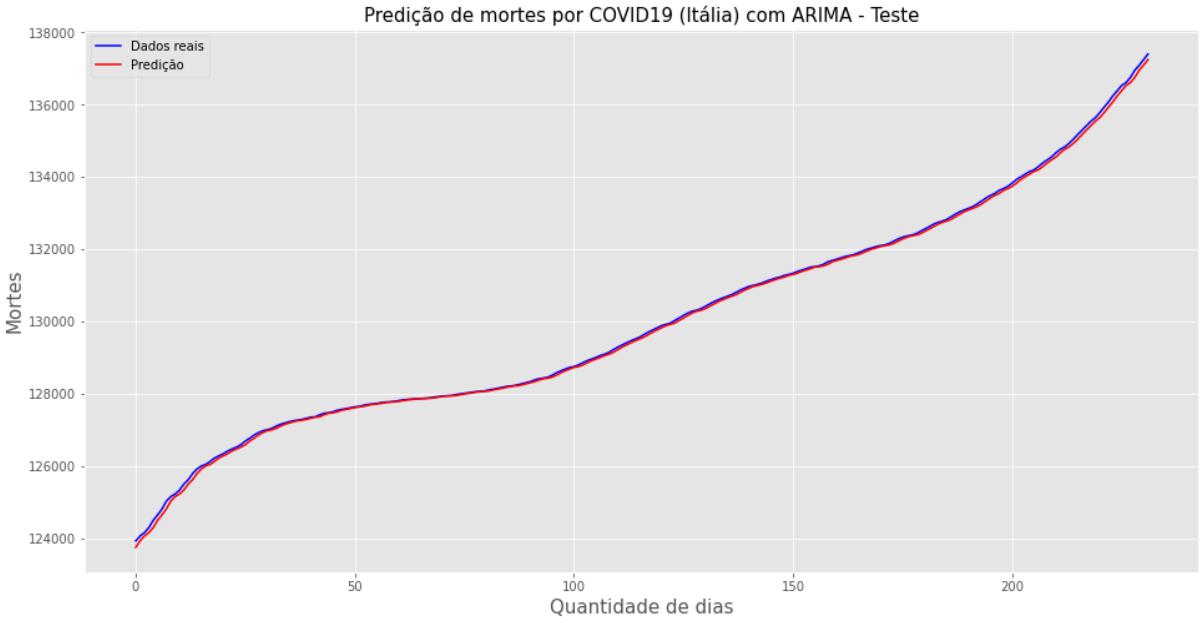


**Figura 42 – Resultado da predição de mortes da Itália com ARIMA - Teste**
**Fonte: Autor**

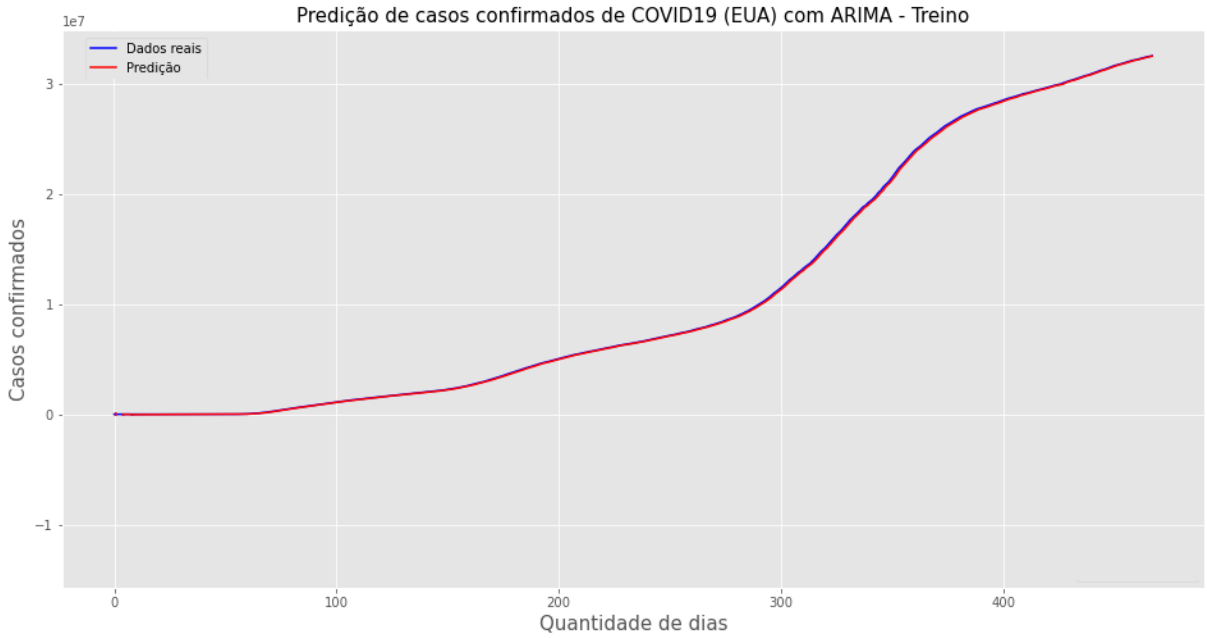


**Figura 43 – Resultado da predição de casos confirmados dos EUA com ARIMA - Treino**
**Fonte: Autor**

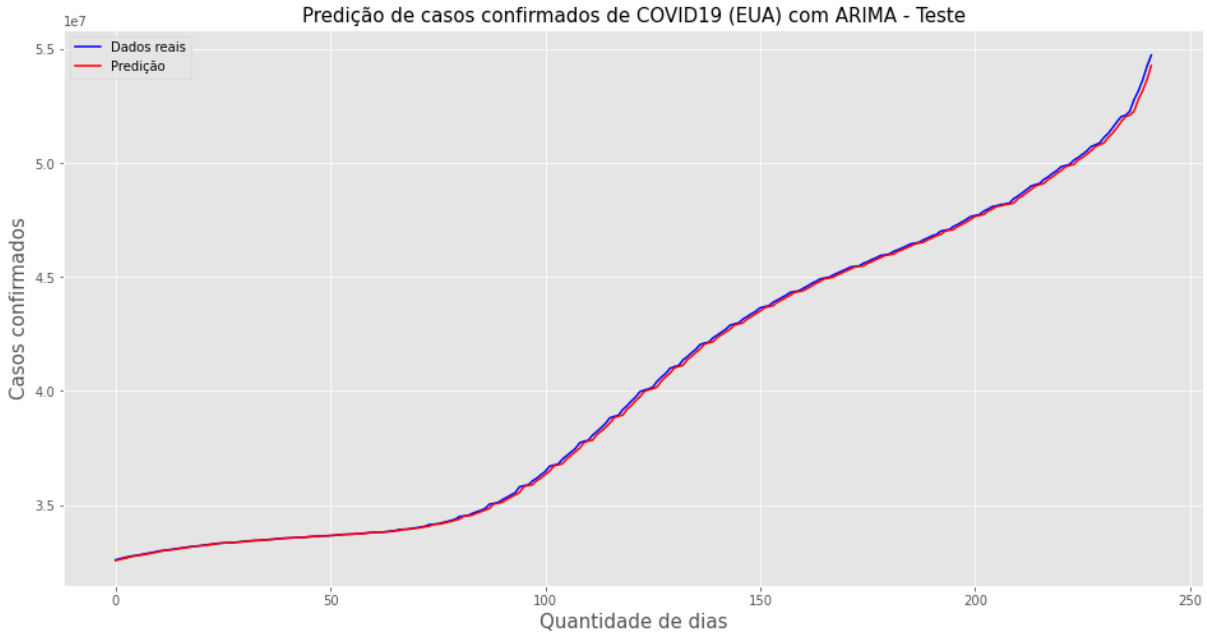


**Figura 44 – Resultado da predição de casos confirmados dos EUA com ARIMA - Teste**
**Fonte: Autor**

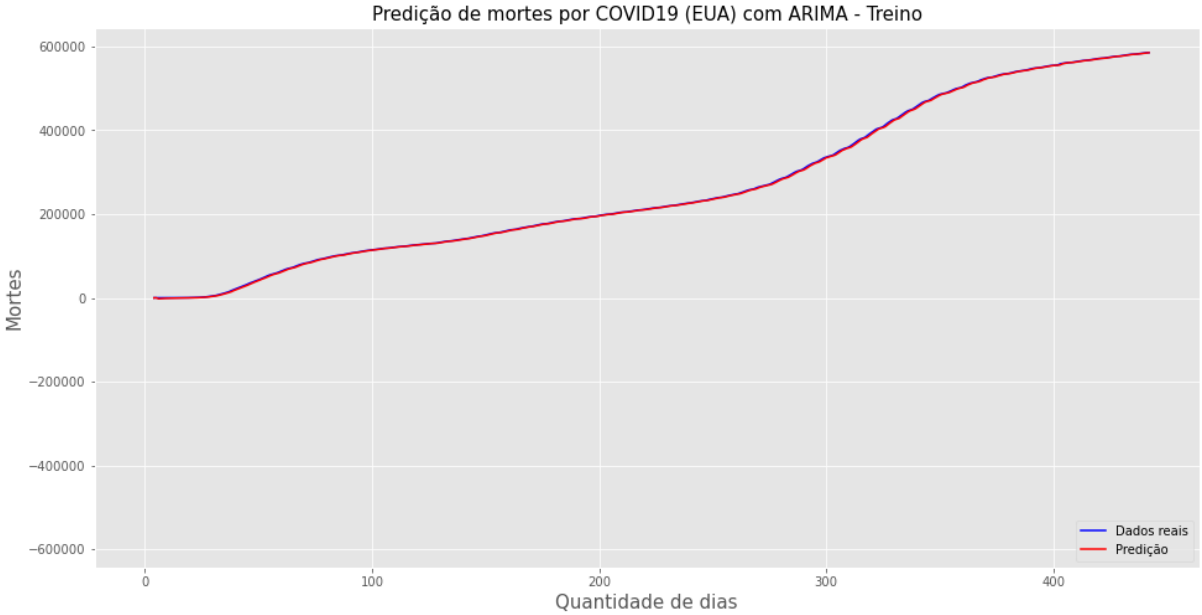


**Figura 45 – Resultado da predição de mortes dos EUA com ARIMA**
**Fonte: Autor**

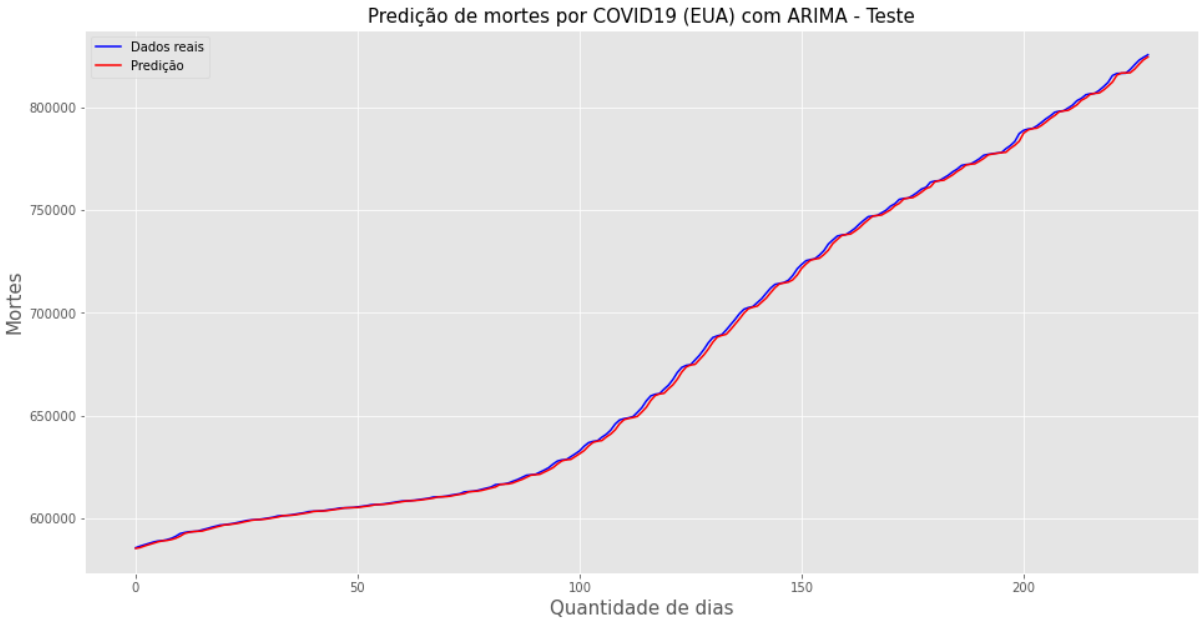


**Figura 46 – Resultado da predição de mortes dos EUA com ARIMA**
**Fonte: Autor**

**MODELO DE SUAVIZAÇÃO EXPONENCIAL**:

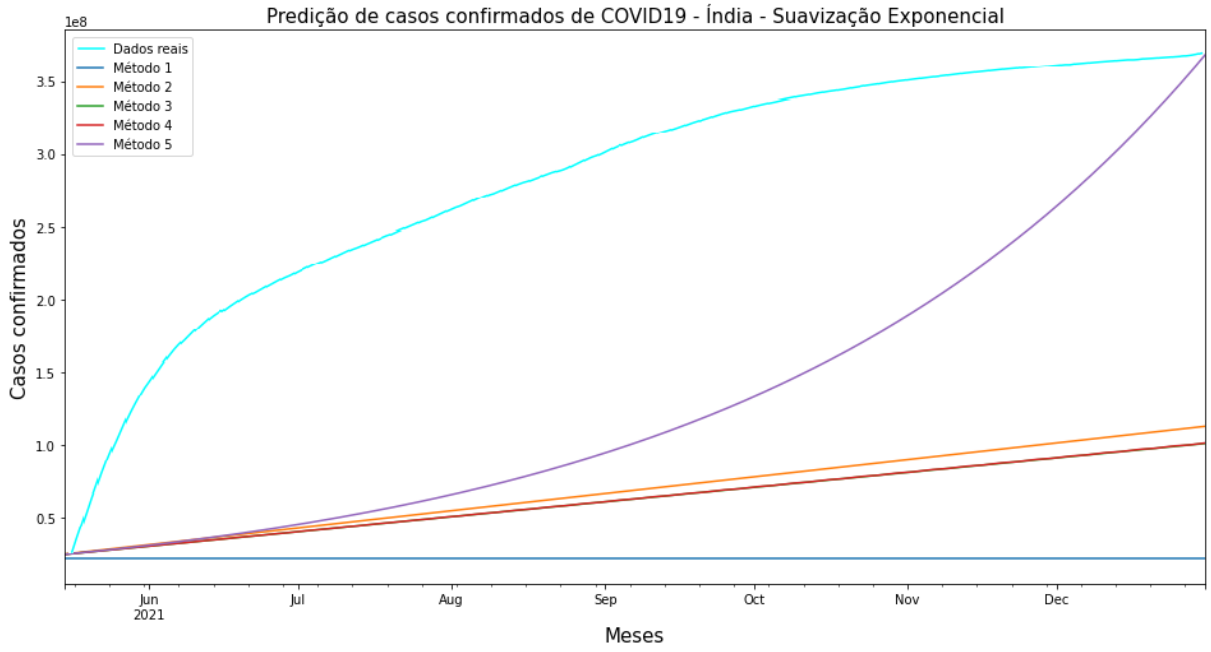


**Figura 47 – Resultado das predições de casos confirmados da Índia com Suavização Exponencial**
**Fonte: Autor**

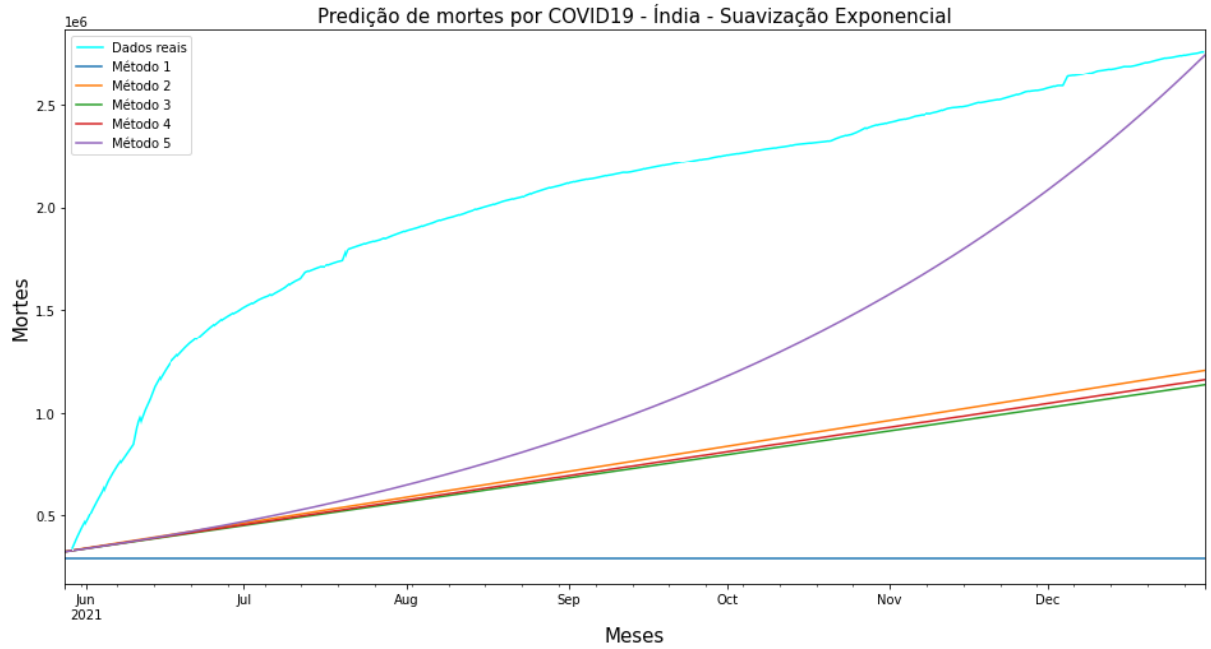


**Figura 48 – Resultado das predições de mortes da Índia com Suavização Exponencial**
**Fonte: Autor**

Acima é possível observar os resultados de predições de cinco métodos do modelo preditivo de Suavização Exponencial do país índia.

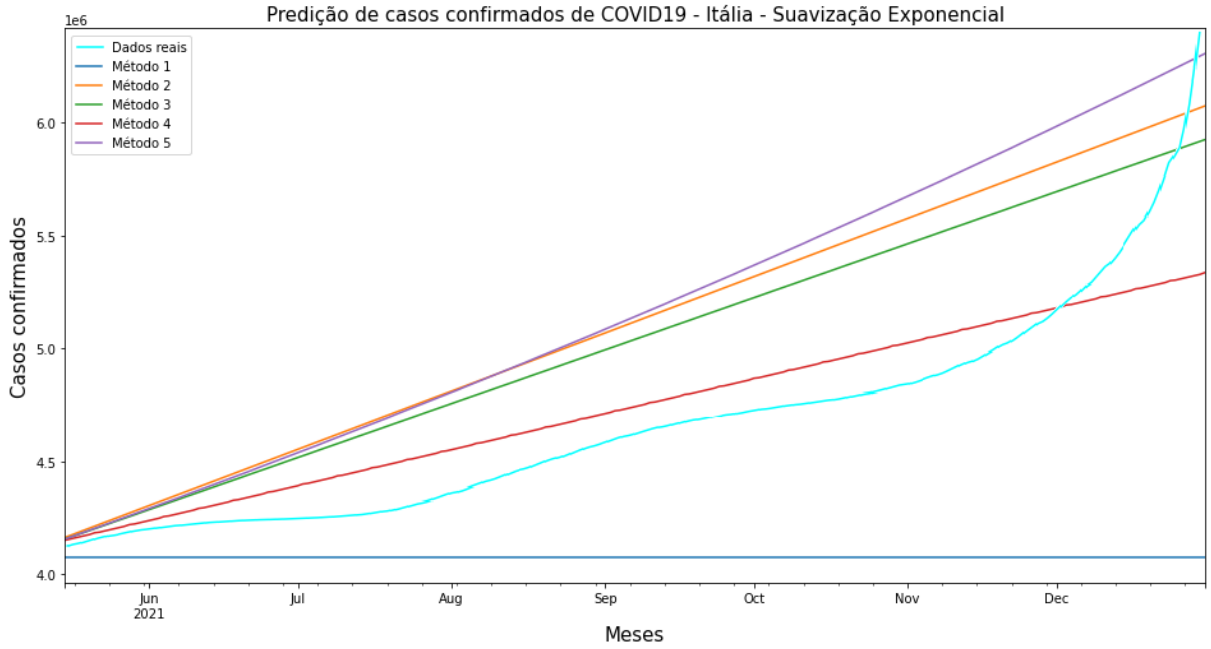


**Figura 49 – Resultado das predições de casos confirmados da Itália com Suavização Exponencial**
**Fonte: Autor**

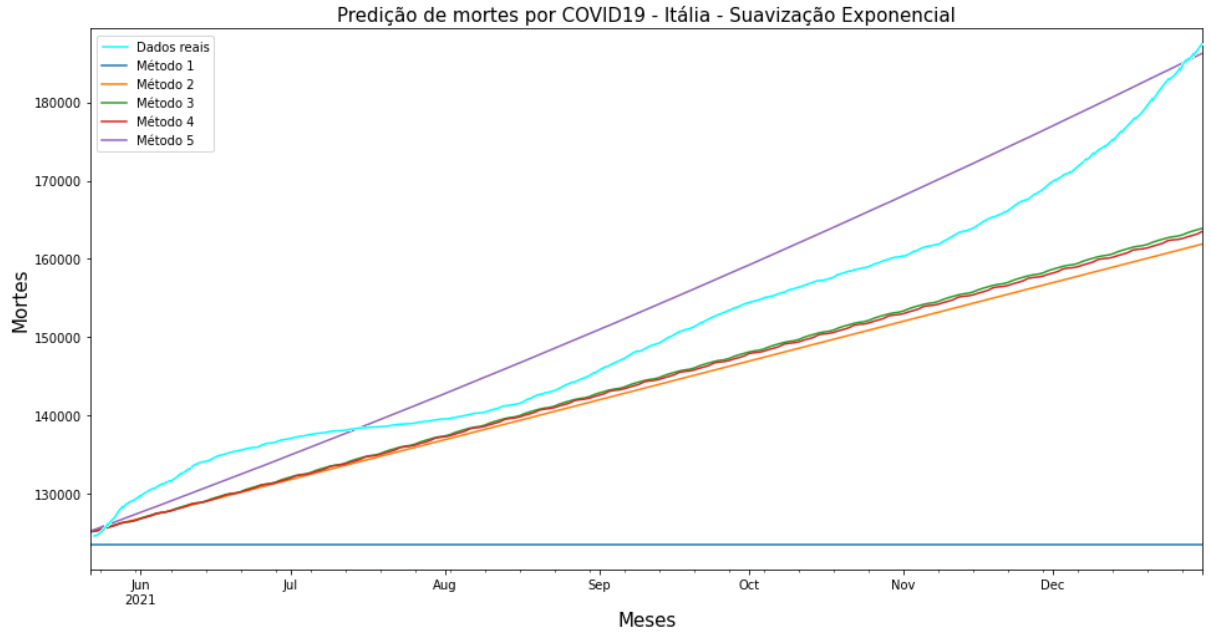


**Figura 50 – Resultado das predições de mortes da Itália com Suavização Exponencial**
**Fonte: Autor**

Acima temos os dados das predições da Itália.

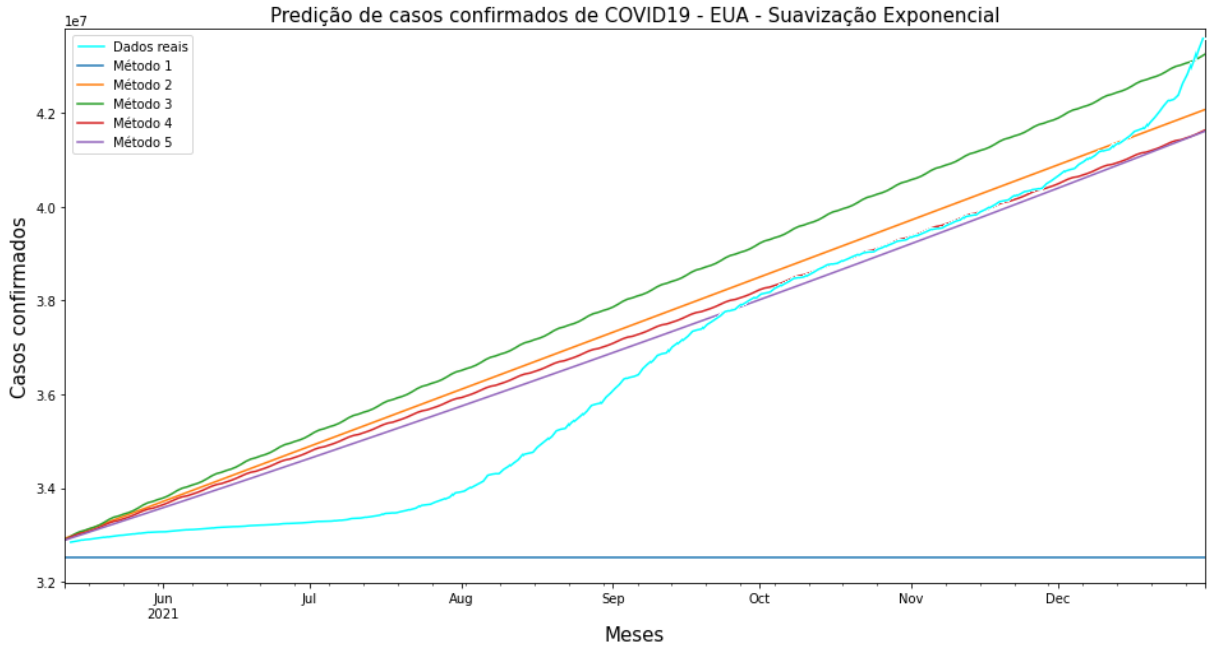


**Figura 51 – Resultado das predições de casos confirmados dos EUA com Suavização Exponencial**
**Fonte: Autor**

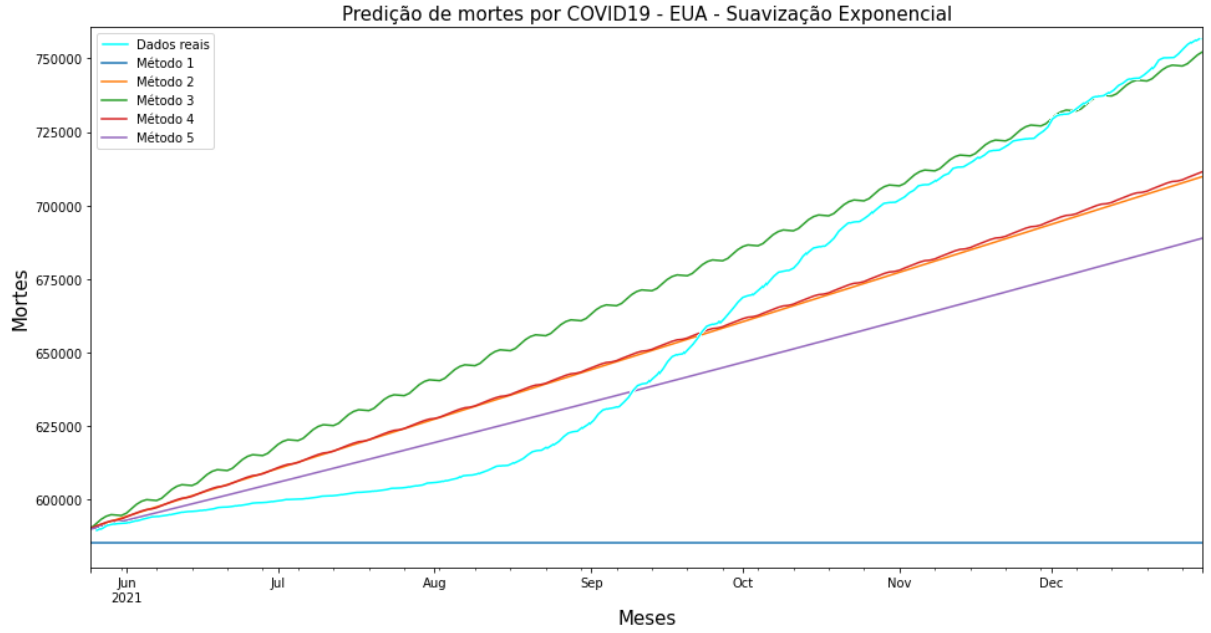


**Figura 52 – Resultado das predições de mortes dos EUA com Suavização Exponencial**
**Fonte: Autor**

Por fim, acima encontram-se os resultados do modelo de suavização para o país dos Estados Unidos da América. Nota-se que este modelo obteve resultados que variaram nas predições conforme cada método.

**ANEXO B -** Links importantes

Link do código do trabalho no GitHub:
https://github.com/sauloleite/COVID-19-time-series-predictions

Link da publicação do Artigo:
https://sbic.org.br/eventos/cbic_2021/cbic2021-72/

**ANEXO C -** Configurações do computador e bibliotecas utilizadas nas predições

**Especificações do computador utilizado nas predições**

*Sistema operacional:*
Windows 10

*Especificações do processador:*
Nome: AMD Ryzen 5 3600

Codinome: Zen2

Soquete: AM4

Fabricação em 7nm

Instruções: 64-bit

Núcleos: 6

Threads: 12

Clock: 3600 MHz

Clock (Turbo): 4200 MHz

Canais de memória: dual-channel

Memórias DDR4 @ 3200MHz

Cache: 32 Kb

PCI Express: 4.0

Canais PCI Express: 40

*Especificações da placa de vídeo:*
Nome: GTX 1660 Super

Processo de fabricação: 12nm

PCI-Express bus: 3.0

Chip Turing: TU116-300

Clock do GPU: 1530 MHz

Clock do GPU (Turbo): 1785 MHz

Tecnologia da RAM: GDDR6

Interface de largura de BUS: 192 bit

Quantidade de RAM: 6GB

Clock das memórias: 1750 MHz

Clock efetivo: 14000 MHz

Largura de banda: 336 GB/s

*Memória RAM:*
16 GB DIMM DDR4 2400Mhz

**Bibliotecas utilizadas**

- Datatime
- Keras
- Matplotlib
- Numpy
- OS
- Pandas
- Seaborn
- Sklearn
- Stasmodels
- TensorFlow
- Warnings