Universidade de Évora - Escola de Ciências e Tecnologia

Mestrado em Modelação Estatística e Análise de Dados

Dissertação

Análise e previsão do número de casos e de mortes por COVID-19 na Guiné-Bissau: Estudo comparativo entre modelos ARIMA e modelos de redes neuronais (LSTM).

Fligencio Ermogenes Gomes

Orientador(es) | Dulce Maria de Oliveira Gomes

Évora 2025

Universidade de Évora - Escola de Ciências e Tecnologia

Mestrado em Modelação Estatística e Análise de Dados

Dissertação

Análise e previsão do número de casos e de mortes por COVID-19 na Guiné-Bissau: Estudo comparativo entre modelos ARIMA e modelos de redes neuronais (LSTM).

Fligencio Ermogenes Gomes

Orientador(es) | Dulce Maria de Oliveira Gomes

Évora 2025

A dissertação foi objeto de apreciação e discussão pública pelo seguinte júri nomeado pelo Diretor da Escola de Ciências e Tecnologia:

Presidente	|	Lígia Henriques-Rodrigues (Universidade de Évora)
Vogais	|	Anabela Afonso (Universidade de Évora) Dulce Maria de Oliveira Gomes (Universidade de Évora) (Orientador)

Évora 2025

"Que todos os nossos esforços estejam sempre focados no desafio à impossibilidade.
Todas as grandes conquistas humanas vieram daquilo que parecia impossível."
(Charles Chaplin)

Agradecimentos

"Quem caminha sozinho pode até chegar mais rápido, mas aquele que vai acompanhado, com certeza vai mais longe." (Clarice Lispecton).

Começo agradecendo a Deus por, ao longo deste processo complicado, ter-me ajudado com a força nesta caminhada nos momentos em que pensei em desistir dos meus sonhos. A minha mãe, **Maria Armando Gomes,** por todo apoio e incentivo que serviram de alicerce para as minhas realizações, pela educação que me proporcionaram e por não medirem esforços para que eu chegasse até esta etapa da minha vida. A você eu devo a vida e todas as oportunidades que nela tive e espero um dia lhe retribuir os seus esforços.

A minha namorada **Sónia Isabel,** pela compreensão, paciência e companheirismo ao longo desses anos. Obrigada por ser a minha melhor amiga e sempre me incentivar a acreditar que eu seria capaz de superar mais essa importante jornada para a nossa vida, muito especial para minha querida cunhada **Dr[a] Cristina Isabel**.

Aos amigos que tive a oportunidade de conhecer e conviver intensamente durante 2 anos da faculdade: **Catarina Meireles, Luís Portugal, Ana Rita, Antero Lopes Cá, Teotóneo António**, e o meu tio **Arlindo Mendes, e meu irmão Luís Nhaga**. O companheirismo e a troca de experiências foram essenciais nesse caminho árduo. Quero deixar um agradecimento muito especial à **Catarina Meireles**, porque sem o suporte dela poderia desistir dos meus sonhos. Aos demais colegas da turma, foi um prazer encontrar vocês nessa caminhada, que possamos nos encontrar em muitas outras oportunidades da vida.

Aos professores do Departamento de Modelação Estatística e Análise de Dados da Universidade de Évora/UE, por todo conhecimento compartilhado e pela incansável dedicação para com os alunos em especial eu. Em especial a minha orientadora, professora **Dulce Gomes**, por ter feito parte deste projeto. Agradeço, também de uma forma muito especial a professora e coordenadora de curso **Anabela Afonso** pela sua paciência e dedicação para me ensinar sem qualquer incomodo.

Por fim, um agradecimento especial à Universidade de Évora pela elevada qualidade do ensino oferecido, pela oportunidade de aprendizado e crescimento tanto profissional quanto pessoal.

Resumo

Prever é algo comum no dia-a-dia de qualquer pessoa. É um factor de grande importância para a tomada de decisões e planeamento estratégico, ainda mais num ambiente de pandemia como foi o caso da Covid-19. Dessa forma, existe interesse que sejam desenvolvidas metodologias cada vez mais eficientes para gerar previsões em diferentes áreas da vida humana. Este trabalho teve por objetivo principal elaborar um estudo comparativo sobre a capacidade de modelação e preditiva entre modelos estatísticos usuais de series temporais e algoritmos de inteligência artificial para prever o número de casos e de mortes por Covid-19 na Guiné-Bissau.

Para este estudo foram escolhidos os modelos do tipo ARIMA (*AutoRegressive Integrated Moving Average*) e SARIMA (*Seasonal AutoRegressive Integrated Moving Average*), e o algoritmo de Redes Neurais Recorrentes-LSTM (*Long Short Term Memory*).

O resultado deste trabalho conside em aprofundar (bem como adquiridos novos) conhecimentos sobre e a modelação de dados através de modelos de séries temporais, nomeadamente através dos modelos do tipo ARIMA e SARIMA e dos algoritmos de Redes Neuronais Recorrentes, nomeadamente o LSTM. Espera-se ainda não só fornecer os mecanismos existentes de ajustamento e previsão de dados sequenciais no tempo, como também abrir novas perspectivas de análise de dados temporais através de novas abordagens na construção dos melhores modelos de previsão, neste estudo aplicados aos dados de Covid-19 na Guiné-Bissau, que consiste na comparação dos resultados dos modelos; com base nos resultados de previsão dos dados em análise, modelos ARIMA/SARIMA apresenta melhor previsão de que o modelo de LSTM para os números de casos e mortes por Covid-19 na Guiné-Bissau; mas que, no futuro, poderão ser aplicados a outras realidades.

Palavras-chave: Covid-19; LSTM; Modelos ARIMA/SARIMA; Previsão; Redes Neurais Artificiais (RNA); Séries Temporais.

Analysis and prediction of the number of cases and deaths per COVID-19 in Guinea Bissau: Comparative study between ARIMA models and neural network models (LSTM).

Abstract

Predicting is something common in anyone's daily life. It is a factor of great importance for decision-making and strategic planning, even more so in a pandemic environment such as the case with Covid-19. Therefore, there is interest in developing increasingly efficient methodologies to generate predictions in different areas of human life. The main objective of this work was to develop a comparative study on the modeling and predictive capacity between usual statistical time series models and artificial intelligence algorithms to predict the number of cases and deaths from Covid-19 in Guinea-Bissau.

For this study, the ARIMA (*AutoRegressive Integrated Moving Average*) and SARIMA (*Seasonal AutoRegressive Integrated Moving Average*) models and the LSTM (*Long Short Term Memory*) Neural Network algorithm were chosen.

The result of this work is to deepen (as well as acquire new) knowledge about and modeling data through time series models, namely through ARIMA and SARIMA models and Recurrent Neural Network algorithms, namely LSTM. It is also expected not only to provide existing mechanisms for adjusting and forecasting sequential data in time, but also to open new perspectives for analyzing temporal data through new approaches in building the best forecasting models, in this study applied to Covid-19 data in Guinea-Bissau, which consists of comparing the results of the models; based on the forecast results of the data under analysis, ARIMA/SARIMA models present better forecasts than the LSTM model for the numbers of cases and deaths from Covid-19 in Guinea-Bissau; but which, in the future, could be applied to other realities.

Keywords: Covid-19; LSTM; ARIMA/SARIMA models; Forecast; Artificial Neural Networks (ANN); Time Series.

Índice

Resumo iii

Abstract ..... **iv**

Lista de Figuras ..... **viii**

Lista de Tabelas ..... **ix**

Lista de Acrónimos ..... **x**

1. Contextualização do Estudo **12**

1.1 Introdução 12

1.2 Motivação e importância do trabalho 14

1.3 Estrutura do trabalho 15

1.4 Caracterização geográfica da Guiné-Bissau 16

1.5 Situação social e económica da Guiné-Bissau 16

1.6 Sistema de saúde da Guiné-Bissau 17

1.7 Panorama geral da pandemia por coronavírus/Covid-19 na Costa Ocidental África "Subsaariana". 17

1.8 Pandemia de Covid-19 e medidas implementadas pelo Governo na prevenção e combate do vírus na Guiné-Bissau 18

2. Séries Temporais **20**

2.1 Definições gerais de séries temporais 20

2.2 Estacionariedade e ruído branco 21

2.3 Função de autocorrelação e autocorrelação parcial 23

2.4 Estimação de medidas de uma série temporal 24

2.5 Séries temporais não estacionárias 24

2.6 Análise de estacionariedade de uma série temporal 26

2.6.1 Teste de Dickey-Fuller 27

2.6.2 Teste de Dickey-Fuller Aumentado 28

2.6.3 Teste de Kwiatkowski, Phillips, Schmidt e Shin 28

3. Modelos de Séries Temporais **29**

3.1 Modelos lineares de séries temporais 29
3.2 Modelos lineares estacionários 29
3.2.1 Modelo autorregressivo - AR (*p*) 29
3.2.2 Modelo de médias móveis - MA(*q*) 31
3.2.3 Modelo autorregressivo e médias móveis - ARMA(*p, q*) 33
3.3 Modelos lineares não estacionários 34
3.3.1 Modelo integrado autorregressivo e médias móveis - ARIMA(*p,d,q*) 35
3.3.2 Modelo integrado autorregressivo e médias móveis sazonal – SARIMA *(p,d,q)(P,D,Q)s* 36
3.4 Etapas na modelação de uma série temporal - metodologia de Box-Jenkins 36
3.5 Previsão e critérios de avaliação 39
4. Redes Neuronais Artificiais e Modelos de Memória de Longo Prazo 43
4.1 Introdução 43
4.2 Redes neuronais artificiais 43
4.3 Treino e aprendizagem de RNAs 47
4.4 Funções de ativação 49
4.3.1 Função Sigmoide 51
4.3.2 Função Tangente Hiperbólica 51
4.3.3 A Função Linear Retificada (*Rectified Linear Unit* - ReLU) 52
4.5 Algoritmos de memória de longo prazo **53**
5. Comparação de modelos de séries temporais e algoritmos de redes neuronais (LSTM) 57
5.1 Introdução **57**
5.2 Número de infeções por Covid-19 ..... **58**
5.2.1 Análise descritiva do número de infeções por Covid-19 58
5.2.2 Modelação e previsão a partir de modelos do tipo ARIMA/SARIMA 60
5.2.3. Modelação e previsão apartir do algoritmo LSTM **63**
5.3 Número de mortes por Covid-19 ..... **65**
5.3.1 Análise descritiva do número de mortes por Covid-19 65

5.3.2 Modelação e previsão a partir de modelos do tipo ARIMA/SARIMA.....67

5.3.3 Modelação e previsão a partir do algoritmo LSTM 70

5.4 Comparação entre os modelos ARIMA e o algoritmo LSTM 72

6. Considerações finais, limitações e trabalho futuro.....73

Referências Bibliográficas. 75

Lista de Figuras

Figura 1.1. Mapa da Guiné-Bissau....16
Figura 4.1. Estrutura básica de um neurónio biológico....44
Figura 4.2. Rede neurónio artificial....45
Figura 4.3. RNA recorrente e unidirecional....46
Figura 4.4. RNA recorrentes com duas camadas ocultas....47
Figura 4.5. Função limiar....50
Figura 4.6 Função limiar por partes....50
Figura 4.7. Função sigmoide....51
Figura 4.8. Função tangente hiperbólica....52
Figura 4.9. Função ReLU....52
Figura 4.10. Rede de memória de longo prazo (LSTM)....55
Figura 5.1. Caixa de bigodes do número de casos por mês....59
Figura 5.2. Média do número de casos por mês....59
Figura 5.3. Variância do número de casos médios por mês....60
Figura 5.4. Séries de número de casos por semanas....60
Figura 5.5. FAC e FACP da série do número de infetados por SARS-CoV-2....61
Figura 5.6. Resíduos do modelo AR(1) ajustado à série do número de casos semanais....61
Figura 5.7. Histograma dos resíduos do número de casos semanais....62
Figura 5.8. Previsão e ajustamento do número de casos semanais por Covid-19....62
Figura 5.9. Ajustamento e previsão do número de casos através do algoritmo LSTM....64
Figura 5.10. Caixa de bigodes do número de mortes por mês....66
Figura 5.11. Média do número de mortes por mês....66
Figura 5.12. Variância do número de mortes por mês....67
Figura 5.13. Série de número dos óbitos por mês....67
Figura 5.14. FAC e FACP do número de mortes semanais....68
Figura 5.15. FAC e FACP da série transformada do número de mortes semanais....68
Figura 5.16. Resíduos do modelo ARIMA(2,1,0) ajustado à série do número de mortes semanais....69
Figura 5.17. Histograma dos resíduos do modelo ARIMA(2,10) ajustado ao número de mortes semanais....69
Figura 5.18. Previsão e ajustamento do número de mortes semanais por Covid-19....70
Figura 5.19. Previsão e ajustamento do número de mortes através do algoritmo LSTM....71

Lista de Tabelas

Tabela 3.1. Propriedades das funções FAC e FACP. 34
Tabela 3.2. Fluxograma de etapas da metodologia de Box-Jenkins para as séries temporais. 42
Tabela 5.1. Medidas descritiva do número de casos de Covid-19 por ano. 58
Tabela 5.2. Valores das medidas de precisão do ajustamento do número de casos semanais. 63
Tabela 5.3. Modelos de treino de LSTM. 64
Tabela 5.4. Medidas de precisão do ajustamento através do algoritmo LSTM para os casos por Covid-19 65
Tabela 5.5. Descrição estatística dos dados não transformados. 65
Tabela 5.6. Precisão da previsão do número de mortes semanais 70
Tabela 5.7. Modelos de treino de LSTM 71
Tabela 5.8. Precisão do ajustamento dos resultados do modelo LSTM relativos às mortes por Covid-19 71
Tabela 5.9. Resumo dos resultados dos diferentes métodos para a série dos casos…… … 72
Tabela 510. Resumo dos resultados dos diferentes métodos para a série das mortes……. 72

Lista de Acrónimos

AR- Autoregressivo

ARIMA- Autoregressivo integrado e médias móveis

DG- Director Geral

FAC- Função de autocorrelação

FACP- Função de autocorrelação parcial

GB- Guiné-Bissau

INE- Instituto Nacional de Estatística

LSTM- Rede de memória de longo prazo

MA- Médias móveis

MAE- Erro absoluto médio

ME- Erro médio

MAPE- Erro absoluto médio percentual

MPE- Erro médio percentual

ONU- Organização das Nações Unidas

OMS- Organização Mundial de Saúde

SARIMA- Autoregressivo integrado e médias móveis com sazonalidade

RMSE- Raiz quadrada do erro quadrático médio

RNN- Rede neuronal recorrente

1

Contextualização do Estudo

1.1 Introdução

Após mais de 100 anos da eclosão da pandemia de gripe de 1918-1919, conhecida como Gripe Espanhola, estivemos, em 2019, perante uma nova pandemia no mundo provocada pelo vírus da SARS-CoV-2, que ficou conhecida por Covid-19. De momento (abril de 2024) a pandemia já foi declarada extinta, mas o vírus continua presente por todo o mundo e a ter efeitos nefastos, como ainda a provocar imensas mortes. Segundo dados mais recentes da Organização Mundial de Saúde (OMS) em termos globais "[…] o número de novos casos diminuiu 58% durante o período de 28 dias de 8 de janeiro a 4 de fevereiro de 2024, em comparação com o período anterior de 28 dias, com mais de 503 000 novos casos notificados.

O número de novas mortes diminuiu 31% em comparação com o período anterior de 28 dias, tendo sido notificadas mais de 10 000 novas mortes. Até 4 de fevereiro de 2024, foram notificados mais de 774 milhões de casos confirmados e mais de sete milhões de mortes em todo o mundo." (https://www.who.int/publications/m/item/covid-19-epidemiological-update-16-february-2024) Devido à capacidade de transmissão do vírus e às suas mutações genéticas o mundo terá de encontrar soluções para que a população possa conviver com este vírus como convive com, por exemplo, o vírus da gripe. Daí, a utilidade em continuar a estudar-se o número de casos e de mortes por Covid-19.

Em dezembro de 2019, na província de Wuhan, na região central da China, foi detectado o primeiro caso oficial do vírus que provoca a doença designada por Covid-19. O mundo surpreendeu-se, assim, com um novo vírus da família dos coronavírus que teve uma rápida propagação espacial pelo mundo e que provocou uma das mais impactantes pandemias do século XXI [1].

No dia 24 de janeiro de 2020, a França reportava o primeiro caso de Covid-19 na Europa, assim começa a sua contagem progressiva de números de infetados de uma forma espacial e exponencial pelo continente europeu (https://news.un.org/pt/story/2022/01/1777352).
Os primeiros dados oficiais das infecções por Covid-19 na África Subsaariana foram registados em Lagos, capital económica da Nigéria, a 27 de fevereiro de 2020 e em Dakar, capital do Senegal, a 3 de março de 2020.

Na Guiné-Bissau os primeiros casos oficiais foram registados no dia 25 de março de 2020 na cidade de Bissau e foram de dois cidadãos estrangeiros, ambos de funcionários de ONU. Face à situação de emergência o Presidente da República Umaro Sissoco Embalo decretou, a 27 de março de 2020, o estado de emergência como forma de impedir a propagação de vírus de uma forma espacial para as outras regiões do país [1].

A 11 de março de 2020, a OMS declara o Covid-19 como pandemia com a justificação de que tinham sido atingidos os níveis alarmantes de propagação e inalação em todo o continente [1]. Nesse mesmo dia, Tedros Adhanom Ghebreyesus, Director-Geral da OMS, afirmou que "o mundo já tem mais de 118 mil infecções com 4291 mortes, a maioria na China, onde a doença surgiu; a Covid-19 está presente agora em 114 países; é a primeira vez que uma pandemia é decretada devido a um coronavírus". (https://news.un.org/pt/story/2020/03/1706881).
A Guiné-Bissau, país onde este estudo se irá centrar, registou entre 26 de março de 2020 e 15 de abril de 2023 um total de 8960 casos de Covid-19, que resultaram em 176 mortes. (https://www.worldometers.info/coronavirus/country/guinea-bissau/)

Os dados que serão utilizados neste trabalho são séries temporais semanais do número de casos e de mortes por Covid-19 na Guiné-Bissau, entre 26/03/2020 e 15/04/2023, que corresponde a 159 semanas. Os dados são provenientes de banco de dados de OMS de Covid-19 no mundo e disponíveis em https://covid19.who.int/region/afro/country/gw.

Estes são os dados disponíveis e, portanto, são com estes que iremos trabalhar. Estamos cientes de que este tipo de dados é muito complexo porque cada país tem diferentes estruturas de saúde e diferentes capacidades logísticas e económicas de lidar com a pandemia. Tendo em conta que a Guiné-Bissau é um dos países mais pobres do mundo, com uma elevada instabilidade política, um baixo nível de condições sanitárias e com uma economia e estuturas deficientes, as interpretações dos modelos ajustados e das correspondentes previsões terão de ser feitas com o cuidado necessário devido a possível falta de confiabilidade dos dados.

1.2 Motivação e importância do trabalho

Para controlar a periodicidade da Covid-19 de forma mais eficiente é necessário monitorizar e prever a prevalência da doença e a sua distribuição espacial e temporal.

Os modelos tradicionais de séries temporais, nomeadamente os modelos ARIMA (*Autoregressive Integrated Moving Average)* e SARIMA *(Integrated auto-regressive and moving averages with seasonality*), têm sido aplicados com sucesso para estimar a incidência e prevalência da mortalidade por gripe, malária, hepatite B, e outras doenças infeciosas ou em estudos temporais epidemiológicos. Estes modelos matemáticos mostraram a sua eficácia na modelação da estrutura de dependência temporal de uma série, dadas as suas alterações de tendências, periódicas e as distorções aleatórias, e na previsão [2]. Mostram-se também apropriados na previsão do impacto do surto da Covid-19, e na tomada de decisões necessárias para responder à crise epidemiológica mundial na distribuição de recursos humanos e na gestão económica [2].

Através destes modelos de previsão, poderá também aferir-se se as medidas aplicadas nos períodos referidos potenciaram o comportamento esperado no surgimento de novos casos e mortes por Covid-19 [3].

Contudo, existem diferentes técnicas e métodos possíveis de ajustamento e previsão para os dados sequenciais discretos, através dos modelos tradicionais de séries temporais, alisamento exponencial, ou através dos algoritmos de inteligência artificial como RNNs-LSTM (*Recurrent Neural Networks - Long Short-Term Memory)*.

Muitas perguntas inquietam os investigadores/estudantes quanto à aplicação de um determinado método de previsão, em relação a outro.

Assim, este trabalho tem como o principal objetivo avaliar e comparar diferentes metodologias que melhor expliquem a variabilidade do número de infectados e de mortes por Covid-19 na Guiné-Bissau, ao longo das semanas em análise. Para tal, irá usar-se os métodos de previsão de séries temporais como ARIMA/SARIMA e comparar estes com o algoritmo de Redes Neuronais, RNNs, em especial o LSTM.

O algoritmo LSTM é baseado em aprendizagem profunda de inteligência da máquina, ou artificial, e consegue identificar estruturas, ou padrões, de dados não-lineares e lineares e, através destas, proceder à previsão de valores futuros. Este algoritmo tem sido também usado na previsão dos resultados de propagação de vírus e na previsão de números infetados e mortes por Covid-19 [4].

Em resumo, a motivação para este trabalho baseia-se nestas questões:

- Como descrever e prever o número de casos e de mortes por Covid-19 na Guiné-Bissau?
- Porque é que se utiliza um modelo em relação a outro?
- Como funcionam os algoritmos de inteligência artificial?

- Qual o desempenho de cada uma das abordagens, tanto as tradicionais como a de RNNs-LSTM, na previsão dos dados?

Estas são algumas das questões que motivaram a realização desta pesquisa de conclusão de mestrado, na procura de compreender a evolução da Covid-19 entre 2020 e 2023 na Guiné-Bissau, o meu país.

Com este trabalho pretende-me não só aprofundar os conhecimentos e métodos sobre os modelos de série temporal, como também adquirir novos conhecimentos sobre redes neuronais, na esperança de num futuro próximo serem aplicados na Guiné-Bissau na compreensão da evolução das demais doenças infeciosas, que assolam a Costa Ocidenteal da África e, assim, contribuir para a melhoria das políticas de saúde pública neste país e na região.

1.3 Estrutura do trabalho

O trabalho está estruturado em 6 capítulos:

- **Capítulo 1:** Contextualização do tema, motivação e objetivos do estudo. Caracterização geográfica da Guiné-Bissau, o sistema de saúde, situação social e económica. Panorama da pandemia na África Subsaariana e na Guiné-Bissau.
- **Capítulo 2:** Conceitos gerais de processos estocásticos e de séries temporais.
- **Capítulo 3:** Modelos lineares para séries temporais, estacionárias e não estacionárias, metodologia Box-Jenkins na modelação de uma série temporal e previsão de valores futuros.
- **Capítulo 4:** Conceitos gerais sobre redes neuronais.
- **Capítulo 5:** Apresentação dos principais resultados do estudo relativo ao ajustamento e previsão através dos modelos ARIMA/SARIMA e algoritmos RNNs, em especial através do LSTM.
- **Capítulo 6:** Considerações finais, as limitações do trabalho e as recomendações para estudos futuros.

1.4 Caracterização geográfica da Guiné-Bissau

A Guiné-Bissau (GB) fica situada na parte Subsaariana do continente africano, é banhada pelo oceano Atlântico, com uma extensão territorial de 36.125km^2, é composta por uma parte continental e outra insular com aproximadamente 88 ilhas e ilhéus, dos quais 17 são habitados. A GB faz fronteira com o Senegal ao Norte e a República da Guiné-Conacri ao Sul e Leste [5], e a sua capital é Bissau, a cidade mais populosa do país.

Figura 1.1. Mapa da Guiné-Bissau.

Fonte: https://news.un.org/pt/story/2015/02/1501211

De acordo com INE-Censos 2009 o país possuía 1.520.830 habitantes, dos quais cerca de 80% viviam nas zonas rurais e 20% no sector autónomo de Bissau. A taxa de crescimento populacional anual é de 3%, com uma taxa de alfabetização de 60% [6].

1.5 Situação social e económica da Guiné-Bissau

Um estudo realizado em 2020 pelo Banco Africano de Desenvolvimento, classificou a GB como sendo um dos países mais pobre do mundo, em que aproximadamente 67% da sua população vive em pobreza abaixo de limiar com 1,70€ por dia/família, e 33% vivem em pobreza extrema. A taxa de desemprego é de 11,6% [7]. A pandemia tornou ainda mais perceptível todas as fragilidades do país e agravou ainda mais os seus problemas económicos e sociais.

1.6 Sistema de saúde da Guiné-Bissau

"O sistema nacional de saúde da GB é composto pelos sectores público, privado, convencionado (Igreja Católica, Estado e organizações não governamentais) e pelo setor tradicional (medicina tradicional)", esta também conhecida como medicina tradicional ou *Kurandeiros/Djambakus*. "O serviço público de saúde está organizado administrativamente em 11 regiões de saúde, que por sua vez são subdivididas por 36 sectores de saúde, pelo sector autónomo de Bissau e 114 distritos sanitários" [8].

"Mais de 40% da população da GB vive uma distância superior a 5 km das poucas instituições públicas de saúde existentes no país" [8]. Os preços elevados no acesso à saúde e a falta de todo o tipo de profissionais de saúde fazem com que uma das populações mais pobre do mundo não tenha acesso a cuidados médicos adequados.

Em 2014 a quantidade de profissionais de saúde existentes na GB era insignificante para a população total do país: $1.7 \approx 2$ médicos e $5.9 \approx 6$ enfermeiros para cada 10 mil pessoas. Em 2017 os médicos especialistas existentes em todo o território nacional eram apenas: três (3) pediatras, quatro (4) obstetras, trinta e quatro (34) parteiras qualificadas e apenas um (1) anestesista [9].

1.7 Panorama geral da pandemia por coronavírus/Covid-19 na Costa Ocidental África "Subsaariana".

De acordo com Monié [1], os primeiros casos de coronavírus na Costa Ocidental da África foram confirmados em Lagos, capital económica da Nigéria, no dia 27 de fevereiro de 2020; neste contágio progressivo, a segunda confirmação do contaminante pelos vírus novos de coronavírus foram confirmados em Dakar-Senegal no dia 02 de março de 2020, assim começa a contagem dos casos de coronavírus na Costa Ocidenta da África.

Segundo Monié [1], os primeiros epicentros da pandemia de coronavírus no mundo foram nas maiores redes produtivas, logísticas e comercias que compõem o espaço global de fluxos económico e social: Wuhan, Beijing, Singapura, Milão, Paris, Londres, Nova Iorque, e São Paulo. Como sistema nacional e internacional de transporte constitui um dos vectores para a propagação de vírus numa escala nacional e internacional para os países desenvolvidos; o sistema mundial de transporte aéreo, com uma média 44000 voos e 2,7 milhões de passageiros por dia, constituiu o principal vetor de difusão do vírus nas escalas intercontinentais.

Estes são motivos pelos quais propagação de vírus na África Subsaariana foi mais tardio por não fazer partes de redes económicos mundiais, nos casos específico da Costa Ocidental da África por ser uma das regiões com os países mais pobres de mundo que poucos voos internacionais de intercâmbios econômicos.

1.8 Pandemia de Covid-19 e medidas implementado pelo Governo na prevenção e combate do vírus na Guiné-Bissau

Face à situação epidemiológica da pandemia de Covid-19 no mundo, o governo de GB aplicou vários níveis de resposta, como o estado de emergência para determinação das medidas e protocolos preventivos de propagação da pandemia. O primeiro estado emergência decretado pelo Presidente da República foi no dia 27 de março de 2020, onde foram decretadas medidas extraordinárias de contenção para a mitigação e propagação de vírus, como a proibição de deslocações não essenciais, de todos os ajuntamentos públicos, as escolas, lojas, e outros locais de vendas não essenciais foram encerradas, decretou-se o fechamento de mercados às 11 horas de manhã em todo o território nacional e os movimentos fronteiriços e os voos foram fortemente restringidos.

O estado de emergência foi renovado pela quinta (5) vez, tendo terminado a 24 de agosto de 2020, tendo sido substituído a 27 de agosto pelo estado de calamidade em todo o território nacional com algumas medidas diferentes, e menos restritivas dos estados de emergência.
O governo estabeleceu esta estratégia de levantamento das medidas de confinamento decretadas nos estados de emergência com o objetivo de iniciar a fase de recuperação e revitalização da vida em sociedade e da economia nacional de país. O estado de calamidade foi prorrogado muitas vezes e terminou a 9 de setembro de 2021, porque os níveis de pobreza estavam muito mais elevados do que estavam antes da pandemia. A partir do dia 21 de outubro de 2021 foi declarado o estado de alerta que vigorou até 5 de março de 2022, que foi o período em que o país recebeu o maior lote de doses de vacinas para a Covid-19, e considerado como a fase de desconfinamento. (https://dadosbissau.maps.arcgis.com/apps/dashboards/85cd1b7a80b94c4dbc32302006374a83)

Os períodos de pandemia foram marcados por dois momentos, o período ativo (estados de emergência e calamidade) e o periodo pós-pandemia (estado de alerta). Durante todo este período, de 26 de março de 2020 a 15 de abril de 2023, o número de infetados por Covid-19 na GB, foi de 8960, dos quais 176 morreram. A 20 de maio de 2023 existiam 508 pessoas infectadas. O total de

pessoas vacinadas, com as duas doses, até 4 de junho de 2023 foi de 41933. (https://www.worldometers.info/coronavirus/country/guinea-bissau/)

2

Séries Temporais

2.1 Definições gerais de séries temporais

Neste capítulo serão apresentados alguns conceitos básicos de séries temporais, os seus objectivos, e abordagens a séries temporais estacionárias e não estacionárias.

Uma série temporal é uma sucessão de observações ordenadas no tempo, em geral igualmente espaçadas, onde a sua ordem desempenha um papel fundamental.
Mais concretamente, uma série temporal é uma realização de um processo estocástico $\{X_t, t \in \mathcal{T}\}$, que assume valores nos reais e onde o tempo $\mathcal{T}$ é discreto.

Ao conjunto $\mathcal{T}$ dá-se o nome de espaço parâmetro. Se $\mathcal{T} = \mathbb{N}\ ou\ \mathcal{T} = \mathbb{Z}$ diz-se que o processo é de tempo discreto, e se $\mathcal{T} = \mathbb{R}\ ou\ \mathbb{R}^+$ diz que o processo é de tempo contínuo. Aos valores que as variáveis aleatórias tomam dá-se o nome de espaço de estados. As séries temporais podem ser contínuas ou discretas em função do espaço de estados. Podem ser também classificadas como univariadas, quando são constituídas por uma só variável, ou como multivariadas, quando se observam mais do que uma variável em cada instante de tempo.

Uma série temporal pode ser decomposta em quatro componentes de variação ao longo de tempo.

- Tendência: a inclinação, positiva ou negativa, que certas séries apresentam ao longo do tempo;
- Sazonalidade: o aumento ou decréscimo que ocorre regularmente em iguais períodos, originando oscilações que se repetem ao longo de tempo;
- componente cíclica: um padrão de flutuação que não apresenta uma periodicidade definida, isto é, a sua duração não é fixa e, portanto, o seu comprimento varia frequentemente de ciclo-a-ciclo, nem apresenta uma causa atribuída a fatores sazonais;

- componente residual: contém qualquer variação não explicada pelas outras componentes, e representa o ruído aleatório [10].

Em geral, considera-se que as componentes cíclicas se encontram incorporada na tendência.

A análise de séries temporais tem como objetivo descrever os dados em estudo, modelá-los e fazer previsão de valores futuros. Para se proceder à modelação e à posterior previsão é necessário descobrir o modelo que melhor descreve o comportamento dos dados que são observados ao longo de um tempo, isto é, estabelecer uma equação que explica a variabilidade (ou comportamento) dos dados ao longo do tempo à custa das relações de dependência entre as observações do presente e passado.

2.2 Estacionariedade e ruído branco

Seja $\{X_t, t \in \mathbb{Z}\}$ uma série temporal.
Existem duas formas de caracterizar uma série temporal. Através da identificação da distribuição de probabilidade conjunta das variáveis aleatórias $\left(X_{t_1}, X_{t_2}, \dots, X_{t_n}\right), t_1, t_2, \dots, t_n \in \mathbb{Z}\ (n \geq 1)$. Ou, mais simples, através dos momentos de 2ª ordem [11].

- Média: $E\left[X_{t_i}\right] = \mu_{t_i},\ t_i \in \mathbb{Z}$

- Variância: $\mathrm{Var}\left[X_{t_i}\right] = \sigma_{t_i}^2,\ t_i \in \mathbb{Z}$

- Covariância $\gamma(t_i, t_j) = E\left[\left(X_{t_i} - \mu_{t_i}\right)\left(X_{t_j} - \mu_{t_j}\right)\right],\ t_i, t_j \in \mathbb{Z}$

sendo $\mathrm{Var}\left[X_{t_i}\right]$ um caso particular da função de autocovariância quando $t_i = t_j$.

Dado que uma série temporal é apenas uma realização de um processo estocástico (uma família de variáveis aleatórias), é necessário que esta se mantenha estável ao longo do tempo. Isto é importante para que a média temporal seja uma boa aproximação da média do processo. Daí a importância da estacionariedade da série.

Os conceitos de série temporal fortemente estacionária e fracamente estacionária são as definições apresentados por Gonçalves e Lopes [11]. Iremos também usar a designação de fracamente estacionária quando nos estivermos a referir ao processo estocástico gerador dessa série temporal.

Uma série temporal $\{X_t, t \in \mathbb{Z}\}$, diz **estritamente estacionária** (ou **estacionáridade forte**) se todas as distribuições finito-dimensionais permanecem as mesmas sob translações no tempo,

$$F(X_{t_1}, X_{t_2}, \dots, X_{t_n}) = F(X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k})$$

para qualquer $t_1, t_2, \dots, t_n \in \mathbb{Z}$ $(n \geq 1)$ e $\forall k \geq 0$.

Uma vez que na prática esta propriedade é de difícil demonstração introduzimos uma definição mais fraca de estacionariedade.

Uma série temporal $\{X_t, t \in \mathbb{Z}\}$, diz **fracamente estacionária** (ou **estacionáridade de 2.ª ordem ou fraca**) se todos os momentos até à 2.ª ordem de $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ existem e são iguais aos momentos correspondentes até à 2.ª ordem de $(X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k})$, para qualquer inteiro positivo n e para quaisquer inteiros $t_1, t_2, \dots, t_n$, ou seja, podemos afirmar que uma série temporal é fracamente estacionária se para todo o t, com $E[X_t^2] < +\infty$, $\forall k \geq 0$, $t \in \mathbb{Z}$

- a média não depende do tempo, $E[X_t] = \mu$;
- a variância não depende do tempo, $Var[X_t] = \sigma^2$;
- e a covariância entre X_t e X_s (autocovariância) depende apenas do desafazamento $|t - s|$, $Cov(X_t, X_s) = Cov(X_{t+k}, X_{s+k}) = \gamma_{|t-s|} = \gamma_k,\ s, k \in \mathbb{Z}$.

Estacionaridade forte implica a estacionaridade fraca, desde que a série possua segundos momentos finitos. Mas a estacionariedade fraca não implica a forte a não ser que a série seja Gaussiana.

Ao longo do trabalho, sempre que nos referirmos a estacionariedade estamos a designar estacionariedade de 2ª ordem.

Um caso particular de uma série estacionária é a chamada série de ruído branco.

Uma série temporal $\{\varepsilon_t, t \in \mathbb{Z}\}$, diz-se que é uma série de **ruído branco** quando possui média e variância constantes ao longo do tempo e possui correlação nula. Ou seja, se só se satisfaz as seguintes propriedades:

- $E[\varepsilon_t] = \mu$ (usualmente considera-se que $\mu = 0$)
- $Var[\varepsilon_t] = \sigma_\varepsilon^2$
- $Cov(\varepsilon_t, \varepsilon_{t+s}) = \gamma_s = 0;\ s = 1, \pm 2, \pm 3, \dots$

2.3 Função de autocorrelação e autocorrelação parcial

Seja $\{X_t, t \in \mathbb{Z}\}$ uma série temporal estacionária. A **função de autocorrelação** (também designada por FAC) é definida como

$$\rho_k = Corr(X_t, X_{t+k}) = \frac{Cov(X_t, X_{t+k})}{\sqrt{Var[X_t]Var[X_{t+k}]}} = \frac{\gamma_k}{\gamma_0}, \qquad k \in \mathbb{Z}$$

As funções de autocovariância (definida anteriormente) e de autocorrelação são muito importantes na análise de uma série temporal porque medem a variabilidade da série temporal ($\gamma_0 = Var[X_t]$) e medem como variam (co-variam) de forma linear duas observações desfazadas de um tempo *k* (também designado por amplitude ou *lag*). Também se pode dizer que a FAC mede se um valor observado no instante *t* se encontra correlacionado com que valores observados no passado. Ou seja, até onde no tempo se regista uma correlação entre as aobservações.

Como à medida que o tempo entre as observações aumenta, diminui a correlação entre as observações (a memória da série diminiu) a FAC de uma série estacionária decresce à medida que o *lag* aumenta. Outra propriedade é que a FAC é uma função par e toma valores, em módulo menores ou iguais a 1.

A **função de autocorrelação parcial** (FACP) de desfasagem (*lag*) *k* é dada por $\{\phi_{kk},\ k \in \mathbb{N}_0\}$ e mede a correlação, linear, entre X_t e X_{t+k}, conhecidos os valores intermédios X_{t+1}, X_{t+2}, ...,X_{t+k-1}. Ou seja,

$$\phi_{kk} = \frac{Cov(X_t, X_{t+k}|X_{t+1}, \dots, X_{t+k-1})}{\sqrt{Var[X_t|X_{t+1}, \dots, X_{t+k-1}]Var[X_{t+k}|X_{t+1}, \dots, X_{t+k-1}]}}$$

A FACP "exprime a variação em X_{t+k} que acompanha, em média, uma variação unitária em X_t quando X_{t+1}, X_{t+2}, ...,X_{t+k-1} permanecem constantes." [13].
As funções de autocovariância e de autocorrelação são funções pares. Ou seja $\gamma_k = \gamma_{-k}$ e $\rho_k = \rho_{-k}, \forall k \in \mathbb{Z}$ e $\phi_{00} = 1$.

2.4 Estimação de medidas de uma série temporal

Na caracterização de séries temporais é extremamente importante a estimação de medidas como a média, a variância e as funções de autocorrelação e autocorrelação parcial. Considerando uma realização $\{X_t;\ t = 1, 2, \dots, n\}$ da série temporal estacionária $X = \{X_t; t \text{ pertencente a } \mathbb{Z}\}$, os estimadores da média, das FAC e FACP são:

- $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$
- $\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X}),\ k = 0, 1, \dots$
- $\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2},\ k = 0, 1, \dots$

Usando o método de Durbin-Levinson obtem-se o seguinte estimador para a FACP [15]

- $\widehat{\emptyset}_{00} = 1$
- $\widehat{\emptyset}_{11} = \hat{\rho}_1$
- $\widehat{\emptyset}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{t=1}^{k}\widehat{\emptyset}_{kt}\,\hat{\rho}_{k+1-t}}{1 - \sum_{t=1}^{k}\widehat{\emptyset}_{kt}\,\hat{\rho}_t},\ k > 1$

2.5 Séries temporais não estacionárias

Muitas séries temporais não são estacionárias porque têm tendência, sazonalidade ou ambas as características. Pode ainda ser estacionária em média, mas não estacionária em variância.
A estacionariedade é importante para a escolha dos modelos e para a estimação desses parâmetros. Assim, sempre que uma série temporal não é estacionária é preciso recorrer a transformações que a transformem em estacionária.

Quando uma série não é estacionária porque a média e/ou a variância não são constantes ao longo do tempo, em primeiro lugar deve estabilizar-se a variância e só depois a média. Caso exista a tendência e sazonalidade, então deve igualmente estalizar-se primeiro a variância, depois eliminar a sazonalidade de seguida a tendência e só depois estabilizar a média.
Esta ordem tem a ver com o facto de que quando existem tendências estas podem não ser facilmente identificadas porque podem ser mascaradas por outros factores. A transformação para eliminar a variância ajuda a visualizar melhor as tendências.

Existem três transformações para estacionarizar uma série temporal: a transformação de Box-Cox, que estabiliza a variância, a diferenciação simples, para eliminar a tendência e a diferenciação sazonal para eliminar a sazonalidade.

Para a estabilizar a variância de uma série não estacionária podemos utilizar os métodos de transformação paramétricos, dentro dos mais conhecidos temos a **transformação de Box-Cox** que consiste na seguinte transformação matemática.

$$Z_t = T(X_t) = \begin{cases} \dfrac{X_t^{\lambda} - 1}{\lambda}; \lambda \neq 0 \\ log(X_t); \ \lambda = 0 \end{cases}$$

onde os valores de λ são, em geral, escolhidos no intervalo $[-5;\ 5]$.

As transformações logarítmicas apenas estão definidas para séries temporais de valores positivos. Para modificar esta situação, pode-se encontrar uma constante b tal que, $X_t + b > 0$ e, posteriormente, pode-se aplicar a transformação nessa mudança de escala.
A transformação logarítmica permite ainda converter um efeito sazonal multiplicativo em aditivo, por isso é uma das transformações mais utilizadas no estudo de séries temporais.

A transformação logarítmica pode ainda ser utilizada para atenuar o efeito de possíveis *outliers*. De realçar sempre que se aplica uma transformação aos dados, todas as previsões geradas pelo modelo selecionado estarão nas unidades transformadas. Por este motivo, depois de ajustar o modelo são estimados os parâmetros, devem ser revertidas as transformações, de modo a obter previsões nas unidades originais [14].

Um dos métodos que permite estabilizar a média consiste no uso de diferenciações simples que consiste na aplicação de operador de diferenciação simples ($\mathcal{D}$), definido de seguida.
Ou seja, se uma série for não estacionária devido à existência de tendência pode aplicar-se, se necessário, uma transformação de Box-Cox e depois a diferenciação simples o número de vezes necessárias até estabilizar a série.

Vamos usar a mesma letra para representar quer a série temporal original, quer a série com transformação de Box-Cox, X_t.
A **diferenciação simples a um passo** define-se como $\mathcal{D}X_t = X_t - X_{t-1},\ t \in \mathbb{Z}$

Se a diferenciação simples a um passo não for suficiente para transformar a série em estacionária pode aplicar-se uma nova diferenciação a esta. Ou seja,

$$\mathcal{D}^2 X_t = \mathcal{D}(\mathcal{D}X_t) = \mathcal{D}(X_t - X_{t-1}) = X_t - 2X_{t-1} + X_{t-2}, \qquad t \in \mathbb{Z}$$

O operador de **diferenciação simples de ordem *d*** para qualquer número inteiro $d \geq 1$, consiste na diferenciação de séries em d vezes até que se atinge a estacionaridade da série. Ou seja,

$$\mathcal{D}^d = \mathcal{D}\left(\mathcal{D}^{d-1}X_t\right), \qquad t \in \mathbb{Z}$$

A diferenciação de uma série já estacionaria é indesejável e deve evitar-se. Se a série transformada $\mathcal{D}^{d_0}X_t$ é estacionária então a série $\mathcal{D}^d X_t$ $(d > d_0)$ também é estacionária, mas com maior variância. Em geral, as diferenciações de primeira e segunda ordens são suficientemente para estacionarizar a série [15].

Quando a série temporal possui componente sazonal aplica-se a diferenciação sazonal.
A **diferenciação sazonal, de período S, a um passo** define-se como $\mathcal{D}_S X_t = X_t - X_{t-S}$. A **diferenciação sazonal, de período S, de ordem *D*** $(D \geq 1)$ consiste em aplicar o operador de diferenciação sazonal de modo interativo. Ou seja,

$$\mathcal{D}_S^D X_t = \mathcal{D}_S(\mathcal{D}_S^{D-1}X_t)$$

Por exemplo, a diferenciação sazonal, de período S, a dois passos traduz-se em

$$\begin{aligned}\mathcal{D}_S^2 X_t &= \mathcal{D}_S(\mathcal{D}_S X_t)\\ &= \mathcal{D}_S(X_t - X_{t-S})\\ &= X_t - 2X_{t-S} + X_{t-2S}\end{aligned}$$

Tal como mencionado relativamente à diferenciação simples, a diferenciação sazonal pode ser aplicada D vezes $(D \geq 1)$ a uma série sazonal apenas o número de vezes necessárias até se obter a estacionariedade da série.

2.6 Análise de estacionariedade de uma série temporal

A estacionariedade de uma série temporal pode ser analisada através do cronograma da série (a sua representação ao longo do tempo) e através do comportamento da FAC. Quando a FAC converge muito lentamente para zero isso é indício de que a série não é estacionária.

Mas para além desta análise mais elementar através dos gráficos gerados pelos dados ao longo de tempo existem alguns testes estatísticos para testar a estacionariedade. Dentro dos quais os mais

utilizados são os testes de Kwiatkowski–Phillips–Schmidt–Shin (KPSS), Dickey-Fuller (DF) e Dickey-Fuller Aumentado (ADF). Os testes DF e ADF testam a existência de raízes unitárias enquanto o teste KPSS testa a existência de tendência.

2.6.1 Teste de Dickey-Fuller

Considera-se que a série pode ser escrita da seguinte forma:

$$X_t = \mu + \phi X_{t-1} + \varepsilon_t$$

Onde μ é uma constante e ε_t é a série de ruído branco.

As hipóteses a testar são:

$$H_0: \phi = 1 \ \ vs \ \ H_1: |\phi| < 1$$

Ou seja, na hipótese nula testa-se que a série tem raíz unitária (a série não é estacionária) e na hipótese alternativa que não existe raíz unitária (a série é estacionária).

Subtraindo X_{t-1} em ambos o lado da equação vem

$$X_t - X_{t-1} = \mu + (\phi - 1)X_{t-1} + \varepsilon_t$$

Ou seja

$$\mathcal{D}X_t = \mu + \beta X_{t-1} + \varepsilon_t$$

com $\beta = (\phi - 1)$.

Assim, passa a testar-se

$$H_0: \beta = 0 \ \ vs \ \ H_1: \beta < 0$$

De acordo com o Hyndman [18], no teste de Dickey-Fuller (DF) deve verificar-se se a sequências dos erros (ε_t) são ou não correlecionados uma vez que quando o pressuposto não for verificado a estatística de teste tende a rejeitar a hipótese nula. Neste caso, deve optar-se por outra versão de teste de raízes unitárias. Uma das sugestões foi aumentar o número de *lags* às diferenças da série, dado orgem ao teste de Dickey-Fuller Aumentado (*Augmented Dickey-Fuller*).

2.6.2 Teste de Dickey-Fuller Aumentado

No caso do teste de Dickey-Fuller Aumentado (ADF), considera-se que

$$\mathcal{D}X_t = \mu + \beta X_{t-1} + \alpha_1 \mathcal{D}X_{t-2} + \cdots + \alpha_{p-1}\mathcal{D}X_{t-p+1} + \varepsilon_t$$

$$\mathcal{D}X_t = \mu + \beta X_{t-1} + \sum_{j=1}^{p-1} \alpha_j \mathcal{D}X_{t-j} + \varepsilon_t$$

onde β representa igualmente o coeficiente de presença de raiz unitária, pelo que as hipóteses a testar são iguais às do teste de Dickey-Fuller.

2.6.3 Teste de Kwiatkowski, Phillips, Schmidt e Shin

O teste de Kwiatkowski, Phillips, Schmidt e Shin (KPSS) considera as seguintes hipóteses a testar:

H_0: A série é estacionária (em tendência) vs H_1: A série tem raiz unitária (não é estacionária)

Assume-se que a série temporal pode ser descrita como

$$X_t = \mu + T_t + r_t + \varepsilon_t$$

com

$$r_t = r_{t-1} + \upsilon_t$$

onde T_t representa a componente de tendência, r_t representa a componente do passeio aleatório (um caso particular dos modelos do tipo ARIMA introduzidos no Capítulo 3), υ_t é uma série i.i.d de média nula e ε_t é a série dos erros estacionária de média nula.
O teste KPSS testa se existe raíz unitária em r_t quando a tendência é nula. Para isso usa os testes dos Multiplicadores de Lagrande [16].

3

Modelos de Séries Temporais

3.1 Modelos lineares de séries temporais

De seguida vamos introduzir os modelos lineares para a modelação e previsão de séries temporais estacionárias e não estacionárias que iremos aplicar neste trabalho.
Iremos também apresentar as principais etapas de modelação, baseadas na metodologia de Box-Jenkins, os critérios de selecção e validação, as previsões e seus intervalos.

3.2 Modelos lineares estacionários

3.2.1 Modelo autorregressivo de ordem p - AR (p)

Os modelos autoregressivos fundamentam-se nos pressupostos de que as observações num dado instante t relacionam-se, de forma linear, com as observações nos instantes anteriores. Diz-se que $\{X_t, t \in \mathbb{Z}\}$ admite uma representação autorregressiva de ordem p (AR(p)) se satisfaz a seguinte equação

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \cdots + \phi_p X_{t-p} + \varepsilon_t$$

ou

$$X_t = \mu + \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t$$

onde μ é uma constante, $\phi_j, j = 1, 2, \ldots, p$ são os parâmetros de autoregressão até à ordem p (com $\phi_p \neq 0$) e ε_t é a série de ruído branco (resíduos) com a média nula e independentes de X_{t-s} para todo o $s \geq 1$.

Usando o operador atraso *B*, definido como $BX_t = X_{t-1}$ e $B^k X_t = X_{t-k}$, pode escrever-se o modelo AR(p) na forma

$$X_t = \mu + \phi_1 BX_t + \phi_2 B^2 X_t + \phi_3 B^3 X_t + \cdots + \phi_p B^p X_t + \varepsilon_t$$

ou na forma mais reduzida

$$\phi_p(B) X_t = \mu + \varepsilon_t$$

onde $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \cdots - \phi_p B^p$ é um polinómio autorregressivo de ordem p.

"Diz-se que uma série é invertível (ou causal) se possuir representação autoregressiva" [14]. Pelo que uma série AR(p) é sempre **invertível**, mas nem sempre é estacionária.
É **estacionária** "se e só se o polinómio $\phi_p(B)$ não possuir raízes no círculo unitário" [14].

Admitindo que a série é estacionária prova-se facilmente que

$$\mathrm{E}[X_t] = \frac{\mu}{1 - (\phi_1 + \phi_2 + \cdots + \phi_p)}$$

$$\mathrm{Var}[X_t] = \frac{\sigma_\varepsilon^2}{1 - (\phi_1 \rho_1 + \phi_2 \rho_2 + \cdots + \phi_p \rho_p)}$$

com ρ_k a FAC da série X_t no *lag* k.
E também se prova que as funções de autocovariância, FAC e FACP, verificam as igualdades

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p}, \qquad k = 1, 2, \ldots$$
$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \qquad k = 1, 2, \ldots$$
$$\phi_{kk} = 0, \qquad k > p$$

Nas séries autorregressivas estacionárias a dependência do passado para explicar o presente diminui à medida que o passado se torna mais afastado.

Dois teoremas muito importantes estabelecem o comportamento da FAC e da FACP de um AR(p) estacionário, provando que a FAC tende exponencialmente, e ou de modo sinusoidal, para zero, à medida que o valor do *lag k* aumenta e que a FACP se anula a partir da ordem p+1. [14, pag.44]

Apresentando apenas o caso particular de ordem 1. Um modelo **AR(1)** de média não nula

$$X_t = \mu + \phi X_{t-1} + \varepsilon_t$$

admite solução estacionária se $|\phi| < 1$ e prova-se que as suas média, variância, FAC e FACP são, respetivamente, da forma

$$\mathrm{E}[X_t] = \frac{\mu}{1-\phi}$$

$$\mathrm{Var}[X_t] = \frac{\sigma_\varepsilon^2}{1-\phi^2}$$

$$\rho_k = \phi^k, \qquad k = 0, 1, \ldots$$

$$\phi_{11} = \phi \qquad \text{e} \qquad \phi_{kk} = 0, \ k > 1$$

Por definição $\rho_0 = 1$ e $\phi_{00} = 1$.

3.2.2 Modelo de médias móveis- MA(q)

Diz-se que $\{X_t, t \in \mathbb{Z}\}$ admite uma representação de médias móveis de ordem de ordem q (MA(q)) se satisfaz a seguinte equação

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_3 \varepsilon_{t-3} - \cdots - \theta_q \varepsilon_{t-q}$$

Ou, escrevendo à custa do operador atraso

$$X_t = \mu + \theta_q(B)\varepsilon_t$$

onde ε_t é um ruído branco de média nula e $\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \cdots + \theta_q B^q$ (com $\theta_q \neq 0$) é o polinómio de médias móveis de ordem q.

Como esta série é definida só à custa do ruído branco é sempre **estacionária**. E é **invertível** "se e só se o polinómio onde $\theta_q(B)$ não possuir raízes no círculo unitário" [14].

Num modelo MA(q) admite-se que a série X_t, no instante *t*, se relaciona, de forma linear, com os erros presentes e passados.

Prova-se após alguns cálculos que

$$\mathrm{E}[X_t] = \mu$$

$$\mathrm{Var}[X_t] = \sigma_\varepsilon^2 \left(1 + \sum_{i=1}^{q} \theta_i^2\right)$$

$$\rho_k = 0, \qquad k > q$$

Os dois teoremas relativos aos comportamentos da FAC e da FACP de um modelo MA(q) dizem que a FAC se anula a partir da ordem q+1 e que a FACP tende exponencialmente, ou de forma sinusoidal para zero, à medida que o valor de k (*lag*) aumenta. [14, pag.40]

Particularizando, um modelo do tipo MA(1) de média não nula

$$X_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1}$$

é sempre estacionário e invertível se $|\theta| < 1$ e prova-se que as suas média, variância, FAC e FACP são, respetivamente, da forma [14].

$$\mathrm{E}[X_t] = \mu$$

$$\mathrm{Var}[X_t] = \sigma_\varepsilon^2(1 + \theta^2)$$

$$\rho_1 = -\frac{\theta}{1+\theta^2} \quad \text{e} \quad \rho_k = 0,\ k > 1$$

$$\phi_{kk} = -\frac{\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}}, \qquad k \geq 1$$

Por definição $\rho_0 = 1$ e $\phi_{00} = 1$.

3.2.3 Modelo autorregressivo e médias móveis - ARMA(*p, q*)

Os modelos autorregressivos e médias móveis de ordens p e q são a combinação dos modelos AR e MA de ordens p e q, respetivamente.
Admitem, portanto, a seguinte representação

$$X_t = \mu + \phi_1 \mathrm{X}_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

com $\phi_p \neq 0$ e $\theta_q \neq 0$. De igual modo, μ representa a constante e ε_t a série de ruído branco de média nula.

Também podem ser escritos à custa dos polinómios autorregressivo e de médias móveis $\phi_p(B)$ e $\theta_q(B)$

$$\phi_p(B) X_t = \mu + \theta_q(B) \varepsilon_t$$

Estes modelos são estacionários e invertíveis cumpridas as condições anteriormente enunciadas para os modelos AR(p) e MA(q).

Para estes modelos tem-se, para séries estacionárias, que tanto a FAC como a FACP tendem exponencialmente, ou de forma sinusoidal para zero, à medida que o valor de k (*lag*) aumenta.

As expressões da média, variância, FAC e FACP para estes modelos obtêm-se facilmente das já obtidas para os modelos AR(p) e MA(q).
Apenas particularizando para o caso **ARMA(1,1)**:

$$X_t = \mu + \phi \mathrm{X}_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$$

prova-se [14] que,

$$\mathrm{E}[X_t] = \frac{\mu}{1-\phi}$$

$$\mathrm{Var}[X_t] = \frac{1 - 2\phi\theta + \theta^2}{1 - \phi^2}\, \sigma_\varepsilon^2$$

$$\rho_k = \frac{(\phi - \theta)(1 - \phi\theta)}{1 - 2\phi\theta + \theta^2}\phi^{k-1}, \qquad k \in \mathbb{N}$$

"e a função de autocorrelação parcial decai para zero através da combinação de séries geométricas e/ou sinusoidais amortecidas a partir da ordem 1." [14].

Os modelos do tipo ARMA(p,q) são importantes na modelação porque evitam o uso de modelos AR(p) ou MA(q) com um número excessivo de parâmetros. Ou seja, através dos modelos ARMA(p,q) obtêm-se modelações mais parcimoniosas.

A tabela abaixo apresenta um resumo do comportamento das FAC e FACP dos modelos estudados.

Processo	FAC	FACP
AR(p)	de modo sinusoidal, para zero à medida que o valor de k aumenta	Anula-se a partir da ordem p+1.
MA(q)	Anula-se a partir da ordem q+1	tende exponencialmente, ou de forma sinusoidal, para zero, à medida que o valor de k aumenta
ARMA(p,q)	tende exponencialmente, ou de forma sinusoidal, para zero à medida que o valor de k aumenta	tende exponencialmente, ou de forma sinusoidal, para zero à medida que o valor de k aumenta

Tabela 3.1: Propriedades das FAC e FACP.

Fonte [12]

3.3 Modelos lineares não estacionários

Na prática, a grande maioria das séries temporais são não estacionárias. Se este cenário acontece é necessário transformar a série (removendo o que a faz não estacionária: tendência e/ou sazonalidade). Como foi mencionado atrás se a série for não estacionária pode ser aplicada uma ou várias transformações para a tornar estacionária.

Seja X_t uma série não estacionária em tendência e seja $Y_t = T(X_t)$, onde $T(X_t)$ é a transformação de Box-Cox. Suponhamos que se aplicarmos uma diferenciação simples a um passo a Y_t, ou seja, $\mathcal{D}Y_t = (1 - B)Y_t$, tornamos a série estacionária.
Neste caso, podemos dizer que a série $\mathcal{D}Y_t$ pode admitir uma representação do tipo ARMA(p,q).
O modelo do tipo ARIMA é, portanto, um tipo de modelo para séries temporais lineares não estacionárias (em tendência) que ao serem transformadas através de diferenciações simples (após uma transformação de Box-Cox, se necessário) se transformam em estacionárias e que admitem uma estrutura do tipo ARMA.

3.3.1 Modelo integrado autorregressivo e médias móveis - ARIMA(*p,d,q*)

Seja X_t uma série não estacionária e vamos, por simplicidade, designar igualmente por X_t a série original ou com a transformação de Box-Cox.

O modelo integrado autorregressivo e médias móveis de ordens p e q (ARIMA(p,d,q)) pode ser expresso da seguinte forma

$$\phi_p(\mathrm{B})(1-B)^d X_t = \mu + \theta_p(B)\varepsilon_t$$

onde d representa o número de diferenciações simples necessárias para transformar a série em estacionária ($\mathcal{D}^d X_t, d \geq 1$), μ representa a constante e ε_t a série de ruído branco de média nula.

No modelo ARIMA a FAC tem valores positivos e de uma forma decrescem lentamente para o zero.

Um **passeio aleatório** é um caso particular de um modelo ARIMA(*p,d,q*), quando *p*=*q*=0 e *d*=1. Ou seja, um passeio aleatório é um ARIMA(0,1,0).

Os modelos do tipo ARIMA(*p,d,q*) não possuem componente sazonal, logo não são adequados para séries temporais que possuam sazonalidade. Para isso existem os modelos mais completos designados por modelos SARIMA.

3.3.2 Modelo integrado autorregressivo e médias móveis sazonal – SARIMA *(p,d,q)(P,D,Q)s*

O modelo integrado autorregressivo e médias móveis sazonal de ordens, respetivamente, autoregressivas e médias móveis (p, P) e (q, Q) e com período de sazonalidade S, SARIMA *(p,d,q)(P,D,Q)s*, admite a representação

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^S)^D X_t = \mu + \theta_q(B)\Theta_Q(B^S)\varepsilon_t$$

onde $\phi_p(B)$ e $\theta_q(B)$ são os polinómios autorregressivo e de médias móveis da parte não sazonal e $\Phi_P(B^s)$ e $\Theta_Q(B^S)$ os polinómios autorregressivo e de médias móveis da parte sazonal de período S.

As ordens p e q representam as ordens de dependência dentro do período de sazonalidade S e as ordens P e Q ordens de dependência entre os períodos. Os valores de d e D representam, respetivamente, o número de diferenciações simples e sazonais a aplicar para transformar a série em estacionária.

Estamos igualmente a escrever como X_t a série original e a série com transformação de Box-Cox, caso esta seja necessária para a transformação da série em estacionária. Designamos por μ a constante e por ε_t a série de ruído branco de média nula.

Todos os modelos anteriores são casos particulares do modelo SARIMA.

Normalmente, a necessidade de uma diferenciação sazonal pode ser indicada através da FAC da série temporal, nomeadamente, quando esta decai lentamente nos *lags* múltiplos do período e é insignificante nos restantes [17].

3.4 Etapas na modelação de uma série temporal - metodologia de Box-Jenkins

A metodologia de Box-Jenkins, introduzida em 1970, é uma metodologia de ajustamento de uma série temporal por etapas, sendo as principais a identificação do modelo, a estimação dos seus parâmetros e validação do modelo.

A **identificação do modelo** linear a ajustar (do tipo SARIMA) a uma série temporal é feita através do comportamento das suas FAC e FACP estimadas.

Quando a série temporal não exibe tendência e/ou sazonalidade, ou seja, é estacionária, as FAC e FACP dão indicação de uma estrutura do tipo AR, MA ou ARMA e uma indicação das respectivas ordens (Tabela 1). Quando a série não é estacionária as FAC e FACP já não têm essa leitura.

Assim, a primeira etapa (**etapa 1.1**), dentro da identificação, consiste em avaliar a estacionariedade da série. Para isso, representa-se a série (cronograma) e obtêm-se as FAC e FACP estimadas. Visualmente pode deduzir-se se a série é ou não estacionária, mas também podem (e devem) ser aplicados testes de hipóteses para avaliar a estacionarieade.

Caso a série temporal não seja estacionária deverá aplicar-se as transformações necessárias para a transformar em estacionária (**etapa 1.2**): transformação de Box-Cox, diferenciações sazonais e/ou simples (por esta ordem).

Com a série estacionária e as suas FAC e FACP estimadas escolher um (ou mais) possíveis modelos a ajustar e as respectivas ordens (p, q, P, Q) (**etapa 1.3**). Para isso usam-se as propriedades resumidas na Tabela 1.

A segunda etapa consiste na **estimação dos parâmetros do(s) modelo(s)**. Existem vários métodos de estimação dos parâmetros do modelo, como o método dos momentos (também conhecido por Estimador de Yule-Walker), o método dos mínimos quadrados (condicionados ou não condicionados) e método da máxima verosimilhança (exata ou condicionada). Para mais detalhe ver [14].

A terceira etapa consiste na **validação do modelo**. Dentro desta começa-se pela avaliação da significância dos parâmetros (**etapa 3.1**). A significância dos parâmetros consiste em fazer testes de hipóteses aos parâmetros. Ou seja, considerando um modelo ARMA(p,q) de média nula, testa-se a significância da constante

$$H_0: \mu = 0 \ \ vs \ \ H_1: \mu \neq 0$$

e a significância dos restantes parâmetros $\beta_i = \left(\phi_1, \dots, \phi_{p,} \theta_1, \dots, \theta_q\right)$

$$H_0: \beta_i = 0 \ \ vs \ \ H_1: \beta_i \neq 0, \qquad i \geq 1$$

Para mais detalhe (estatística de teste, etc) ver [14].

Após a avaliação da significância dos parâmetros, da retirada dos parâmetros não significativos e da re-estimação passa-se para a quarta etapa, o **estudo dos resíduos**.
Os resíduos do modelo estimado são a diferença entre os valores observados e os valores estimados. A série dos resíduos deverá assemelhar-se a um ruído branco, de média nula e deverão também ser aleatórios. Também se pode testar a normalidade dos resíduos apesar desta não ser uma condição necessária, apenas contribui para uma maior rapidez no cálculo dos intervalos de previsão [18].

Para testar a nulidade da média dos resíduos aplica-se um **test t-student**. Através da análise das FAC e FACP estimadas dos resíduos pode testar-se as hipóteses de estas serem nulas

$$H_0: \rho_k = 0 \ \ vs \ \ H_1: \rho_k \neq 0, \qquad k > 1$$

e

$$H_0: \phi_{kk} = 0 \ \ vs \ \ H_1: \phi_{kk} \neq 0, \qquad k > 1$$

Para testar estas hipóteses usam-se, respetivamente, o **teste de Bartlett** e o **teste de Jenkins e Daniels** [15].
Ou então, mais adequado, testar as hipóteses conjuntas

$$H_0: \rho_1 = \ \rho_2 = \cdots = \ \rho_k = 0$$

usando, por exemplo, o **teste de Ljung-Box** [14].

Para avaliar a aleatoriedade dos resíduos usualmente aplicam-se os o **teste do sinal das diferenças** (*"Difference-Sign Test"*), o **teste dos instantes de inflexão** ("*Difference-Sign Test*") e o **teste das ordens** ("*Rank Test*"). Para mais detalhe ver [14].

Podemos ainda acrescentar uma quarta etapa que é a etapa da **seleção de modelos**.
Quando se tem mais do que um "bom" modelo (pressupostos todos validados) é preciso escolher de entre esses qual o melhor. O melhor modelo depende do objetivo do estudo. Pode ser, por exemplo, o modelo que produz melhor descrição e/ou melhores previsões.

Existem 4 critérios usuais de seleção de modelos:

- Critério de informação de Akaike (AIC);
- Critério Bayesiano de informação de Akaike (BIC);
- Critério de informação de Akaike Corrigido (AICc);

- Critério Bayesiano de Schwartz (SBC),

sendo que o "melhor" modelo é aquele que apresentar o menor valor destes critérios. Para mais detalhe ver, por exemplo, [20].

3.5 Previsão e critérios de avaliação

Dada uma série temporal $\{X_t, t = 1, 2, \dots, n\}$ e ajustado e validado o "melhor" modelo, o "melhor" **preditor a *k* passos** define-se como

$$\hat{X}_{n+k} = E[X_{n+k}|X_1, \dots, X_n]$$

De modo a avaliar a precisão das previsões são calculados os erros de previsão (diferença entre o valor observado e a correspondente previsão dada pelo modelo). O objetivo é minimizar esses erros, dados por $e_{t+k} = X_{t+k} - \hat{X}_{n+k}$.

Por fim, obtêm-se os intervalos de confiança das previsões.
Seja $\{X_t, t = 1, 2, \dots, n\}$ uma série temporal modelada por um modelo do tipo SARIMA onde os erros ϵ_t são ruído branco de média nula e com distribuição normal.
Neste caso, prova-se ([15]) que o **intervalo de previsão a *k* passos**, com uma confiança de $(1-\alpha)100\%$, para $\hat{X}_{n+k}$, conhecidos os valores da série até ao instante *n*, é dado por

$$]\hat{X}_{n+k} - Z_{1-\frac{\alpha}{2}} * \sigma_{\hat{X}_{n+k}};\ \hat{X}_{n+k} + Z_{1-\frac{\alpha}{2}} * \sigma_{\hat{X}_{n+k}}[$$

onde $\sigma_{\hat{X}_{n+k}}$ é o erro padrão de $\hat{X}_{n+k}$ e $Z_{1-\frac{\alpha}{2}}$ é o percentil $\left(1-\frac{\alpha}{2}\right)100\%$ da distribuição normal padrão.
No caso de os erros serem normais e a variância dos erros desconhecida (o caso usual) substitui-se o percentil da distribuição normal padrão pelo percentil da distribuição t-Student. Como em geral as séries têm uma dimensão elevada é indiferente, em termos práticos usar uma ou outra distribuição.

No caso de os erros não serem normais pode prosseguir-se de igual modo com a previsão, desde que estes não se afastem demasiado da normalidade.

Para a validação das previsões usa-se, em geral, um conjunto de treino (parte da série temporal usada para a modelação) e um conjunto de teste (parte da série temporal usada para a comparação dos valores observados com as previsões). As melhores previsões são aquelas que resultam nas menores medidas de precisão.
Usualmente usam-se as seguintes **medidas de precisão**: **MAE**; **MSE**; **RMSE** e **MAPE**.
Dada uma série temporal $\{X_t, t = 1, 2, \dots, n\}$ define-se:

- **Erro absoluto médio – MAE**

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|X_t - \hat{X}_t\right|$$

Ou seja, é a diferença média (em valor absoluto) entre o valor observado e o valor estimado (ou previsto). Ou seja, a média do valor absoluto dos erros.
É uma medida que nos dá a magnitude do erro, é fácil de calcular e de interpretar e como está em valor absoluto, os valores sobreprevistos (ou sobreestimados) não anulam os valores subprevistos (subestimados).

A desvantagem é que não fornece nenhuma informação sobre a direção ou padrão dos erros, trata todos os erros igualmente, o que significa que uma previsão com erros positivos consistentes e outra com erros negativos consistentes teriam o mesmo erro absoluto médio [31].

- **Erro quadrático médio – MSE**

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{X}_t)^2$$

É uma medida igualmente fácil de calcular e penaliza os erros muito grandes.
As desvantagens são que esta medida não é muito fácil de interpretar porque as unidades vêm ao quadrado, é muito sensível aos valores discrepantes (ou *outliers*) e não distingue entre sobreprevisão (ou sobreestimação) e subprevisão (subestimação) [30].

- **Raiz quadrada do erro quadrático médio – RMSE**

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{X}_t)^2}$$

Como aplica a raiz quadrada ao MSE fica fácil de interpretar e já penaliza os grandes erros.
As desvantagens é que continua a ser muito sensível aos valores discrepantes e continua a não fornece informações sobre a direção ou o padrão dos erros de previsão e não distingue sobreprevisão (ou sobreestimação) e subprevisão (subestimação) [30].

- **Erro absoluto médio percentual - MAPE**

Esta medida é uma das mais usadas por ser de fácil compreensão. Expressa a percentagem média dos erros (em valor absoluto) cometidos nas previsões (ou ajustamento) da série temporal.

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{X_t - \hat{X}_t}{X_t}\right| * 100\%$$

A vantagem desta medida é que os seus valores são expressos em percentagem, o que torna mais compreensível a sua interpretação e também permite fazer a comparação entre diferentes modelos porque não depende da escala dos dados.
A desvantagem está quando a série possui valores próximos ou iguais a zero, neste caso o valor do erro absoluto médio percentual tende para infinito [19]. Continua a ser uma medida muito sensível a valores discrepantes.

Para todas estas medidas, quanto menor for o seu valor melhor será a previsão, ou ajustamento, [19].
Estas medidas são também usadas para validar o ajustamento do modelo e poder escolher de entre vários modelos possíveis qual o que melhor se ajusta aos dados.

De uma forma resumida a fase de metodologia de Box-Jenkins pode ser apresentada através deste Fluxograma para descrever os processos de modelação de séries temporais.

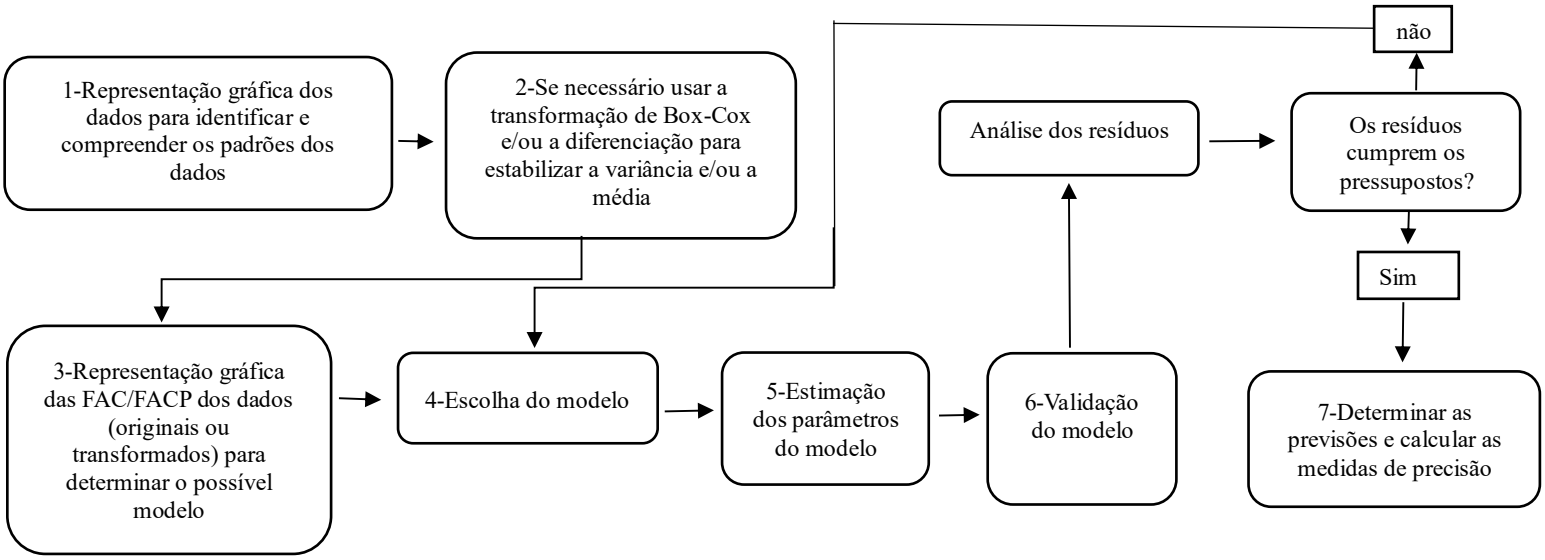

Tabela 3.2. Fluxograma das etapas da metodologia de Box-Jenkins na análise de séries temporais.

4

Redes Neuronais Artificiais e Modelos de Memória de Longo Prazo

4.1 Introdução

Neste capítulo iremos introduzir os principais fundamentos e conceitos das Redes Neuronais Artificiais (RNAs), unidirecionais e recorrentes, de *Machine Learning*, de *Deep Learning*, e de um caso particular de uma RNA recorrente que funciona em *Deep Learning*.

4.2 Redes neuronais artificiais

As Redes Neuronais Artificiais (RNAs) foram pensadas à semelhança do funcionamento do cérebro humano, com todos os seus neurónios.
Os neurónios são células responsáveis por transmitir os impulsos nervosos que recebem através dos sentidos. Mais concretamente, "os neurónios são células nervosas, estrutura básica do sistema nervoso, que processam e transmitem informação através de sinais eletro-químicos." [20].

Como pode ser observado na Figura 4.1, o neurónio é constituído pelas seguintes células

- ✓ Dendrites: têm como a função receber informações, ou impulsos nervosos, oriundos de outros neurónios e conduzi-las até o corpo celular;
- ✓ Corpo Celular: recebe a informação enviada pelas dendrites e decide a ação tomar;
- ✓ Axónio: transmite os impulsos nervosos provenientes do corpo celular;
- ✓ Terminação do axónio: "contêm sinapses, estruturas especializadas onde são libertadas susbtâncias químicas, neurotransmissores, que estabelecem a comunição com as dendrites ou corpo celular de outros neurónios" [20].

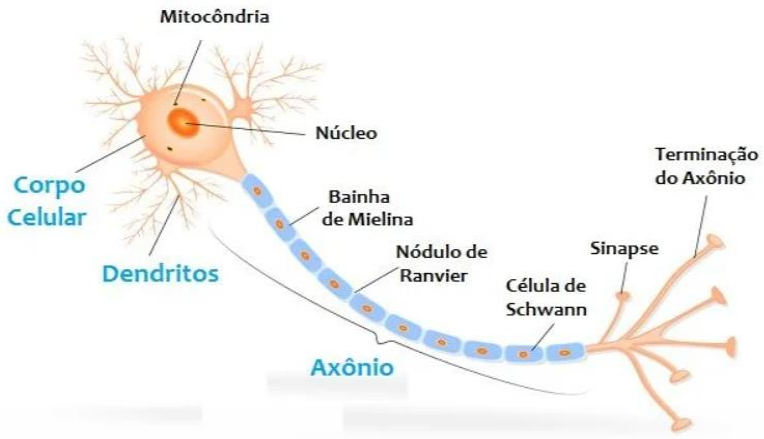

Figura 4.1. Estrutura básica de um neurónio biológico.

Fonte: https://www.todamateria.com.br/neuronios

As RNAs são assim sistemas constituídos à semelhança de um neurónio biológico por nodos (ou "neurónios" artificiais) que são unidades de processamento de informação que, através de uma operação matemática (linear ou não linear) produzem uma ação (uma saída) a partir de um conjunto de estímulos (conjunto de informação de entrada). A saída é uma função da soma ponderada das entradas mais um viés [29; 24].

Mais concretamente, os neurónios artificiais são constituídos por 4 componentes: as entradas (também chamados na linguagem da computação *inputs*) – estas entradas podem ser os próprios dados originais, mas também podem ser as saídas produzidas por outros neurónios -, os pesos, uma função de ativação e um viés.

As RNAs funcionam, assim, do seguinte modo:

- ✓ cada neurónio recebe **dados de entrada**, ou *inputs*, (dados originais ou informação resultante da saída de outros neurónios).
- ✓ Como nem toda a informação de entrada tem a mesma relevância para a saída, a cada entrada está associado um **peso** que pondera a importância dessa entrada para a saída.
  Na etapa de treino da rede estes pesos vão sendo ajustados em função do objectivo do estudo (reconhecimento de padrões, previsões, etc.). Os valores iniciais dos pesos da rede são números aleatórios uniformemente distribuídos num certo intervalo definido.
- ✓ Cada neurónio tem um **viés** (*bias***)** de modo a evitar o sobreajustamento (*overfitting*) e o subajustamento (*underfitting*).

- ✓ Por fim, as entradas ponderadas e o viés são somados resultando numa soma ponderada. Esta soma ponderada é transformada à custa de uma função matemática (**função de ativação**) que vai então permitir que a rede "aprenda" com os dados.

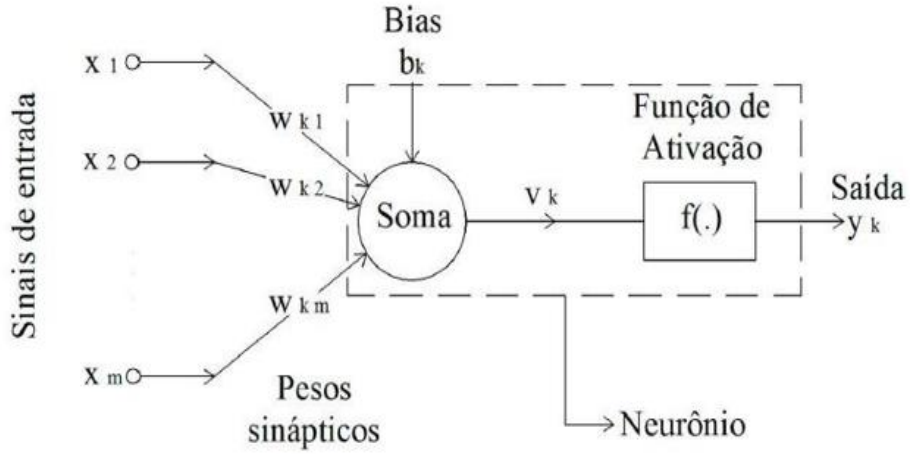

Figura 4.2. Rede neural artificial.

Fonte: Neuronio-artificial-proposto-por-McCulloch-e-Pitts-em-1943-Araujo-2013_fig2_26874316

As RNAs possuem 3 **camadas**:

- ✓ a camada de entrada (ou de *input*), que é a camada que recebe os dados,
- ✓ a camada oculta (*Hidden*), que é a camada intermediária onde vão ocorrer os processamentos (os pesos, o viés e a função de ativação)
- ✓ a camada de saída (ou de *ouput*), que é a camada que recebe o resultado do processamento da rede.

A camada oculta pode ser formada por uma **única camad**a ou por **múltiplas camadas**. No caso de a camada oculta ser constituída por várias camadas temos a chamada **aprendizagem profunda**, também designado por **RNAs profundas**, ou simplesmente *Deep Learning*.

Quando uma rede possui apenas uma única camada oculta isso implica que os padrões de saída são similares aos padrões de entrada. Neste caso, a rede não "aprende" e, portanto, não consegue produzir saídas diferentes.

As RNAs podem ainda dividir-se, em função da ligação entre os nós, em redes unidirecionais ou em redes multidirecionais (Figura 4.3 e Figura 4.4)

Nas RNAs **unidirecionais** (*Feed-Forward*) o fluxo segue apenas um sentido, que vai da entrada para a saída. Ou seja, as entradas e as saídas são independentes umas das outras. A rede não se re-alimenta. Isto significa que a saída de qualquer camada não afeta a mesma camada.
Nas RNAs **multidirecionais** (usualmente designadas por RNAs **Recorrentes**) o fluxo não segue um só sentido. O que significa que saída de qualquer camada pode afetar essa mesma camada. Ou seja, são introduzidas repetições (*loops*, grafo de conexão) na rede. A rede é, portanto, re-alimentada, o que lhe dá uma "memória".

Nas RNAs Recorrentes (ou RNN, acrónimo do inglês, **Recurrent Neuronal Network**) as saídas dependem dos valores da sequência anterior. Daí a sua importância nas séries temporais.

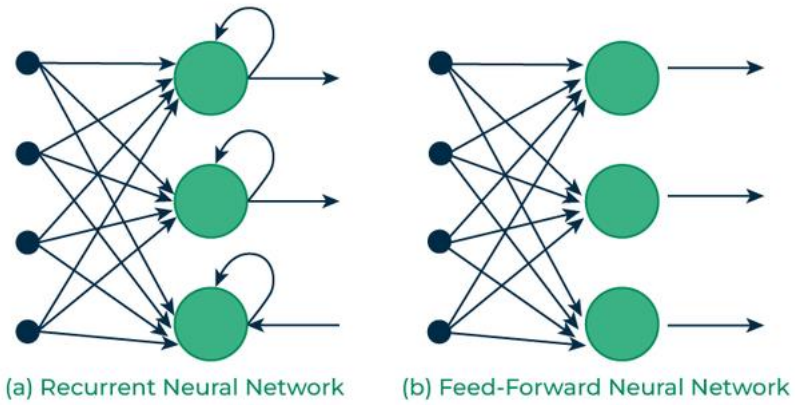

Figura 4.3. RNA recorrente e unidirecional.
Fonte: https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/

Outra coisa que distingue estas duas redes é que as RNAs unidirecionais possuem pesos diferentes em cada nó enquanto as RNNs recorrentes possuem o mesmo peso dentro de cada camada. Estes pesos ainda são ajustados através dos processos de propagação retroativa e descida de gradiente para facilitar a aprendizagem por reforço. (https://www.ibm.com/br-pt/topics/recurrent-neural-networks).

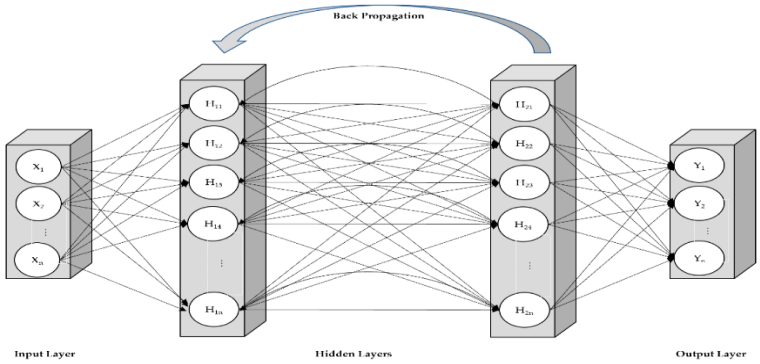

Figura 4.4. RNA recorrente com duas camadas ocultas.

Fonte: https://www.mdpi.com/2078-2489/11/5/243

4.3 Treino e aprendizagem de RNAs

Para que os algoritmos produzam novos dados ou previsões (saídas) é preciso que a máquina aprenda (ou apreenda o padrão dos dados de entrada). Ou seja, no *machine learning* a máquina treina um algoritmo através de um conjunto inicial de dados. Para isso, também aqui os dados são divididos em dados de treino e dados de teste.

Nas fases de treino de redes, diversos exemplos são apresentados para a rede poder ajustar os seus pesos até que a rede aprenda, para se poder representar as informações ou dar as respostas; os pesos são determinados para gerar as soluções para novos valores fornecidos na entrada na rede para o treino [21].

No treino das RNAs podem ser aplicados diferentes algoritmos, que diferem pelo modo como é gerado o ajuste dos pesos da rede. O algoritmo de treino (e, por conseguinte, a aprendizagem da rede) divide-se em dois tipos: supervisionado e não supervisionado.
No caso da **aprendizagem não supervisionada** é dada à rede valores de entrada brutos e não etiquetados que são organizados em diferentes classes, de acordo com as suas semelhanças, através do ajustamento dos seus pesos de modo a produzir saídas consistentes. Este tipo de aprendizagem adequa-se, por exemplo, no *clustering* temporal.

No caso da **aprendizagem supervisionada** é dada à rede (na fase de treino) valores de entrada e de saída etiquetados. Esta etiquetagem (classificação prévia) é dada pelo utilizador. Daí ser conhecida esta aprendizagem por supervisionada. As saídas resultantes são comparadas com o conjunto de teste. É calculado um erro, a rede é re-alimentada com esse erro (ajustando os pesos) de modo a minimizar o erro final.

A aprendizagem supervisionada pode ser classificada em duas categorias: aprendizagem dinâmica e aprendizagem estática [21].

- ✓ Aprendizagem dinâmica: ocorrem alterações na estrutura de rede como a redução do número de camadas, nós em camadas e número de conexões.
- ✓ Aprendizagem estática: somente os valores dos pesos sofrem alterações. A estrutura da rede é mantida constante ao longo da aprendizagem da rede.

Vamos apenas abordar a **aprendizagem supervisionada estática**. Neste tipo de aprendizagem, ou de treino da rede, existem duas etapas: a etapa de *forward* e a etapa *backward*.
Ou seja, na primeira etapa a aprendizagem supervisionada estática relaciona os valores de saída calculados com o valor de entrada esperado (etiquetado) e avalia os erros decorrentes da diferença entre ambos. Na segunda etapa esses erros re-alimentam a rede, alterando os pesos de forma a minimizar o erro e encontrar a melhor relação entre as informações de entrada e de saída [21].
Ou seja, os erros $\varepsilon_j(t)$ são descritos como

$$\varepsilon_j(t) = d_j(t) - y_j(t)$$

onde $y_j(t)$ é o valor calculado pela RNA no instante t e $d_j(t)$ o valor da saída esperado, para os j elementos de entrada.
Em cada fase de treino são feitas algumas alterações nos pesos, proporcionando uma minimização ajustável dos erros que vão convergindo em direção do valor esperado ($d_j(t)$) [21]. A alteração nos pesos pode ser descrita pela seguinte função:

$$w_{j,i}(t+1) = w_{j,i}(t) + \eta\varepsilon_j(t)x_i(t)$$

onde η a taxa de aprendizagem que determina a velocidade com que os pesos serão ajustados em direção ao menor erro e $x_i(t)$ é a entrada para o nó i no tempo t.

Para avaliar o desempenho da rede usam-se os critérios usuais como o Erro Quadrático Médio (EQM).
Durante a fase de aprendizagem (ou de treino da rede), o EQM normalmente começa com valores altos e vai decrescendo conforme aumenta o número de etapas. No final, o EQM mantém-se praticamente constante e a rede neural converge [22].
O conjunto de teste avalia se a rede ficou ou não demasiado treinada (**sobreajustamento**). Neste caso, a rede perde a capacidade de ter bom desempenho com novos dados. Os erros de treino são

muito pequenos, mas os erros usando novos dados são muito grandes. No sobreajustamento o algoritmo capta não só a tendência dos dados, mas também os seus ruídos [23].

Um **subajustamento** também não é bom. Isto significa que a rede não foi bem treinada. Neste caso, o desempenho não vai ser bom nem com os dados de treino, nem com novos dados.

4.4 Funções de ativação

A função de ativação representa o efeito da entrada de informação na rede, pelo que pode ser considerada como o estado de activação dos nós. Ou, por outras palavras, são consideradas como sendos os impulsos nervosos das RNAs.

Em termos matemáticos descreve-se assim o neurónio artificial. Seja $(x_1, x_2, \dots, x_T)$ uma sequência de entrada, $(h_1, h_2, \dots, h_T)$ a sequência dos neurónios e $(y_1, y_2, \dots, y_T)$ a sequência de saída.
Uma RNA define-se como [25].

$$h_t = \varphi(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$

$$y_t = W_{yh}h_t + b_y$$

onde $\varphi(\cdot)$ é a função (não linear) de ativação, W é a matriz dos pesos e b é o viés.

Estas funções dividem-se em funções limiares (*binary step*), lineares e não lineares.

As **funções limiares (*binary step*)** são sugeridas para classificações binárias e redes simples sem camadas ocultas. É estabelecido um limite em que a função é ativada.
Esta função toma o valor 0 quando o valor de saída do nó é negativo e 1 quando é positivo ou nulo. Ou seja,

$$\varphi(\mu) = \begin{cases} 1 \; se \; \mu \geq 0 \\ 0 \; se \; \mu < 0 \end{cases}$$

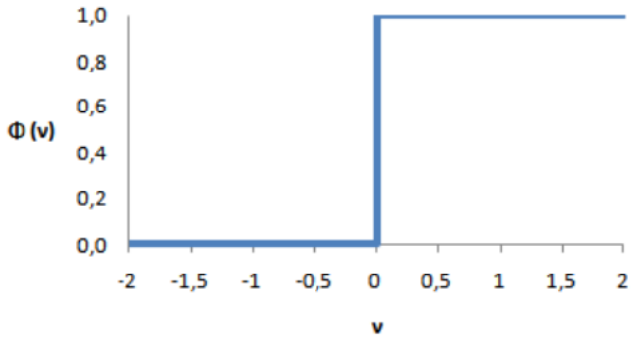

Figura 4.5. Função limiar.

Fonte. https://www.maxwell.vrac.puc-rio.br/16580/16580_4.PDF

Também se define a **função limiar (*binary step*) por partes**. Esta função pode ser vista como uma aproximação de um amplificador não linear. É representada pela equação

$$\varphi(\mu) = \begin{cases} 1 & se \quad \mu \geq +1/2 \\ \dfrac{\mu}{2} & se \quad -1/2 < \mu < 1/2 \\ 0 & se \quad \mu \leq -1/2 \end{cases}$$

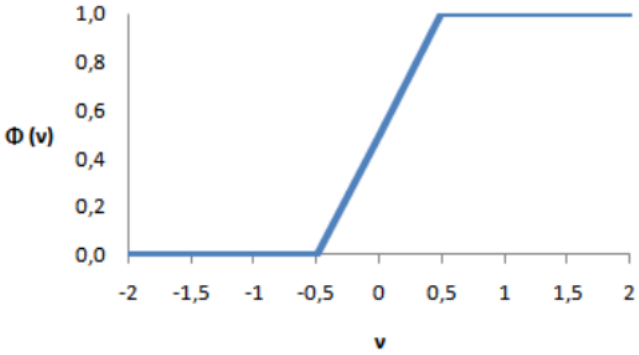

Figura 4.6. Função limiar por partes.

Fonte. https://www.maxwell.vrac.puc-rio.br/16580/16580_4.PDF

As **funções lineares** têm pouco interesse nesta metodologia porque da sua aplicação resultaria que os valores de saída seriam diretamente proporcionais aos valores de entrada. São aplicadas nos casos de regressão linear. A função linear toma valores nos números reais e

$$\varphi(\mu) = a\mu + b$$

e resulta na reta linear usual. As funções lineares correspondem a uma não-ativação.

São as **funções não lineares** que permitem que o algoritmo "aprenda" a descobrir padrões e a estabelecer relações complexas entre os dados. Vamos apresentar alguns tipos.

4.3.1 Função Sigmoide

Também conhecida como função logística. É a função mais abrangente de todas as funções para o funcionamento da rede. Pode ser caracterizada como uma função crescente com balanceamento adequado entre as funções linear e não linear, e que assume valores no intervalo [0,1].
A função sigmóide é representada por:

$$\varphi(\mu) = \frac{1}{1 + e^{-\alpha\mu}}$$

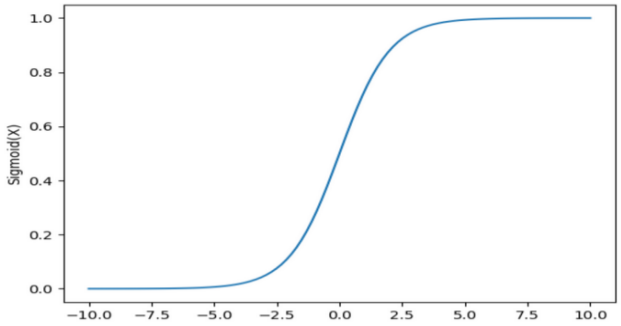

Figura 4.7. Função sigmoide.

Fonte: https://acervolima.com/implementar-a-funcao-sigmoide-usando-numpy/

sendo o α o parâmetro da inclinação na função sigmóide.

4.3.2 Função Tangente Hiperbólica

Contrariamente à função sigmoide, esta função toma valores entre -1 e 1.

$$F(\mu) = tanh(\mu) = \frac{e^{\mu} - e^{-\mu}}{e^{\mu} + e^{-\mu}}$$

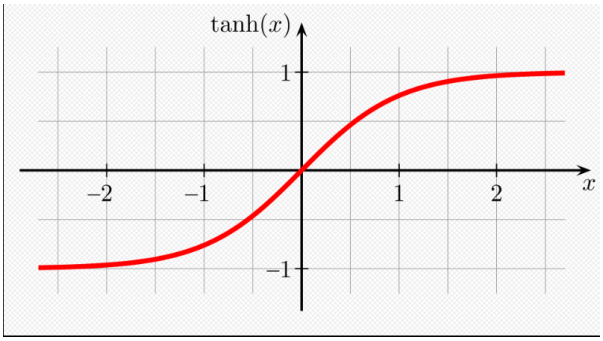

Figura 4.8. Função Tangente hiperbólica.

Fonte:https://pt.wikipedia.org/wiki/Tangente_hiperb%C3%B3lica#/media/Ficheiro:Hyperbolic_Tangent.svg

4.3.3 A Função Linear Retificada (*Rectified Linear Unit* - ReLU)

A função ReLU é definida como

$$\varphi(x) = max(0, x)$$

e toma valores entre 0 e $+\infty$. Ou seja, toma o valor 0 (ou seja, o neurónio não é ativado) para os valores negativos e toma o próprio valor para valores positivos.

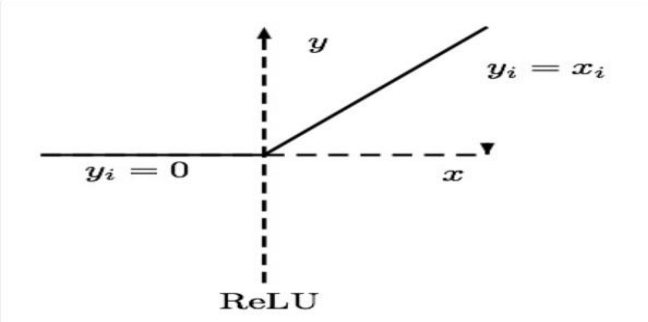

Figura 4.9. Função ReLU.

Fonte: https://matheusfacure.github.io/2017/03/05/ann-intro/

"Como a função ReLU toma o valor 0 para todos os valores negativos, e o próprio valor para valores positivos resulta numa função computacionalmente leve, entretanto não é centrada em zero. Como seu resultado é zero para valores negativos, ela tende a "apagar" alguns neurónios

durante um passo *forward*, o que aumenta a velocidade do treino, mas por outro pode fazer com que esses neurónios "morram" e não aprendam nada se eles só receberem valores negativos. Além disso, ela pode produzir ativações explosivas, já que não possui um limite positivo. Mesmo com suas limitações, a função ReLU é hoje uma das funções de ativação mais utilizadas no treino de redes neurais, e não costuma ser utilizada na camada de saída" (https://iaexpert.academy/2020/05/25/funcoes-de-ativacao-definicao-caracteristicas-e-quando-usar-cada).

A função ReLU permite o treino das redes neuronais supervisionadas sem a necessidade de treino não supervisionado, em comparação com as demais funções, e permite um treino mais rápido e efetivo para um conjunto de dados complexos [25].

No trabalho Gomes, Ludermir e Lima de 2011, podem ser encontradas estas funções de ativação como muitas outras, especialmente focadas na previsão de séries temporais económicas [32].

Em resumo, a arquitetura de uma RNA está relacionada com o tipo de problema a ser estudado pela rede e para a qual é necessário definir:

- ✓ Tipo de rede;
- ✓ Número de camadas da rede;
- ✓ Número de nós em cada camada;
- ✓ Tipo de conexão entre os nós;
- ✓ Tipo de treino;
- ✓ Dados de treino e os dados de teste;
- ✓ Normalização dos dados.

Mais detalhes sobre o funcionamento das redes neuronais e, em particular, das redes neuronais recorrentes, pode ser encontrado em [26].

4.5 Algoritmos de memória de longo prazo

As RNAs recorrentes (RNNs) de multicamadas inserem-se na chamada **aprendizagem profunda** (***Deep Learning***).

Dentro das RNNs de aprendizagem profunda temos uma variante designada por **LSTM** (*Long Short-Term Memory*). Este tipo de redes tem uma "memória" interna pelo que são as que melhor se adequadam no estudo de séries temporais, onde as observações possuem uma estrutura de

autocorrelação. Este tipo de rede é adequado para captar memória curta, mas também memória a longo prazo.
As LSTM possuem células (ou unidades de memória) de modo a poderem guardar, ou esquecer, informação ao longo do tempo. Cada uma destas células possui três portas principais, ou portões (*gates*): a porta de entrada (*input gate*), a porta de esquecimento (*forget gate*) e a porta de saída (*output gate*).

Numa RNA recorrente "tradicional", os valores entram na porta de entrada onde é atribuído um peso a cada valor, em função da sua importância, e um viés (transformação linear). Depois esta informação é transformada à custa da função de ativação (a transformação não linear) e passada para a(s) camada(s) intermédia(s) (ou oculta(s)). Esta informação volta a ser pesada (ponderada por novos pesos), passada pela função de activação e volta a entrar na rede até que o erro seja mínimo e daí resultam os valores de saída.

De igual modo, nas redes LSTM mais complexas temos a **porta de entrada** que não só recebe, num dado instante de tempo *t*, os valores originais, como os valores de saída da camada anterior (camada oculta, designada por h_{t-1}) como os valores atualizados da sua própria camada (ou seja, os valores são multiplicados pelas informações de entrada e somados ao estado atual da célula), como também neste caso a porta de entrada decide qual a informação que deve ser esquecida e a informação que deve ser armazenada. Isto acontece porque existe agora uma **porta de esquecimento** que permite que a rede guarde informação importante ou que elimine informação que não serve.

De igual modo, na porta de esquecimento cada valor é ponderado em função da informação a ser esquecida. Depois esses valores são multiplicados pelo estado atual e passados por uma função de ativação (de esquecimento).
Por fim, a informação é enviada para a **porta de saída** onde se avalia se o valor é o final ou se vai enviado para a próxima camada oculta.

Neste tipo de redes existem dois tipos de funções de ativação, a função sigmóide e a tangente hiperbólica, como se pode observar na figura 4.10.

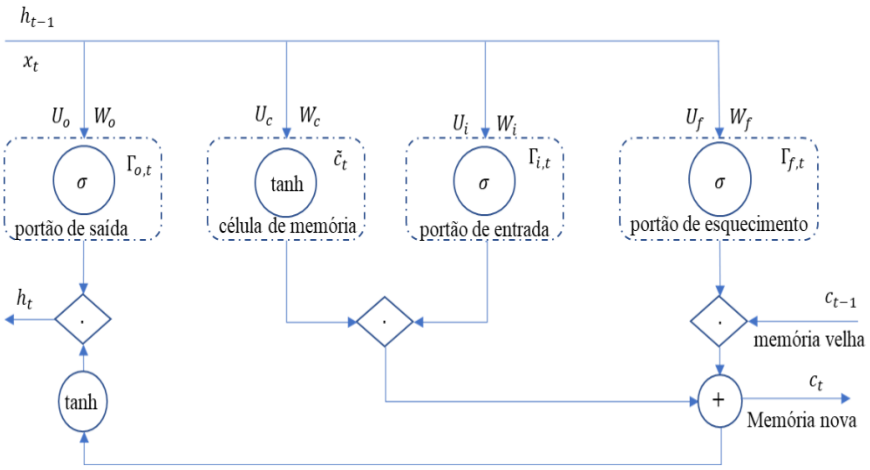

Figura 4.10. Rede de memória de Longo Prazo (LSTM).

Fonte: [29].

Seja $x = (x_1, \ldots, x_T)$ a sequência de entrada na rede até ao tempo *T*, com x_t a entrada na rede no instante $t \epsilon T$. W representa a matriz dos pesos entre as entradas da rede e as portas, *U* é a matriz dos pesos entre os valores dos estados intermédios e as portas. $\Gamma_{0,t}$, $\Gamma_{i,t}$ e $\Gamma_{f,t}$ representam, respectivamente, as portas de saída, de entrada e de esquecimento no instante *t*; $\tilde{c}_t$ representa a célula de memória no instante *t*; h_t é o estado oculto no instante t; σ e $tanh$ (tangente hiperbólica) são as funções de ativação, não lineares, geralmente implementadas como funções sigmóides e tangentes hiperbólicas.

As portas de saída, de entrada e de esquecimento, no instante *t*, são definidas como

$$\Gamma_{j,t} = \sigma(W_j x_t + U_j h_{t-1} + b_j)$$

onde o índice *j* representa a entrada (*j=i*), o esquecimento (*j=f*) e a saída (*j=0*) e b_j o correspondente vetor de viés,

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

e

$$h_t = \Gamma_{0,t}\, tanh(c_t)$$

Informação mais detalhada sobre o funcionamento da rede LSTM pode ser encontrada, por exemplo em [27, 28].

Existe ainda um tipo de rede LSTM mais simples, a rede **GRU** (*Gated Recurrent Units*). Esta rede recorrente (mas fechada) tem apenas duas portas, contrariamente à rede LSTM, a porta de eliminação (*reset*) e a porta de atualização. A primeira decide "quanto do estado oculto anterior deve ser mantido, e a porta de atualização decide quanto da nova entrada deve ser incorporada ao estado oculto. O estado oculto também atua como o estado da célula e a saída, portanto, não há uma porta de saída separada. A GRU é mais fácil de implementar e requer menos parâmetros do que a LSTM." [27].

"Geralmente, o LSTM é mais poderoso e flexível do que o GRU, mas também é mais complexo e propenso a *overfitting*. A GRU é mais rápida e eficiente do que a LSTM. Ambas podem aprender dependências de longo prazo e capturar padrões sequenciais nos dados." [https://pt.linkedin.com/advice/0/what-differences-similarities-between-lstm-gru?lang=pt].

5

Comparação de modelos de séries temporais e algoritmos de redes neuronais (LSTM)

5.1 Introdução

Neste capítulo serão apresentados os resultados de análise dos dados semanais do número de casos e de mortes, provocados pela Covid-19 na Guiné-Bissau entre 26/03/2020 a 15/04/2023, que corresponde a 159 semanas de dados, através de modelos de séries temporais do tipo ARIMA/SARIMA e através de algoritmos de redes neuronais recorrentes (RNNs), em especial do LSTM. Os dados utilizados neste trabalho foram retirados dos bancos de dados da OMS da pandemia de Covid-19 no mundo https://covid19.who.int/region/afro/country/gw. Este portal da OMS é o principal recurso de dados mundial sobre a pandemia de Covid-19 no mundo. Os dados disponíveis eram dados diários. Como em muitos dos dias quer o número de infecções, quer o número de mortes, eram zero tivemos a necessidade de os agrupar por semana de modo a podermos aplicar os modelos de séries temporais. Para além disso, na análise descritiva, os dados foram agrupados em dados mensais e anuais, de modo a ter uma panorâmica mais geral relativamente ao número de casos e de mortes por Covid-19 na GB.

Antes da modelação e previsão do número de casos e de mortes semanais por Covid-19 foi feita uma análise exploratória de modo a descrever os dados em estudo. Esta análise exploratória incluiu as análises habituais, gráficos e cálculo de medidas, mas também a análise da existência de outliers, análise de valores omissos, análise de padrões de tendência e/ou de sazonalidade.

As análises dos dados foram feitas no *software Rstudio* (versão 4.2.2) para os modelos do tipo ARIMA/SARIMA e *Python* (versão 3.12) para a versão de inteligência artificial em especial

LSTM. No *RStudio* usou-se a biblioteca *forecast* e no *Python* o *Keras* e o *TensorFlow Lite*. Em todas as análises inferenciais usou-se um nível de significância de 5%.

5.2 Número de infeções por Covid-19

De modo melhor resumir os dados, analisou-se o número de infeções (ou casos), por ano e por mês, por Covid-19. Contudo, para implementação dos modelos ARIMA e LSTM os dados foram analisados por semana.

Como só dispomos de 3 anos e 1 mês de dados (que são159 semanas) e, para além disso, os últimos dados são muito atípicos relativamente aos anteriores não resultou viável no caso do ajustamento dos modelos ARIMA usarmos o conjunto de treino para a modelação e o conjunto de teste para a validação das previsões. Isso apenas foi possível fazer-se no caso do LSTM.

5.2.1 Análise descritiva do número de infeções por Covid-19

Na tabela 5.1 apresentam-se as principais medidas descritivas do número de casos de Covid-19, em cada um dos anos em estudo.

Casos	**2020**	**2021**	**2022**	**2023**
Nº casos (Nº semanas)	2447 (40)	4019 (52)	2481 (52)	667 (16)
Média	62.74	77.29	47.71	41.70
Mediana	25	26.5	17	0
Desvio Padrão	87.12	99.48	89.12	79.94
Mínimo	0	0	0	0
Máximo	382	395	430	264

Tabela 5.1. Medidas descritivas do número de casos de Covid-19 por ano.

Foi em 2021 que se registou, em média, o maior número de casos de infecção pelo vírus SARS-CoV-2, mas foi em 2022 que se registou o pico máximo. No ano de 2023 em 50% das semanas não foi registado nenhum caso. Tal pode ter ficado a dever-se não há ausência de casos, mas há falta de registo dos mesmos.

Representando a distribuição do número de casos por mês (figura 5.1) podemos verificar a existência de valores discrepantes em alguns meses, com o mês de janeiro a apresentar os valores mais elevados. Praticamente em todos os meses os dados têm uma distribuição assimétrica positiva, muito acentuada nos meses de janeiro (devido à existência de valores discrepantes), maio e agosto.

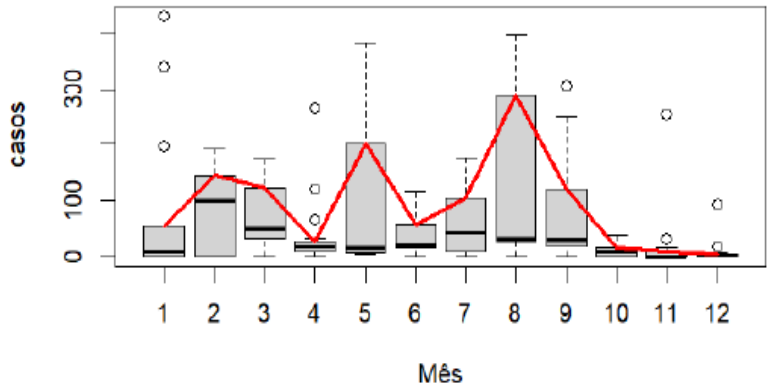

Figura 5.1. Caixa de bigodes do número de casos por mês.

Fazendo agora a média do número de casos por mês (figura 5.2) constata-se que houve um pico nos meses de agosto, com valores muito baixos entre outubro e dezembro. Os meses distinguem-se em termos da sua variabilidade, mas não diferem muito em termos medianos, com o mês de fevereiro a apresentar a mediana mais elevada.

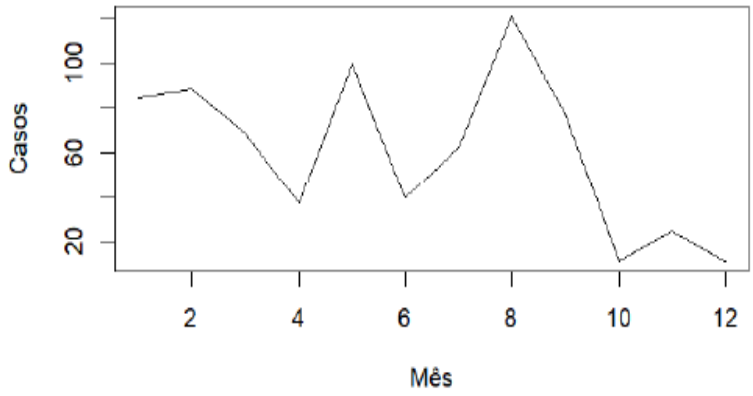

Figura 5.2. Média do número de casos por mês.

A figura 5.3 ilustra a variabilidade do número de casos médios por mês, igualmente com um valor máximo no mês de agosto e valores mínimos nos meses de outubro a dezembro.

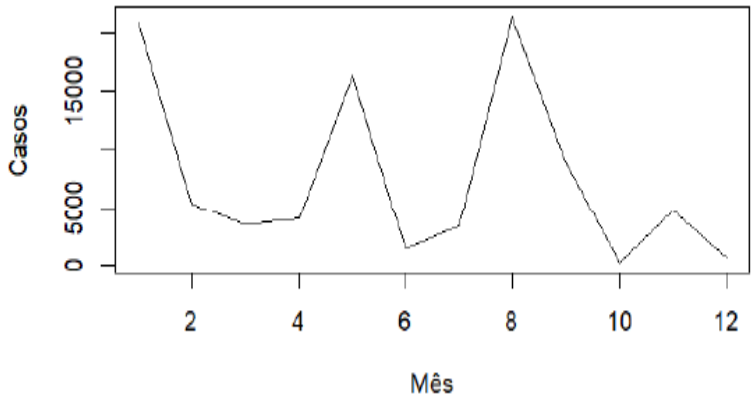

Figura 5.3. Variância do número de casos por mês.

5.2.2 Modelação e previsão a partir de modelos do tipo ARIMA/SARIMA

A figura 5.4 ilustra do número de infetados semanais pelo vírus SARS-CoV-2. Ao longo do tempo houve valores muito elevados (picos de infeção) seguidos de valores muito baixos. A partir de meados de 2022 houve muitas semanas sem que se tivesse registado qualquer valor, mas com alguns picos ocasionais.

Mais detalhadamente, o número dos infetados por vírus tem uma maior proporção até à semana 103 (que corresponde a março de 2022), com uma média de 78,36 casos por semana. Nas restantes 57 semanas o número dos infetados estabiliza, existindo alguns com alguns picos, com uma média de 27,07 casos por semana.

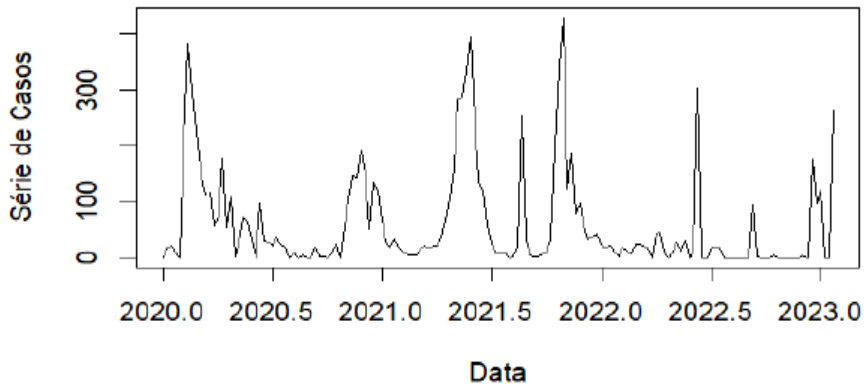

Figura 5.4.Série de número de casos por semanas.

Ao longo das análises foram feitos diversos testes (descritos na secção 2.6) para testar a tendência e a sazonalidade dos dados, tendo-se obtido *p-values* iguais a 0.01, tanto para o teste de ADF como para o teste de KPSS, pelo que podemos afirmar que os dados não apresentam tendência,

ou seja, que são estacionários. O que também se comprova pelo comportamento da FAC (figura 5.5). Tendo em conta esta consideração não foi feita qualquer transformação aos dados.

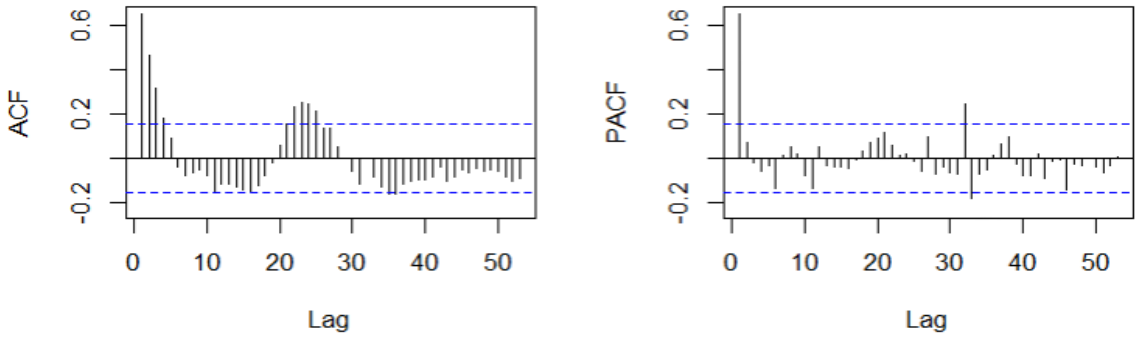

Figura 5.5. FAC e FACP da série do número de infetados por SARS-CoV-2.

Escolheu-se um modelo do tipo ARIMA(1,0,0), ou AR(1), para modelar estes dados e procedeu-se à estimação dos seus parâmetros. Tal ficou a dever-se ao facto da FAC convergir para zero à medida que os *lags* aumentam e a FACP se anular a partir da ordem 2.

Para o número de casos o modelo AR(1) ajustado foi o seguinte

$$X_t = 20{,}39 + 0.67X_{t-1}$$

com variância dos resíduos de 4596.

Depois de analisada a significância dos parâmetros passou-se à análise dos resíduos (figura 5.6). Pode verificar-se que praticamente todos os valores da FAC e da FACP estão dentro das bandas de confiança a 95% pelo que os resíduos se comportam como um ruído branco. Também comprovado pelo teste de Ljung-Box (*p-value* = 0.9953, pelo que não se rejeita a hipótese nula de não correlação dos dados).

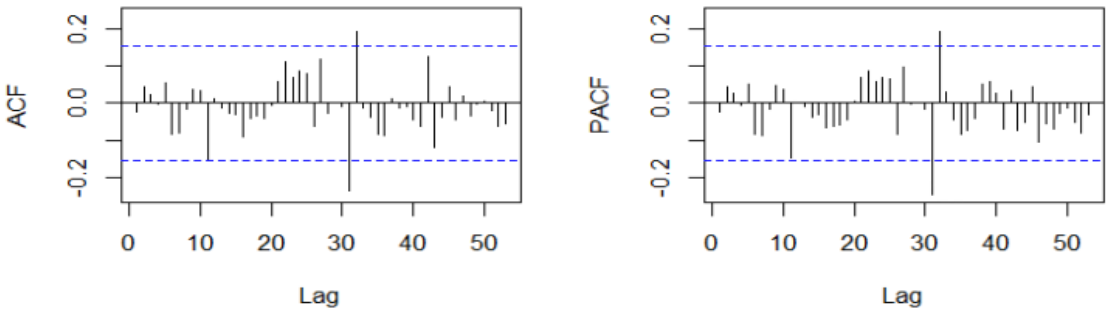

Figura 5.6. Resíduos do modelo AR(1) ajustado à série do número de casos semanais.

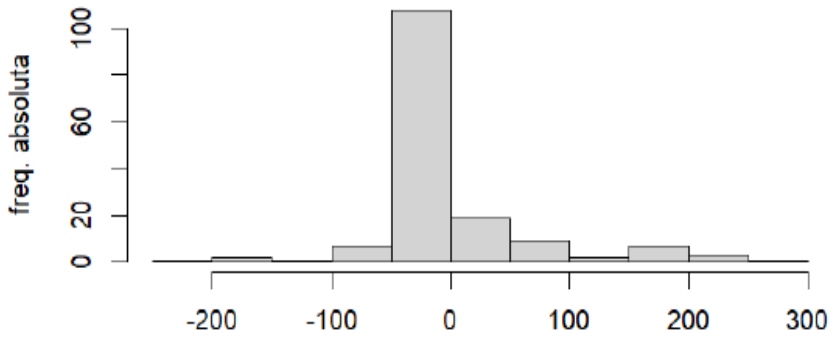

Figura 5.7. Histograma dos resíduos do número de casos semanais.

De seguida, aplicou-se o teste t-student para testar a hipótese de que a média dos resíduos é nula (H_0), obtendo-se um *p-value* de 0.948, pelo que não existe evidência estatística de que a média dos resíduos seja significativamente diferente de zero.

Validado o modelo procedeu-se à previsão para os anos de 2024 e 2025.

Na Figura 5.8 mostra-se o ajustamento do modelo AR(1) estimado aos dados reais (linha a vermelho) e a previsão para os anos de 2024 e 2025 (linha a azul), como os respectivos intervalos de previsão a 80% (cinzento escuro) e a 95% (cizento claro). As previsões indicam que nos anos 2024 e 2025 o número de casos tende a estabilizar em valores próximos de zero.

As previsões são feitas admitindo que o futuro é parecido com o presente e o passado recente. Como a partir de meados de 2022 muitas semanas registaram valores muito pequenos ou nulos (não sabemos se por falta de registos ou porque realmente não ocorreram infeções) é normal que as previsões se comportem assim.

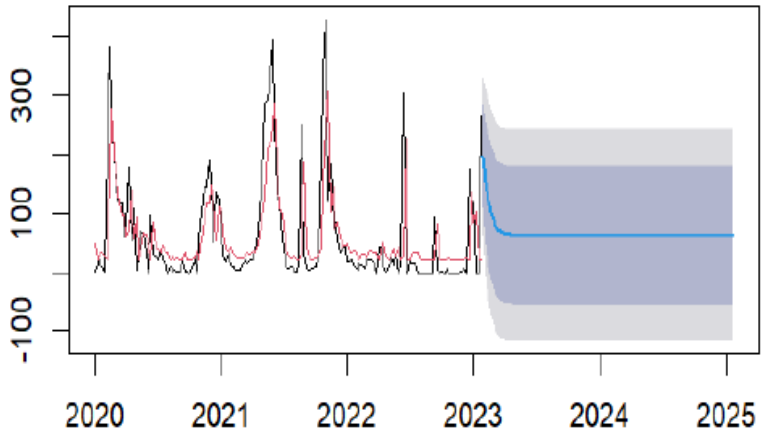

Figura 5.8. Previsão e ajustamento do número de casos semanais por Covid-19.

Tendo em consideração as medidas de precisão do ajustamento através do modelo AR(1) estimado (tabela 5.2) temos, por exemplo, um erro médio (ME) de ajustamento bastante baixo o que revela

que o modelo estimado se ajusta bem aos dados. Ou seja, o modelo consegue capturar os valores observados mesmo tendo em conta a pequena dimensão da série e a existência de valores muito baixos (por vezes zero) do número de casos de Covid-19. Por este motivo (presença de zeros), não foi possível calcular-se uma das medidas mais adequadas, que é o MAPE (e por isso o valor **NA** apresentado na tabela abaixo)

	ME	**RMSE**	**MAE**	**MAPE**
ARIMA/SARIMA	0.3491134	67.36887	41.57604	**NA**

Tabela 5.2. Valores das medidas da precisão do ajustamento do número de casos semanais.

5.2.3. Modelação e previsão a partir do algoritmo LSTM

A segunda parte do trabalho consiste em aplicar o algoritmo LSTM para a modelação e previsão do número de infeções e de mortes por Covid-19 na Guiné-Bissau. Para tal, usou-se o *software Python,* versão 3.12.

Em primeiro lugar, começou por re-escalonar-se os dados relativos ao número de casos (como posteriormente será efectuado para o número de mortes). Existem dois tipos de re-escalação dos dados: a normalização e a standardização (ou padronização).

Na **normalização** os dados são redimensionados (por exemplo, deixam de estar entre o mínimo e o máximo para estarem dentro do intervalo entre 0 e 1). Na **padronização** os dados são centrados na média e divididos pelo desvio-padrão. Neste caso assume-se que os dados se ajustam bem a uma distribuição normal.

Como os nossos dados são assimétricos positivos fizemos uma normalização. Mais concretamente

$$y_t = \frac{x_t - min}{max - min}$$

onde x_t é a observação original no instante *t* e o *min* e o *max* são, respectivamente, os valores mínimo e máximo globais observados. Pelo que os dados transformados passaram a pertencer ao intervalo entre 0 e 1.

No segundo passo foi preciso determinar os dados de treino e os dados de teste. Para os dados de treino usou-se 65% da informação (103 observações iniciais) e para os dados de teste 35% (56 observação finais). A mesma abordagem será feita para o número de mortes (analisados em 5.2 e 5.6).

No processo de treino e teste usou-se o algoritmo supervisionado. Para a implementação do modelo LSTM foram importadas diversas bibliotecas (*tensorflow.keras.models*; *keras.layers;*

keras.models). A função de ativação da rede usada foi a ReLU. O número de nodos ou neurónios que são ativados foi de cinquenta, o número de entrada da informação é 1. Resultou num total de 50851 nós, como se pode observar pela tabela 5.3.

As perdas e a exatidão dos dados treinados e de teste foram representadas utilizando os comandos *plt.plot()* e o *model.history*. Isto é importante para saber como é que o modelo convergiu ao longo das análises. Tanto o ajustamento como as previsões foram obtidas na escala original dos dados, depois de efetuadas as transformações inversas. As sequências de camadas apresentadas na tabela 5.3 são geradas automaticamente pela rede tendo em conta o número de entradas determinado pelo analista.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 1, 50)             10400

 lstm_1 (LSTM)               (None, 1, 50)             20200

 lstm_2 (LSTM)               (None, 50)                20200

 dense (Dense)               (None, 1)                 51

=================================================================
Total params: 50851 (198.64 KB)
Trainable params: 50851 (198.64 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Tabela 5.3. Modelos de treino de LSTM.

Com base na Figura 5.9, relativa ao número de casos semanais de Covid-19, podemos observar o ajustamento aos dados de treino e aos dados de teste através do algoritmo de LSTM. Na tabela 5.4 temos os valores das medidas de precisão.

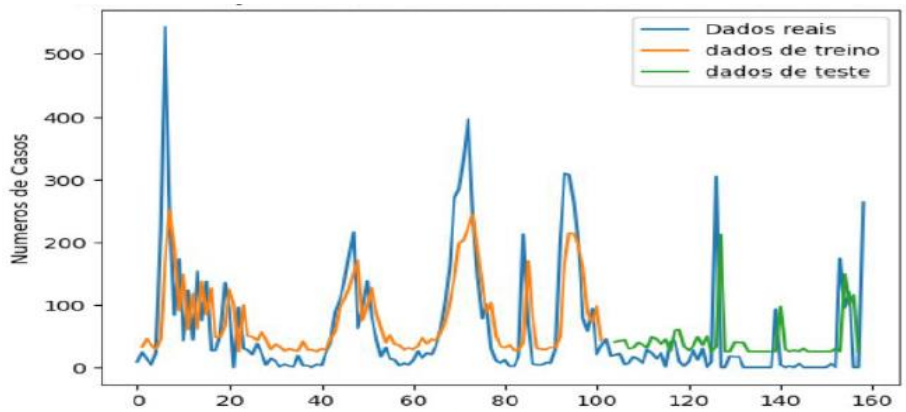

Figura 5.9. Ajustamento e previsão do número de casos através do algoritmo LSTM.

Em resumo, dos métodos de ajustamento utilizados constata-se um pior desempenho do modelo LSTM em relação ao ARIMA, onde apresentou, por exemplo, um erro de ajustamento de **ME**= 53.7. Os dados não só têm poucas observações como variam bastante, oscilando entre valores demasiado elevados comparados com valores demasiado baixos, por vezes nulos ou quase nulos. Estes aspectos dificultam a aprendizagem do algoritmo e daí os elevados valores das medidas de precisão. Contudo, podemos afirmar que o modelo capta bem a variabilidade dos dados.

	RMSE	**MAE**	**MAPE**	**ME**
LSTM	135.2	125.1	113.4	53.7

Tabela 5.4. Medidas de precisão do ajustamento através do algoritmo LSTM para os casos por Covid-19.

5.3 Número de mortes por Covid-19

De modo a melhor resumir os dados analisou-se o número de óbitos (ou mortes), por ano e por mês, por Covid-19. Contudo, para implementação dos modelos ARIMA e LSTM os dados serão analisados por semana.

Como só dispomos de 3 anos e 1 mês de dados (que corresponde 159 semanas) não usamos o conjunto de treino para a modelação e o conjunto de teste para a validação das previsões, isto no caso da modelação do tipo ARIMA.

5.3.1 Análise descritiva do número de mortes por Covid-19

Na tabela 5.5 apresentam-se as principais medidas estatísticas dos dados relativos às mortes por Covid-19. Como se pode observar, o ano em que se registou o maior número de mortes foi o ano de 2021, com uma média de 2 mortes por mês, como um máximo de 14 mortes.

Mortes	**2020**	**2021**	**2022**	**2023**
Nº mortes (Nº de semana)	70 (4)	104 (52)	25 (52)	1 (16)
Média	1.16	2	0.50	0.06
Mediana	1	1	0	0
Desvio Padrão	1.45	3.21	1.09	0.25
Mínimo	0	0	0	0
Máximo	5	14	6	1

Tabela 5.5. Descrição estatística dos dados não transformados.

Na figura 5.10 representou-se o número de mortes por mês, tendo-se verificado, tal como no número de casos, um grande número de valores discrepantes (com os valores mais elevados nos meses de agosto e setembro). Os meses de agosto e setembro foram também os meses que apresentaram as medianas mais elevadas.

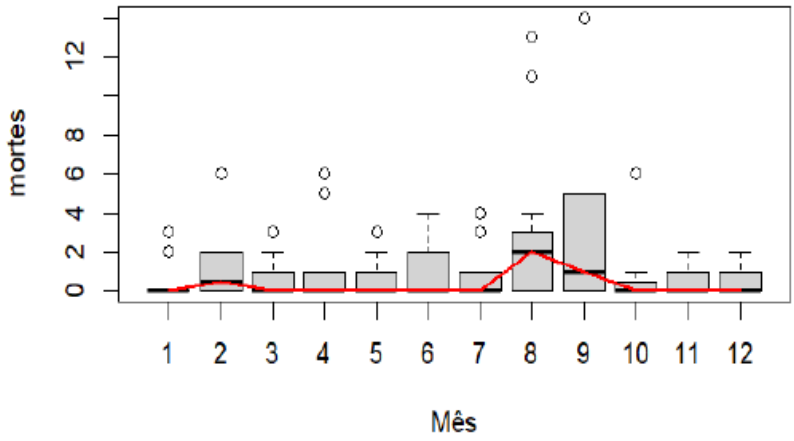

Figura 5.10.Caixa de bigodes do número de mortes por meses.

Na figura 5.11, constata-se que o número de mortes por Covid-19 não apresenta uma média constante ao longo dos meses, com uma subida brusca nos meses de julho, setembro e outubro, como já tínhamos visto na figura anterior (relativamente às medianas).

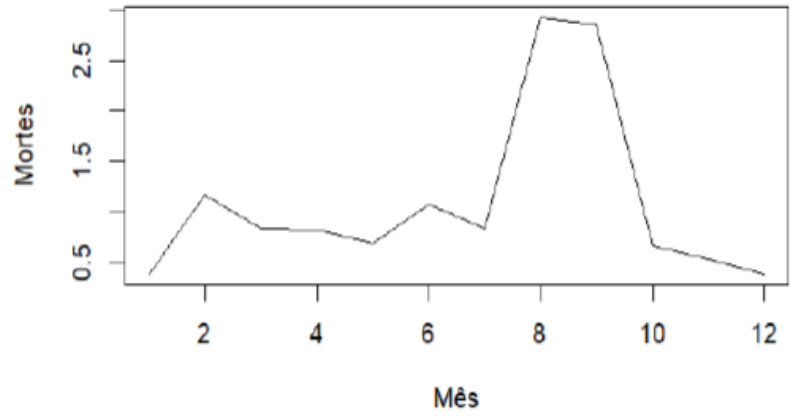

Figura 5.11. Média do número de mortes por meses.

Na Figura 5.12, constata-se que o número de mortes apresenta igualmente uma variância não constante ao longo dos meses, com valores superiores a 5 nos meses de julho, agosto e setembro (sendo o mês de agoro aquele que registou a subida mais brusca), e inferior a 5 nos restantes meses.

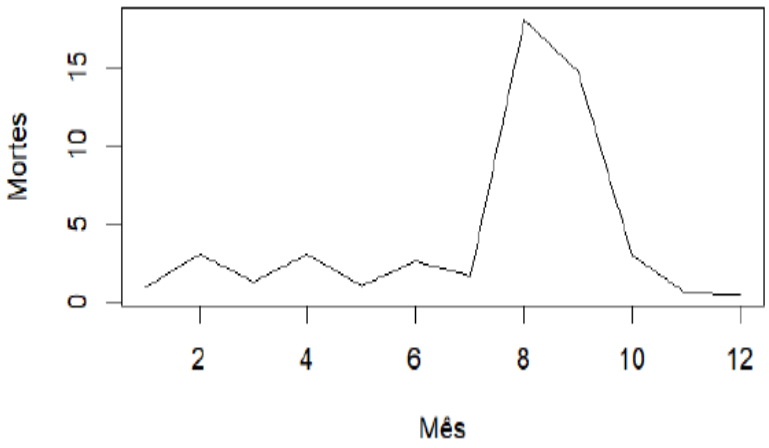

Figura 5.12. Variância do número de mortes por meses.

5.3.2 Modelação e previsão a partir de modelos do tipo ARIMA/SARIMA

O número de mortes por vírus tem uma maior frequência até à semana 92 (que corresponde a dezembro de 2021), com uma média de 2 óbitos por semana. Nas restantes 67 semanas as mortes ou deixam praticamente de ocorrer (ou deixam de ser registadas) existindo, contudo, algumas ocorrências (com uma média de 0,5 óbitos por semana).

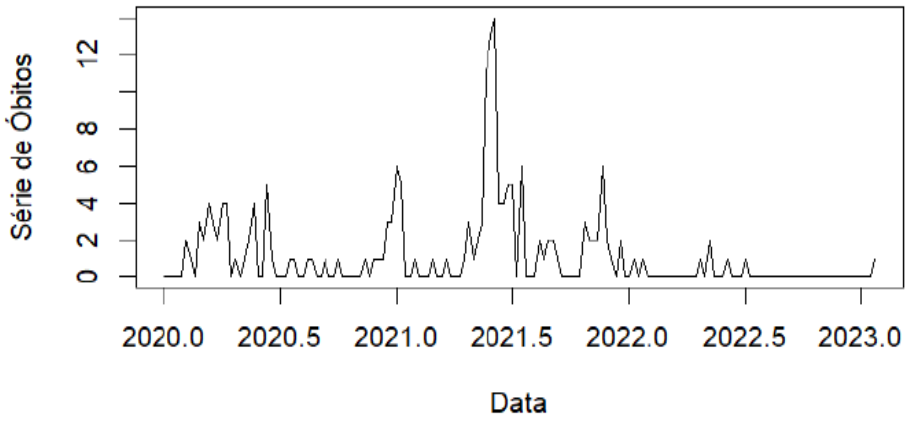

Figura 5.13.Série do número de mortes semanais.

Como se pode observar pela figura 5.13 o número de mortes tem valores muito baixos nos anos de 2022 e 2023, com um pico no ano 2021, o ano de maior número de mortes.

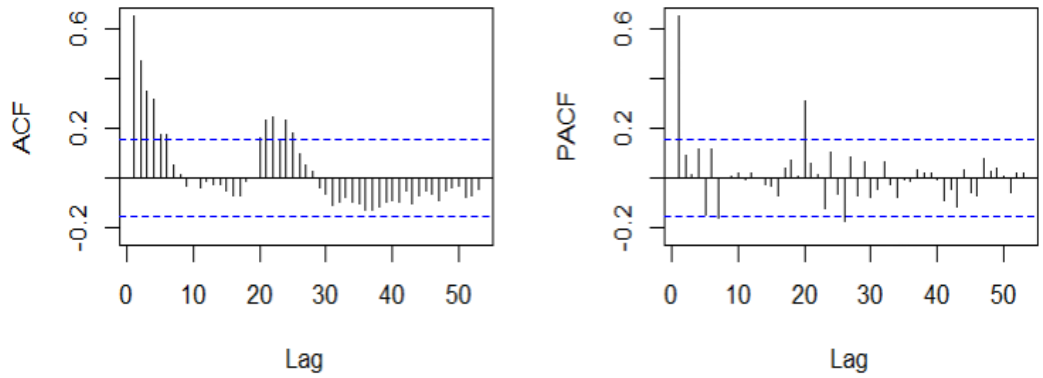

Figura 5.14. FAC e FACP do número de mortes semanais.

Usualmente numa modelação estatística para séries temporais os dados devem ser divididos em treino e teste. Contudo, para o presente trabalho os dados não podem ser divididos em treino e teste devido ao elevado número de zeros que as bases de dados apresentam, tantos as dos casos como as das mortes.

Foram realizados os testes mencionados na secção 2.6 para testar a hipótese da existência de tendência, tendo-se obtido evidência estatística suficiente para afirmar que a série não é estacionária (apresentando uma tendência decrescente). Tendo em conta esta evidência de não estacionariedade da série foi efetuada uma transformação de Box-Cox ($\lambda = 0.2897$) e, de seguida, fez-se uma única diferenciação simples para estabilizar a tendência da série (média e variância).

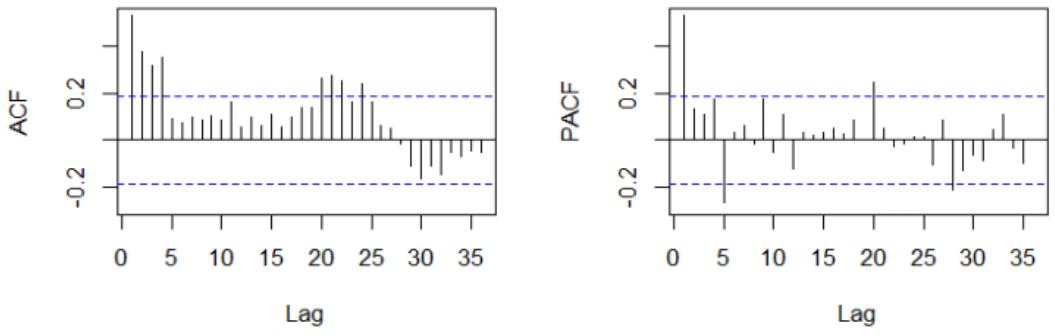

Figura 5.15. FAC e FACP da série transformada do número de mortes semanais.

Através da análise das FAC e FACP começámos por ajustar um modelo AR(5) aos dados transformados. Contudo, os coeficientes do modelo de ordens 3 a 5 revelaram-se não significativos. Ficando, portanto, um modelo do tipo AR(2) de parâmetros $\tilde{\phi}_1 = -0{,}5492$ e $\tilde{\phi}_2 = -0{,}2818$ (com variância dos resíduos igual a 3,902).

Assim, obtivemos para os dados das mortes um modelo ARIMA(2,1,0) que se escreve da seguinte forma

$$X_t = (1 + \tilde{\phi}_1)X_{t-1} + (\tilde{\phi}_2 - \tilde{\phi}_1)X_{t-2} - \tilde{\phi}_2 X_{t-3}$$

ou seja,

$$X_t = 0{,}4518X_{t-1} + 0{,}2674X_{t-2} + 0{,}2818X_{t-3}$$

cujos resíduos se comportam como um ruído branco (Figura 5.16) de média nula (*p-value* igual a 0.9897 relativo ao teste t-student)

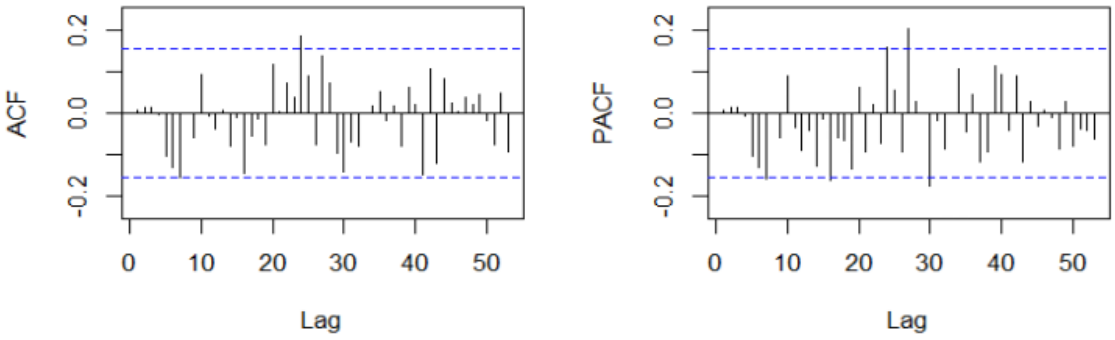

Figura 5.16. Resíduos do modelo ARIMA(2,1,0) ajustado à série do número de mortes semanais.

Os resíduos apresentam uma distribuição aproximadamente simétrica (Figura 5.17)

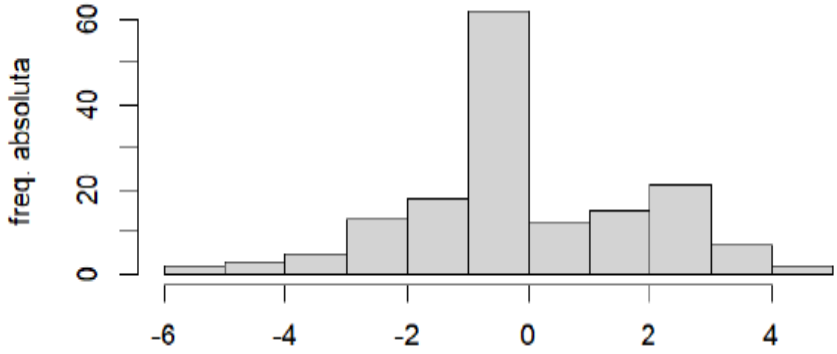

Figura 5.17. Histograma dos resíduos do modelo ARIMA(2,1,0) ajustado ao número de mortes semanais.

Foi também realizado o teste de autocorrelação de Ljung-Box, para testar a não autocorrelação dos resíduos (*p-value* = 0.9186). Podemos afirmar que o modelo ajustado parece adequado.
Com base na figura 5.19, mostra-se que o modelo se ajusta bem aos dados. Ou seja, capta bem a variabilidade destes ao longo dos anos em estudo. Já as previsões para 2024 e 2025 são quase (ou mesmo) nulas pois as previsões são feitas partindo do pressuposto que os valores futuros são muito semelhantes ao presente e passado mais recente. Como os valores de 2023 são muito baixos

e em muitos vezes nulos, os valores previstos são os espectáveis caso o padrão anterior se mantivesse.

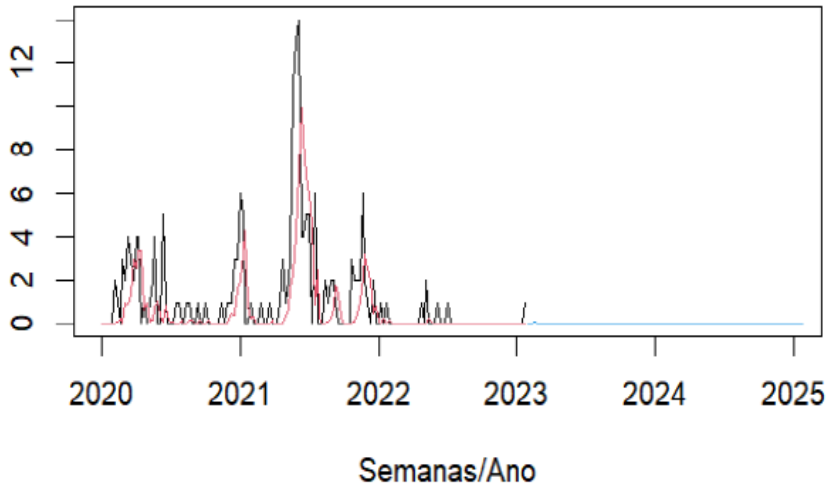

Figura 5.18. Previsão e ajustamento do número de mortes semanais por Covid-19.

Tendo em consideração as medidas de precisão do ajustamento o modelo ARIMA(2,1,0) estimado (tabela 5.6) temos, por exemplo, um erro médio (ME) de ajustamento bastante baixo o que revela que o modelo estimado se ajusta bem aos dados. Ou seja, o modelo consegue capturar os valores observados mesmo tendo em conta a pequena dimensão da série e a existência de valores muito baixos (por vezes zero) do número de óbitos por Covid-19. Por este motivo (presença de zeros), não foi possível calcular o MAPE (e por isso o valor de NA apresentado na tabela abaixo).

	ME	**RMSE**	**MAE**	**MAPE**
ARIMA/SARIMA	0.4206067	1.827938	0.9637228	***NA***

Tabela 5.6. Precisão da previsão do número de mortes semanais.

5.3.3 Modelação e previsão a partir do algoritmo LSTM

O método de aprendizagem e treino utilizado é o modelo supervisionando, tal como citado anteriormente para os dados do número de casos. Para os dados de mortes são implementados os mesmos padrões de análise e as mesmas bibliotecas do *python*. O total de parâmetros treinados para os dados são de 113776, como se pode observar pela tabela 5.7.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 1, 75)             23100

 lstm_1 (LSTM)               (None, 1, 75)             45300

 lstm_2 (LSTM)               (None, 75)                45300

 dense (Dense)               (None, 1)                 76

=================================================================
Total params: 113776 (444.44 KB)
Trainable params: 113776 (444.44 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Tabela 5.7. Modelos de treino de LSTM.

Com base na Figura 5.19, para a previsão de mortes por Covid-19 através do modelo de LSTM, o número de óbitos por vírus tem uma maior proporção entre as semanas de 20 e 80 (correspondente aos anos 2020-2021), com uma média de 2 óbitos por semana.

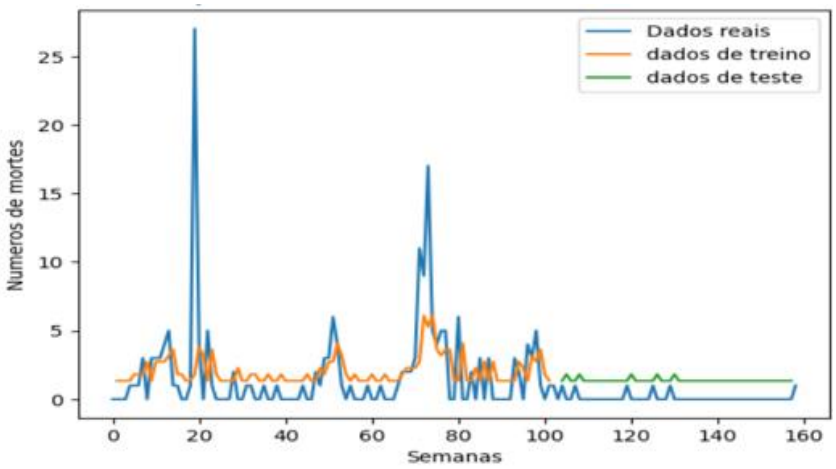

Figura 5.19. Previsão e ajustamento do número de mortes através do algoritmo LSTM.

Em resumo, dos métodos de ajustamento utilizado constata-se igualmente um pior desempenho do algoritmo LSTM, relativamente ao modelo ARIMA. O comentário feito para o caso dos infectados aplica-se ao caso das mortes.

	RMSE	MAE	MAPE	MSE	ME
LSTM	2.28	1.87	3.40	7.94	2.05

Tabela 5.8. Precisão do ajustamento dos resultados do modelo LSTM relativos às mortes por Covid-19.

5.4 Comparação entre os modelos ARIMA e o algoritmo LSTM

De forma a resumir o desempenho dos modelos ARIMA e LSTM são apresentadas as medidas de precisão para os dados ajustados relativamente aos infetados e às mortes por Covid-19.
Como no caso dos modelos do tipo ARIMA não foi possível usar um conjunto de teste para validação das previsões as medidas de precisão destas não serão aqui apresentadas para o caso o LSTM, pois não existe como comparar as duas abordagens em termos de capacidade de previsão. Em todo o caso, dada a possível falta de qualidade dos dados, bem como o número muito reduzido de casos e de mortes reportados nas últimas semanas de dados também não será de esperar boa capacidade de previsão.

Modelos	RMSE	MAE	ME	MAPE
ARIMA	67.36	41.57	0.35	**NA**
LSTM	135.2	125.1	53.7	113.4

Tabela 5.9. Resumo dos resultados dos diferentes métodos para a série dos casos.

Modelos	RMSE	MAE	ME	MAPE
ARIMA	1.8	0.9	0.4	**NA**
LSTM	2.2	1.8	2.1	3.4

Tabela 5.10. Resumo dos resultados dos diferentes métodos para a série das mortes.

Com base nestes resultados podemos considerar que o modelo tradicional de série temporal apresenta um melhor desempenho relativamente ao algoritmo LSTM, tanto para o caso dos infetados como para o caso das mortes.

Apesar dos valores do ME terem os valores muito inferior sendo possíveis de comprovar que modelos ARIMA podemos manter esta afirmação sobre a maior capacidade de previsão para o modelo ARIMA.

6

Considerações finais, limitações e trabalho futuro

As análises e previsões de séries temporais têm evoluído bastante com o passar do tempo. Onde no princípio se utilizavam apenas as observações para compreender a evolução de um fenómeno, passando depois pela utilização de simples médias, depois para a utilização de métodos matemáticos (lineares e não-lineares) e, hoje em dia, com os algoritmos de inteligência artificial. Para este trabalho foram aplicados dois tipos de técnicas de análise de séries temporais, os modelos do tipo ARIMA/SARIMA e as redes neuronais recorrentes, nomeadamente o LSTM.

Neste trabalho foram aplicadas estas metodologias ao conjunto do número de casos e mortes por Covid-19 na Guiné-Bissau entre 26/03/2020 a 15/04/2023. Estes dados foram retirados dos bancos de dados da OMS. Apesar da OMS ser o principal recurso de dados mundial sobre a pandemia de Covid-19 no mundo, os dados relativos à Guiné-Bissau podem não corresponder à realidade do que efectivamente ocorreu durante o período da pandemia. Isto tendo em conta que a Guiné-Bissau é um dos países mais pobre do mundo, com uma elevada instabilidade política, um baixo nível de condições sanitárias e com uma economia e estuturas deficientes, logo uma falta de capacidade logística para lidar com a pandemia e ter dados mais fiáveis em todo o território nacional. Daí, que as interpretações dos modelos ajustados e das correspondentes previsões tenham de ser feitas com o cuidado necessário devido a possível falta de confiabilidade dos dados que são utilizados para este trabalho.

Para além da falta de confiabilidade nos dados, estes não só têm poucas observações como variam bastante, oscilando entre valores demasiado elevados comparados com valores demasiado baixos, por vezes nulos ou quase nulos. Estes aspectos dificultam a escolha de conjuntos de treino e de teste adequados, bem como a aprendizagem do algoritmo de redes neuronais.

Apesar de tudo isto, podemos afirmar que ambas as metodologias conseguiram captar bastante bem a variabilidade dos dados.

É de salientar que o uso dos métodos tradicionais de séries temporais e de inteligência artificial para análise e previsão de dados sequenciais são muito importantes em diversas áreas pelo que para a minha formação foi muito importante aprofundar o conhecimento e a aplicação dos modelos do tipo SARIMA e entre outros métodos de séries temporais, bem como aprender novas metodologias como são as redes neuronais. Também gostaria de salientar o aprofundamento da aprendizagem na linguagem *Rstudio* e *Python*, bem como para a aprendizagem sobre inteligência artificial, sobre a qual nada sabia (para além do que ouvia nas notícias) antes de ter iniciado esta tese de mestrado.

Como trabalho futuro, tem-se como sugestão nessa linha de pesquisa a investigação de fatores relacionados com os dados epidemiológicos com o intuito de melhor compreender quais as regiões e/ou os estratos da população da Guiné-Bissau mais afectadas pelo Covid-19.
Outra possibilidade de trabalho futuro seria a combinação de modelos/metodologias de séries temporais (por exemplo, decomposição clássica ou alisamento exponencial) juntamente com os algoritmos de redes neuronais.

Por fim, tem-se como oportunidade relevante a utilização da metodologia para o desenvolvimento de pesquisas com outros conjuntos de dados em demais áreas, nomeadamente para os estudos de epidemiologia. Entende-se que o conhecimento aqui adquirido pode ser adequado a demais estudos para as áreas de estatística e de inteligência artificial para posteriores pesquisas nos estudos das doenças epidemiológicas existentes na África subsaariana, em especial para Costa Ocidental da África.

Referências Bibliográficas

[1] Monié, F. (2020). *Sub-Saharan Africa facing the COVID-19 pandemic: spatial diffusion, impacts and challenges.* http://journals.openedition.org/espacoeconomia/13629; DOI: https://doi.org/10.4000/espacoeconomia.13629

[2] Ceylan, Z. (2020). *Estimation of COVID-19 prevalence in Italy, Spain, and France*. Science of The Total Environment, v. 729.

[3] Shumway, R.H.; Stoffer, D.S. (2017). *Time series analysis and its applications*. *Fourth edition.* Springer.

[4] Staudemeyer R.C.; Morris E.R. (2019). *Understanding LSTM -a tutorial into Long Short-Term Memory recurrent neural networks.* Schmalkalden University of Applied Sciences, Germany; Singapore University of Technology and Design, Singapore.

[5] Silva, B.A. (2010). *Urbanização na Guiné-Bissau*: Morfologia e estrutura urbana da sua capital. Dissertação (Mestrado em Arquitetura e Urbanismo) – *Universidade Lusófona de Humanidades e Tecnologias, Lisboa.*

[6] Instituto Nacional De Estatistica- INE (2009). *Recenciamento geral da população e habitação na Guiné-Bissau*. Ministerio da Economia, do Plano e da Integração Regional.

[7] Malomalo, B.; Caetan, L.P.; Oliveira, R.L.; Tamba, T.; Gomes, P.M.; Vaz, F.; Monteiro, I. (2020). *O Covid-19 em Guiné-Bissau: conjuntura econômica, social e política do país e a garantia dos direitos sociais.*

[8] de Melo, N.A.; Ferreira, P.A.S.C.; Sá, E.; Salla, M.R. (2019). *The relationship between basic sanitation and public health in Bissau, Guinea-Bissau. Universidade Federal de Uberlândia.* Faculdade de Engenharia Civil. Departamento de Hidráulica e Saneamento. Uberlândia, MG, Brasil.

[9] Gabinete Integrado das Nações Unidas para a Consolidação da Paz na Guiné-Bissau-UNIOGBIS (2017). *Relatório sobre direito a saúde na Guiné-Bissau*. Bissau. Disponível em: <https://bit.ly/2LIIMWX>. Acesso em: 28 fev. 2019.

[10] Morettin, P; Toloi, C.M. (2006). *Análise de séries temporais*. 2ª Edição Revista e Ampliada. 2ª Edição, Editora Edgar Blucher, São Paulo.

[11] Gonçalves, E.; Lopes, N.M. (2008). *Séries temporais Modelações lineares e não lineares*. Sociedade Portuguesa de Estatística. 2ª Edição.

[12] Gomes, D. (2022). *Aulas de séries temporais.* Departamento de Matemática. Escola de Ciência e Tecnologia-Universidade de Évora.

[13] Bueno, R. (2012), *Econometria de séries temporais*, 2º edição, Cengage Learning, São Paulo.

[14] Pires, A.P. (2001/01). *Nota de séries temporais 2000/01*. Instituto Superior Técnico, Universidade de Lisboa.

[15] Alpuim, T. (1998). *Séries temporais.* Associação dos Estudantes da Faculdade de Ciências de Lisboa, 2ª Edição.

[16] Sjoster, L. (2022). *A Comparative study of the KPSS and ADF test in terms of size and Power.* Bachalor`s thesis in Statistics-UPPSALA Universitet.

[17] Hyndman, R.J.; Athanasopoulos, G. (2021). *Forecasting:* Principles and practice (3ª ed.) Australia: OTexts.

[18] Pertanika J.; Technol, S. (2002). *The Performance of AICC as an Order Selection Criterion in ARMA time series models*. Department of Mathematics, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.

[19] de Myttenaere, A.; Golden, B.; Le Grand, B.; Rossi, F. (2015). *Using the mean absolute percentage error for regression models.* Université Paris 1 Panthéon - Sorbonne - Centre de Recherche en Informatique 90 rue de Tolbiac, 75013 Paris – France. HAL Id: hal-01162980 https://hal.science/hal-01162980

[20] Moreira, C. (2013). Neurónio. *Revista de Ciência Elementar* 1(01):0006. Casa das Ciências-Porto.

[21] Pereira, E.B.; Fiorin, D.V.; Martins, F.R.; Schuch, N.J. (2011). *Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares*. *Revista Brasileira de Ensino de Física.* www.sbfísica.org.br.

[22] Rauber, T.W. (2005). *Redes neurais artificiais.* Departamento de Informática, Universidade Federal do Espírito Santo.

[23] Programmer, L. (2016). *Deep Learning: Recurrent Neural Networks in Python, LSTM, GRU, and more RNN machine learning architectures in Python and Theano*, Kindle, United States of America.

[24] Muhuri, P.S.; Chatterjee, P.X.; Roy, K.; Esterlin, A. (2020). *Using a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) to Classify Network Attacks*. Department of Computer Science, North Carolina A & T State University, Greensboro, NC 27411, USA.

[25] Glorot, X.; Bordes, A.; Bengio, Y. (2011). *Deep sparse rectifier neurais networks*. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics.

[26] Lal, U.; Chikkankod, A.V.; Longo, L. (2024). Fractal dimensions and machine learning for detection of Parkinson's disease in resting-state electroencephalography. *Neural Computing and Applications (2024) 36:8257–8280*.

[27] Bianchi, F.M.; Maiorino, E.; Kampffmeyer, M.C.; Rizzi, J.; Antonello, R. (2017). *Recurrent Neural Networks for Short-Term Load Forecasting*. An Overview and Comparative Analysis, Springer Briefs in Computer Science.

[28] da Silva, R.N.; Fantini, D.G.; Mendes, R.C.F.; Junior, A.C.P.B. (2022). *Previsão de velocidade do vento com série temporal multivariada utilizando rede neural recorrente. IX Congresso Brasileiro de Energia Solar – Florianópolis.* Universidade de Brasília, Campus Darcy Ribeiro, Brasília - DF

[29] Venkateswaran, B.; Ciabuschi, F. (2017). *Smart models using CNN, RNN, deep learning, and artificial intelligence principles*. Neurais networks with R. Packt, Birmingham, UK Mumbai.

[30] Das, B.K.; Jiang, J.; Rao, J.N.K. (2004). *Mean squared error of empirical predicto*r. Calcutta University, University of California, Davis and Carleton University. arXiv: math/0406455v1 [math.ST] 23 Jun 2004

[31] Camargo, J.S. (2021). *Método bagging para aprimoramento de previsões de séries temporais***.** Dissertação de Mestrado do Programa Interinstitucional de Pós-Graduação em

Estatística. Instituto de Ciências Matemáticas e de Computação da Universidade Federal de São Carlos (Brasil).

[32] Gomes, G.S.S., Ludermir, T.B.&Lima, L.M.M.R.(2011). *Comparison of new activation functions in neural network for forecasting financial time series*. Federal University of Pernambuco Centre of Informatics Cidade Universitria Recife/PE, 50732-970, Brasil.