

MESTRADO
CIÊNCIAS EMPRESARIAIS

TRABALHO FINAL DE MESTRADO
DISSERTAÇÃO

**UTILIZAÇÃO DE MODELOS ESTATÍSTICOS E MACHINE LEARNING
PARA A PREVISÃO DE VENDAS NO SECTOR DO RETALHO – UM
ESTUDO COMPARATIVO**

CHRISTIAN RAPOSO LOPES

SETEMBRO - 2019

MESTRADO EM
CIÊNCIAS EMPRESARIAIS

TRABALHO FINAL DE MESTRADO
DISSERTAÇÃO

**UTILIZAÇÃO DE MODELOS ESTATÍSTICOS E MACHINE LEARNING
PARA A PREVISÃO DE VENDAS NO SECTOR DO RETALHO – UM
ESTUDO COMPARATIVO**

CHRISTIAN RAPOSO LOPES

ORIENTAÇÃO:

PROFESSOR DOUTOR JORGE CAIADO

SETEMBRO - 2019

Agradecimentos

Agradeço a todos aqueles que contribuíram, de forma direta ou indireta, para o finalizar desta fase académica. Em especial ao Filipe pelo facultar dos dados utilizados neste estudo.

Um agradecimento especial para a Raquel pelo suporte e apoio prestados ao longo destes dois anos difíceis.

Resumo

A previsão das necessidades futuras torna-se fundamental para o correto planeamento e operação das empresas de retalho. Ao nível organizacional, as previsões de vendas funcionam como parâmetros de entrada para as diversas áreas funcionais das empresas. Previsões pouco precisas ou ineficientes podem levar à realização de elevado ou baixo nível de inventário, afetando de forma direta a rentabilidade e a posição competitiva das organizações. Os dados de vendas provenientes deste sector, exibem fortes padrões de tendência e variação sazonal, colocando desafios ao desenvolvimento de modelos de previsão eficazes. O objetivo do trabalho assenta na comparação do desempenho da previsão entre os métodos ARIMA e MLP para o estudo de caso de uma empresa portuguesa do sector do retalho, tendo em consideração cinco grupos de produtos. Pretende-se também investigar qual o impacto do pré-processamento dos dados de entrada no rigor das previsões, e de que forma a utilização combinada dos métodos ARIMA e MLP aumenta a eficácia desta previsão. Os resultados apontam para um melhor desempenho do método MLP em comparação com o método ARIMA, quando utilizado o erro de previsão EAM. Foi ainda demonstrado que o pré-processamento dos dados impacta de forma positiva o desempenho da previsão, contribuindo para a redução significativa dos erros de previsão. Também se conclui que a previsão combinada das metodologias ARIMA e MLP produz resultados positivos para duas das cinco categorias de produtos analisadas, mostrando ser um conceito com elevado interesse para uma investigação futura.

Palavras-Chave

ARIMA; MLP; Séries Temporais; Tendência; Sazonalidade; Previsão de Vendas; Erros de Previsão.

Abstract

Forecasting the future needs is one of the most important decisions for the effective planning and operation of the retail industries. At the organizational level, sales forecasts act as input parameters for the several functional areas of the business. Inaccurate or inefficient forecasts can lead to high or low stock levels, directly affecting the organizations profit and competitive position. Retail sales data often exhibit a strong trend pattern and seasonal variation, creating challenges for the development of effective forecast models. The objective of this study is to compare the forecast performance of ARIMA and MLP methods through a case study of a Portuguese retail company, considering five different groups of products. It is also intended to investigate the impact of the data preprocessing on the accuracy of forecasts, and if the combination of ARIMA and MLP methods can increase the effectiveness of this forecast. The results show that the forecasting performance, evaluated via EAM error, of the MLP method is better than the one obtained for ARIMA method. It has also been shown that data preprocessing positively impacts the forecasting performance, contributing to a significantly reduction of the forecasting errors. It is also concluded that the combined forecasting of the ARIMA and MLP methodologies yields positive results for two of the five groups of products, proving to be a concept of high interest for future research.

Keywords

ARIMA; MLP, Time Series; Trend; Seasonality; Forecasting; Forecast Error Measures.

Índice

LISTA DE TABELAS	2
LISTA DE FIGURAS	2
LISTA DE ABREVIATURAS	2
1. INTRODUÇÃO	3
1.1. Guia de leitura	4
2. REVISÃO DE LITERATURA	6
2.1. Modelação de séries temporais sazonais com padrões de tendência	7
2.2. Modelo sazonal ARIMA	12
2.3. Modelo MLP	13
2.4. Questões de investigação	15
3. DADOS E METODOLOGIA	16
3.1. Dados do estudo	16
3.2. Erros de previsão	18
3.3. Pré-processamento dos dados	20
3.4. Análise dos dados, modelação e previsão (<i>software</i> utilizado)	21
3.5. Metodologia ARIMA	21
3.5.1 Implementação da metodologia ARIMA	23
3.6. Metodologia MLP	26
3.6.1 Arquitetura da rede	26
3.7. Combinação de previsões dos métodos ARIMA e MLP	28
4. ANÁLISE DE RESULTADOS	30
5. CONCLUSÕES, LIMITAÇÕES E INVESTIGAÇÃO FUTURA	35
REFERÊNCIAS BIBLIOGRÁFICAS	37
ANEXOS	42

Lista de Tabelas

Tabela I – Comparação do desempenho da previsão dos métodos ARIMA, MLP e combinação ARIMA e MLP (janeiro 2018 a dezembro 2018).	32
Tabela A1 – Modelos ARIMA utilizados na previsão de vendas. Teste ADF aos dados de entrada e teste de Ljung-Box aos resíduos dos modelos selecionados.	42
Tabela A2 – Desempenho da previsão do método ARIMA para o período de janeiro 2018 a dezembro 2018.....	43
Tabela A3 – Desempenho da previsão do método MLP para o período de janeiro 2018 a dezembro 2018 (considerando a arquitetura de rede com menor valor de erro EAM).	44
Tabela A4 – Previsão de vendas para as diversas metodologias em análise entre janeiro 2018 e dezembro 2018 (valores utilizados na Figura 8).	46

Lista de Figuras

Figura 1 – Modelo computacional de um neurónio artificial, adaptado de Haykin (2009).	14
Figura 2 – Rede neuronal MLP com uma camada escondida.	15
Figura 3 – Vendas mensais dos cinco grupos de produtos entre janeiro de 2014 e dezembro de 2018: (a) Jornais e Revistas, (b) Guloseimas, (c) Serviços, (d) Tabaco e (e) Papelaria.....	17
Figura 4 – Representação do modelo multiplicativo sazonal ARIMA.	21
Figura 5 – Gráficos FAC (à esquerda) e FACP (à direita) para os dados originais da série temporal Papelaria.	25
Figura 6 – Gráficos FAC (à esquerda) e FACP (à direita) dos resíduos dos dados originais da série temporal Papelaria.	25
Figura 7 – Propagação de sinal num modelo MLP, adaptado de Haykin (2009).	26
Figura 8 – Previsão de vendas mensais para as diferentes séries (entre janeiro 2018 a dezembro 2018): (a) Jornais e Revistas, (b) Guloseimas, (c) Serviços, (d) Tabaco e (e) Papelaria.	34

Lista de Abreviaturas

ADF	<i>Augmented Dickey-Fuller</i>
AIC	<i>Akaike's Information Criterion</i>
AICc	<i>Akaike's Information Criterion corrected</i>
ARIMA	<i>Autoregressive Integrated Moving Average</i>
BIC	<i>Bayesian Information Criterion</i>
EAM	Erro Absoluto Médio
EPAM	Erro Percentual Absoluto Médio
EQM	Erro Quadrático Médio
FAC	Função de AutoCorrelação
FACP	Função de AutoCorrelação Parcial
MLP	<i>Multi-Layer Perceptron</i>

1. Introdução

A previsão de vendas é um dos assuntos muito importante para efeitos de tomada de decisão no setor empresarial. Em particular, é fundamental para as decisões e planeamento estratégico das empresas de retalho. A importância de previsões precisas para o controlo eficiente dos inventários tem sido reconhecida há muitos anos. Previsões fracas ou ineficientes podem levar à realização de elevado ou baixo nível de inventário, afetando de forma direta a rentabilidade e a posição competitiva das organizações (Ramos *et al.*, 2015). A falta de *stock* causa perdas potenciais nas vendas e leva à insatisfação dos clientes. Estudos recentes demonstram que os consumidores, quando confrontados com uma situação de falta de *stock*, preferem não realizar a compra de um produto substituto, mas antes trocar de loja (Huang *et al.*, 2014). A modelação ou a criação de métodos de previsão das necessidades dos consumidores surge então como possível solução para o aumento das vendas, redução dos custos e aumento do lucro das empresas.

Tradicionalmente, a previsão de vendas tem por base a utilização de métodos estatísticos. Estes métodos baseiam-se no histórico de dados passados para a previsão de dados futuros. A sua formulação baseia-se em técnicas simples e de fácil aplicação, existindo modelos fundamentados em técnicas de média móvel, alisamento exponencial e métodos autorregressivos de média móvel, sendo este último considerado um dos mais robustos e eficazes. Apesar de serem ferramentas muito poderosas, estas assentam em pressupostos lineares no tipo de dados e na sua distribuição, podendo apenas tratar de um número limitado de variáveis (Cui *et al.*, 2006).

Métodos inovadores baseados em *Machine Learning* têm sido propostos na literatura académica como alternativa aos tradicionais. Estes métodos são apoiados em programação dinâmica e inteligência artificial, permitindo aos investigadores explorar grandes quantidades de dados na procura por padrões de consistência. De acordo com Makridakis *et al.* (2018), a inteligência artificial, em especial as redes neuronais artificiais, surgem como uma ferramenta versátil para a construção de modelos de previsão, estando entre os métodos mais inovadores para modelar a previsão do consumo.

A motivação e justificação do estudo passou pela necessidade de entender que métodos poderão ser mais eficazes na previsão de vendas das empresas de retalho em Portugal. Aqui, diversos fatores encontrados nos dados passados como a sazonalidade, tendência, etc., podem afetar a previsão dos dados futuros. Para isso, um pré-processamento dos dados pode ser a solução indicada para o aumento da eficácia destes métodos.

O propósito do estudo assenta na comparação do desempenho entre os métodos estatísticos (lineares) e os métodos de *Machine Learning* (não lineares), na previsão de vendas. Para isso, foi realizado um estudo de caso de uma empresa portuguesa do sector do retalho, onde foram analisados diferentes grupos de produtos. Para este estudo, foram utilizados dois métodos de análise: o método estatístico ARIMA (*Autoregressive Integrated Moving Average*), sendo conhecido como um dos métodos mais robustos pela literatura; e o método de *Machine Learning* MLP (*Multi-Layer Perceptron*), sendo também amplamente proposto na literatura, tendo obtido maior eficácia, entre os diversos métodos de *Machine Learning* no estudo realizado por Makridakis *et al.* (2018).

Os resultados do estudo indicam que o método MLP é o mais eficaz para a previsão de vendas de séries temporais com tendência e padrões de sazonalidade. Este demonstrou ser o mais preciso para três dos cinco grupos de produtos em estudo. Os resultados mostraram ainda o desempenho positivo apresentado pela utilização combinada do método ARIMA e MLP, tendo obtido os melhores resultados para os restantes grupos de produtos. De salientar que o pré-processamento dos dados é, na grande maioria dos casos, uma opção eficaz quando se trata de problemas de previsão.

1.1. Guia de leitura

A presente dissertação divide-se em cinco capítulos, sendo o primeiro este capítulo introdutório.

Capítulo 1

Procede-se ao enquadramento do presente trabalho, realizando uma introdução ao tema escolhido, caracterizando o problema da previsão de vendas no setor do retalho e apresentando os principais resultados obtidos.

Capítulo 2

Neste capítulo, apresenta-se uma revisão bibliográfica sobre os assuntos mais abordados, realizando-se um levantamento dos métodos mais utilizados e a sua eficácia na previsão de séries temporais. São também apresentadas as questões de investigação que se deseja dar resposta com a realização do trabalho.

Capítulo 3

Detalham-se os dados do estudo, apresentam-se os erros de previsão, descreve-se a metodologia utilizada e os passos seguidos na investigação.

Capítulo 4

Procede-se à análise e comparação dos resultados obtidos a partir do *software* de programação estatística R. Expõem-se os resultados com base nas diferentes metodologias em análise.

Capítulo 5

Neste capítulo final descrevem-se as principais conclusões e limitações do trabalho que aqui se apresenta. Indicam-se ainda algumas perspetivas de desenvolvimento de trabalho futuro.

2. Revisão de Literatura

A previsão das necessidades futuras é fundamental para o planeamento e operação das empresas de retalho, tanto ao nível macro como ao nível micro. Ao nível organizacional, as previsões de vendas funcionam como parâmetros de entrada para as diversas áreas funcionais, a saber: vendas, produção, compras, marketing, finanças e contabilidade (Chu e Zhang, 2003). Previsões de vendas precisas têm o potencial de aumentar a rentabilidade dos retalhistas, ao melhorar a eficiência da cadeia de operações, minimizando desperdícios. Thall (1992), e Agrawal e Schorling (1996) nos seus estudos, apontaram para o papel importante que uma previsão eficaz tem no lucro das operações de retalho. Realçaram ainda o facto de previsões fracas resultarem em muito ou pouco nível de *stock*, afetando diretamente a posição competitiva das empresas de retalho. Para Chu e Zhang (2003), previsões de venda eficazes têm grande importância na gestão eficiente do inventário, tanto ao nível desagregado como agregado. Neste sentido, Barksdale e Hilliard (1975) examinaram a relação entre o *stock* e as vendas ao nível agregado, tendo concluído que o sucesso da gestão do inventário depende em larga escala da previsão eficaz das vendas. Ramos *et al.* (2015) acrescentam que séries temporais com dados de venda agregados são preferenciais, dado conterem padrões de sazonalidade e tendência, fornecendo uma boa base de testes para a comparação dos métodos de previsão.

As séries temporais das vendas de retalho exibem uma forte variação sazonal e elevados padrões de tendência, apresentando desafios ao desenvolvimento de modelos de previsão eficazes. Segundo Zhang e Qi (2005), a sazonalidade é definida como um padrão periódico e recorrente causado por fatores como o tempo, as férias, as promoções, assim como o comportamento dos agentes económicos. A tendência encontra-se, como o próprio nome indica, relacionada com a tendência de comportamento de uma variável em relação ao tempo. Apesar das variações sazonais serem a componente mais significativa de uma série temporal sazonal, estas são acompanhadas de tendências de natureza estocástica, que podem ter um impacto importante em diversos métodos de previsão. De acordo com Zhang e Qi (2005), uma série temporal com um padrão de tendência é considerada não estacionária e, na maioria das

situações, necessita de ser tornada estacionária antes do processo de previsão se iniciar. Conforme descrito anteriormente, a previsão precisa de uma série temporal, que exiba sazonalidade e padrões de tendência, é muito importante para a tomada de decisões consistentes no setor do retalho. Segundo Ramos *et al.* (2015), a modelação efetiva de uma série temporal com variações sazonais é um dos maiores desafios que se levanta no seio académico, tendo sido alvo de enorme desenvolvimento nas últimas décadas.

Abordagens tradicionais na modelação de séries temporais sazonais são baseadas em métodos heurísticos como a decomposição clássica, o alisamento exponencial, regressões de séries temporais e modelos autorregressivos de médias móveis, também designados de ARIMA (Chu e Zhang, 2003). Uma das grandes limitações destes modelos é o facto de serem essencialmente lineares, de natureza estacionária. Nos últimos anos, a rápida acumulação de dados dos consumidores e das suas transações, tem resultado em bases de dados de elevada dimensão, criando assim uma oportunidade única aos investigadores de explorarem a utilização de métodos de mineração de dados (vulgarmente conhecidos como *data-mining*), de forma a adquirir conhecimento sobre o comportamento dos consumidores (Cui *et al.*, 2006).

Métodos inovadores como o *Machine Learning*, permitiram aos investigadores realizarem mineração de dados em bases de dados volumosas, fornecendo assim decisões de suporte aos gestores. De acordo com Chu e Zhang (2003), a forma eficaz de modelar uma série temporal sazonal é uma tarefa desafiante não só para os modelos não lineares recentemente desenvolvidos, mas também para os modelos designados tradicionais.

2.1. Modelação de séries temporais sazonais com padrões de tendência

A modelação eficiente de séries temporais sazonais com padrões de tendência tem sido um dos maiores esforços de investigação há décadas. As primeiras pesquisas sobre este tema surgiram na década 20 do século XX, estando a investigação focada em modelos de decomposição e no ajustamento sazonal. Desde essa altura, diferentes modelos de ajustamento têm sido propostos, sendo o mais significativo e o um dos mais populares o designado método X-11, desenvolvido por Bureau e Census na década de 50 e 60 do

mesmo século (Zhang e Qi, 2005). Estas abordagens consideram que as variações sazonais existentes nas séries temporais devem ser removidas, utilizando, para isso, métodos de ajustamento sazonal. O método de decomposição clássica, por exemplo, decompõe a série temporal em sazonalidade, tendência, ciclo e componentes aleatórias. A influência da sazonalidade é então estimada e removida dos dados antes de se estimar as restantes componentes.

Devido à natureza *ad hoc* dos métodos de ajustamento sazonal, alguns modelos ou procedimentos base foram sendo desenvolvidos. Entre eles, encontra-se o trabalho realizado por Box e Jenkins (1976) no modelo sazonal ARIMA, que viria a ter um impacto elevado na aplicação prática da modelação de séries temporais sazonais. Contudo, este requer que os dados sejam diferenciados sazonalmente, por forma a ser atingida uma condição estacionária.

Muitos utilizadores de modelos de previsão em diferentes aplicações têm, de forma satisfatória, adotado a prática do ajustamento sazonal. Conforme descrito por Zhang e Qi (2005), estes acreditam que as flutuações sazonais podem dominar as restantes variações de uma série temporal, causando dificuldade na medição eficaz de outras componentes da série. Segundo Gardner e MacKenzie (1989), os resultados da conhecida competição M, apresentados pelos investigadores Makridakis *et al.* (1982), sugerem que a remoção da sazonalidade dos dados é uma aproximação eficaz na modelação de séries temporais sazonais. Também Miller e Williams (2003) demonstraram que a eficácia da previsão é superior quando se diminui a componente sazonal em torno de zero. Contudo, o ajustamento sazonal dos dados é um tema envolto em controvérsia. Ghysels *et al.* (1996) sugerem que o ajustamento sazonal pode gerar propriedades não lineares indesejáveis em séries temporais com uma única variável. Também Ittig (1997), questionou o método tradicional por gerar índices de sazonalidade, tendo proposto um método não linear para estimar os fatores sazonais. De Gooijer e Franses (1997) afirmaram que apesar de o ajustamento sazonal dos dados ser por vezes útil, é tipicamente recomendado usar os dados sazonais sem este tipo de ajustamento. Zhang e Qi (2005), referiram que as flutuações sazonais não são sempre constantes no tempo e,

em algumas séries temporais, as componentes sazonais e não sazonais não são independentes, portanto, não devem ser separadas.

Outro aspeto em que também não é claro o tipo de modelação a adotar é a designada tendência ou padrão de tendência de uma série temporal. No método popular de Box e Jenkins (1976), a diferenciação é usada para atingir um estado estacionário em torno da média. Contudo, alguns autores como Pierce (1977) e Nelson e Plosser (1982), argumentam que a diferenciação não é sempre o caminho mais adequado para tratar de um padrão de tendência, sugerindo que uma remoção linear da tendência poderá ser mais apropriada.

Os métodos de modelação de séries temporais sazonais com tendência acima mencionados, são paramétricos por natureza. A relação essencialmente linear destes modelos, limita a sua capacidade para modelar problemas de complexidade não linear, encontrados no contexto real. Esta também poderá ser a causa para os resultados mistos existentes na literatura em relação aos méritos do ajustamento sazonal.

Pesquisas recentes sobre o uso de inteligência artificial para as mais variadas aplicações de previsão, têm demonstrado que este método pode ser um candidato importante na previsão de séries temporais sazonais com tendências. Segundo Makridakis *et al.* (2018), o sucesso da inteligência artificial assenta na utilização de algoritmos capazes de aprender por tentativa erro e melhorar o seu desempenho ao longo do tempo. A literatura demonstra que existe uma elevada quantidade de pesquisa na utilização de *Machine Learning*, para a previsão de séries temporais. Conforme descrito por Cui *et al.* (2006), os métodos de *Machine Learning* têm sido adotados como ferramentas eficazes na mineração de dados e na descoberta de padrões não óbvios ou até mesmo escondidos nos dados em análise. Estes métodos incluem regras de associação, árvores de decisão, redes neuronais e algoritmos genéticos. Um modelo que tem recebido uma extensa atenção na área da previsão é o modelo de rede neuronal. Este modelo é inspirado na arquitetura do cérebro humano e na forma como este processa a informação. De acordo com Pillo (2013), as redes neuronais apresentam uma estrutura mais flexível do que os modelos estatísticos, baseando as suas previsões não só nos dados de saída, mas também nos dados de entrada no qual os resultados de saída podem depender.

Conforme descrito por Makridakis *et al.* (2018), a primeira aplicação das redes neuronais para fins de previsão data de 1964, não tendo, contudo, sido muito desenvolvida até à introdução da técnica de retro propagação 20 anos mais tarde. Desde essa altura, têm existido diversos estudos utilizando as redes neuronais, muitos dos quais comparam a sua precisão com a dos métodos estatísticos. A popularidade do modelo de rede neuronal pode ser atribuída à sua capacidade única de simular uma larga variedade de comportamentos não lineares. Nesta linha de pensamento, os autores Hornik *et al.* (1989) afirmaram que, com a arquitetura apropriada, as redes neuronais podem aproximar qualquer tipo de função com a precisão desejada. Zhang (1998) complementa este raciocínio ao dizer que, apesar de serem intrínsecas aos modelos não lineares, as redes neuronais são também capazes de modelar processos lineares.

Devido às funções de aproximação universal das redes neuronais, é natural que estas sejam utilizadas de forma direta na modelação de variações sazonais e padrões de tendência. Gorr (1994), evidencia que as redes neuronais são capazes de detetar a sazonalidade e a tendência de uma série temporal sazonal. Sharda e Patil (1992), examinaram 88 séries temporais sazonais, provenientes da competição M (Makridakis *et al.*, 1982), tendo descoberto que as redes neuronais podem modelar, de forma eficiente, as variações sazonais, sem que para isso seja necessário um pré-processamento dos dados para remoção da sazonalidade. Franses e Draisma (1997), reforçaram ainda que as redes neuronais são capazes de detetar possíveis mudanças de padrões sazonais. Alon *et al.* (2001) concluíram que as redes neuronais são capazes de capturar a dinâmica não linear dos padrões de sazonalidade e tendência, bem como a interação existente entre eles. Outros estudos citados por Zhang e Qi (2005), reportaram também resultados encorajadores no que respeita à utilização de redes neuronais para a previsão de séries temporais sazonais. Contudo, Farway e Chatfield (1995) obtiveram resultados mistos utilizando diferentes estruturas de redes neuronais e diferentes horizontes de previsão. No entanto, estes apenas utilizaram uma série temporal para a realização de previsões futuras. Baseado num estudo de 68 séries temporais provenientes da competição M (Makridakis *et al.*, 1982), Nelson *et al.* (1999) concluíram que as

redes neuronais treinadas com dados com prévia remoção da sazonalidade, conseguem prever melhor que aquelas treinadas com dados não ajustados. Chu e Zhang (2003) e Zhang e Qi (2005) confirmaram as conclusões desse estudo.

Tendo em consideração as grandes capacidades e o elevado desempenho atribuído aos métodos de *Machine Learning*, Makridakis *et al.* (2018) alertaram para a limitada evidência objetiva em relação ao seu desempenho como uma ferramenta padrão de previsão. Estes mesmo autores, afirmaram que a sua superioridade é caracterizada por três grandes limitações. A primeira, é as suas conclusões serem baseadas em poucas, ou até mesmo séries temporais únicas, levantando a questão sobre o significado estatístico dos resultados e a sua generalização. A segunda, é que os métodos são utilizados para previsões de curto prazo, sem considerar o médio e longo prazo. A última, é que não são utilizados *benchmarks* para comparar a precisão dos métodos de *Machine Learning* com a dos métodos alternativos.

Apesar de existirem vários estudos que comparam o desempenho dos modelos tradicionais estatísticos com os modelos de redes neuronais, os resultados apresentados são de certa forma mistos. Ahmed *et al.* (2010), sumarizam um grande número de estudos realizados desde 1995, onde obtiveram esta mesma conclusão. Conclusão similar foi também apresentada por Adya e Collopy (1998), quando compararam 48 estudos de redes neuronais com os métodos estatísticos. Contudo, o que caracterizou todos esses estudos foi o número limitado de séries temporais utilizadas nos mesmos. O primeiro estudo em grande escala, utilizando 3003 séries temporais provenientes da competição M3, foi publicado em 2000 por Makridakis e Hibon (2000). Esse estudo incluiu um método de rede neuronal artificial autónoma que obteve resultados em média similares aos dos métodos estatísticos tradicionais. Zhang e Qi (2005) investigaram o uso de redes neuronais artificiais na previsão de séries temporais com fortes variações sazonais e padrões de tendência, tendo concluído que o desempenho geral da amostra de dados de saída não é superior ao apresentado pelo modelo ARIMA. Anos mais tarde, Crone *et al.* (2011) realizaram um estudo para uma competição especializada em redes neuronais, utilizando um subconjunto de dados mensais da competição M3. Neste estudo, eles compararam 22 métodos de redes neuronais com 11 métodos estatísticos.

Os resultados apontaram para o facto de nenhum método de *Machine Learning* ter superado o método estatístico *Theta*, o mais preciso da competição M3.

2.2. Modelo sazonal ARIMA

O modelo sazonal ARIMA pertence a uma família de modelos lineares e flexíveis, podendo ser utilizado na modelação de diferentes tipos de séries temporais sazonais e não sazonais. Segundo Chu e Zhang (2003), é considerado como o método linear mais versátil na previsão de séries temporais sazonais. Este método tem gozado de grande sucesso quer a nível académico quer ao nível industrial durante as últimas quatro décadas. Pode representar muitos tipos de séries temporais estocásticas como puramente autorregressivas (AR), puramente de média móvel (MA) ou uma junção de ambas. A teoria do modelo ARIMA, foi desenvolvida por diversos investigadores, sendo a sua aplicação global devida ao trabalho realizado por Box e Jenkins (1976), que desenvolveram um método prático e sistemático de construção de modelos. Através de um processo iterativo de construção em três etapas: identificação do modelo, estimativa de parâmetros e diagnóstico do modelo, a metodologia de *Box-Jenkins* tem comprovado ser uma aproximação eficiente na modelação de séries temporais.

O modelo multiplicativo sazonal ARIMA, denotado como ARIMA (p, d, q) $X(P, D, Q)_m$ apresenta a seguinte formulação matemática (Ramos *et al.*, 2015):

$$\phi_p(B)\Phi_P(B^m)(1-B)^d(1-B^m)^D y_t = C + \theta_q(B)\Theta_Q(B^m)\varepsilon_t \quad (1)$$

onde

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1 B - \dots - \phi_p B^p; & \Phi_P(B^m) &= 1 - \phi_1 B^m - \dots - \phi_p B^{pm} \\ \theta_q(B) &= 1 + \theta_1 B + \dots + \theta_q B^q; & \Theta_Q(B^m) &= 1 + \theta_1 B^m + \dots + \theta_q B^{qm} \end{aligned}$$

sendo m o período sazonal, B o operador atraso, d o grau de diferenciação simples e D o grau de diferenciação sazonal. Os polinómios $\phi_p(B)$ e $\theta_q(B)$ são os polinómios autorregressivo e de média móvel de ordens p e q , respetivamente. Os polinómios $\Phi_P(B^m)$ e $\Theta_Q(B^m)$ são os polinómios sazonais

autorregressivo e de média móvel das ordens P e Q, respetivamente. A constante C é representada como:

$$C = \mu(1 - \phi_1 - \dots - \phi_p)(1 - \Phi_1 - \dots - \Phi_p) \quad (2)$$

onde μ é a média do processo $(1 - B)^d(1 - B^m)^D y_t$ e ε_t o processo de ruído branco com variância σ^2 . As raízes dos polinómios $\phi_p(B)$, $\Phi_p(B^m)$, $\theta_q(B)$ e $\Theta_q(B^m)$ devem estar fora do círculo unitário de modo a assegurar as condições de estacionaridade e invertibilidade (Ramos *et al.*, 2015).

2.3. Modelo MLP

Segundo Ahmed *et al.* (2010), o modelo MLP é uma das mais populares arquiteturas de rede existentes, tanto na resolução de problemas de classificação como de regressão. O modelo MLP é o exemplo mais comum de uma rede neuronal de alimentação direta (do inglês *feedforward*). Este tipo de rede é constituído por várias camadas de neurónios ligados entre si por sinapses com pesos associados. Numa rede de alimentação direta, a informação move-se apenas num sentido, para a frente. Iniciando nos neurónios da camada de entrada, passando pelos neurónios das camadas escondidas, até aos neurónios da camada de saída, não existindo ciclos ou *loops* na rede. A função da camada de entrada é fornecer à rede informação sobre o mundo exterior. Nenhum cálculo é realizado pelos neurónios de entrada, estes apenas fornecem informação aos neurónios da camada escondida. Esta camada, por sua vez, é responsável pela realização de cálculos, tendo em consideração os dados da camada anterior, sendo ainda responsável pela transferência da informação para a camada seguinte. A camada de saída, recebe a informação proveniente da camada escondida anterior e realiza cálculos similares aos desta. O resultado final destes cálculos será considerado como o resultado de saída do modelo MLP.

Para entender o funcionamento deste modelo é necessário perceber a forma como cada neurónio recebe a informação, proveniente de outros neurónios ou fontes externas, realiza os cálculos e gera um resultado. Cada dado de entrada tem um peso associado (p), que é atribuído com base na sua importância relativamente a outros dados de entrada. Cada neurónio, aplica uma

função de transferência à soma ponderada dos dados de entrada, criando assim um resultado de saída. A Figura 1 apresenta uma constituição simplificada do modelo computacional de um neurónio artificial.

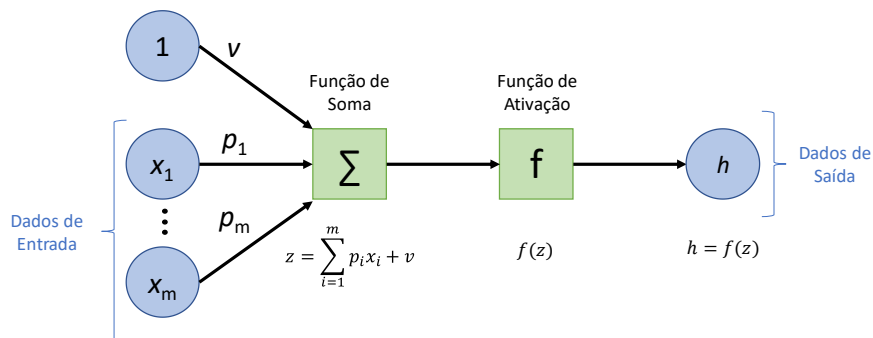


Figura 1 – Modelo computacional de um neurónio artificial, adaptado de Haykin (2009).

O modelo ilustrado na Figura 1 é constituído pelos dados numéricos de entrada x_1 até x_m e pelos seus pesos associados p_1 a p_m . Apresenta ainda como parâmetro de entrada a constante 1 (valor de ativação) com peso associado v , designado de viés. A função do viés é fornecer a cada nó um valor que permita maior flexibilidade de aprendizagem de padrões de comportamento. A função de soma traduz a soma ponderada dos dados de entrada. O valor da função de soma passa então por uma função de ativação, com o propósito de introduzir não linearidade ao resultado de saída do neurónio, limitando o valor a um certo intervalo. Este processo é importante, uma vez que a maioria dos dados do mundo real são não lineares, permitindo assim que os neurónios aprendam a partir dessas representações não lineares.

Os neurónios são então organizados em camadas de modo a compor uma rede. De acordo com Reed e Marks (1999), um modelo MLP padrão é uma cascata de camadas singulares, onde existe uma camada de entrada, uma camada de saída, e uma ou mais camadas escondidas. A Figura 2, apresenta um exemplo de uma rede neuronal multicamada, constituída por uma camada de entrada, uma camada escondida e uma camada de saída. A forma como os dados de entrada são processados e transferidos para a próxima camada é a mesma da apresentada na Figura 1.

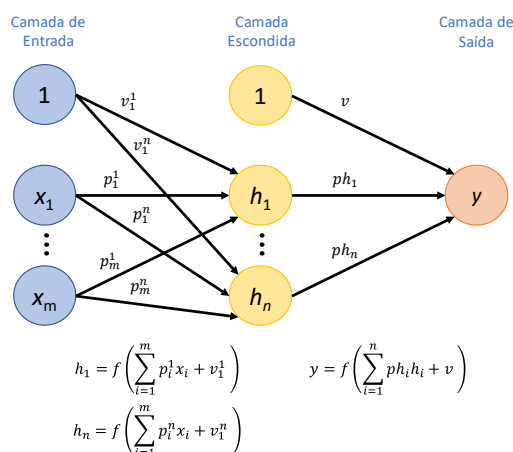


Figura 2 – Rede neuronal MLP com uma camada escondida.

2.4. Questões de investigação

Esta investigação teve como objetivo fornecer algumas evidências empíricas sobre a eficácia dos métodos de previsão ARIMA e MLP na modelação de séries temporais com padrões de tendência e sazonalidade, provenientes de uma empresa de retalho em Portugal. Tendo em consideração o objetivo identificado, foram colocadas as seguintes questões de investigação:

1. Qual o método de previsão, de entre os métodos ARIMA e MLP, que permite modelar de forma mais eficaz o comportamento de uma série temporal com padrões de tendência e sazonalidade?
2. O pré-processamento dos dados, via remoção da tendência e ajustamento sazonal, aumenta a eficácia dos métodos de previsão?
3. A utilização combinada das previsões dos métodos ARIMA e MLP traduzir-se-á numa solução mais precisa?

A primeira questão foi motivada pelo estudo de Makridakis *et al.* (2018), que evidenciaram a facilidade como se sobrevalorizam os métodos de redes neuronais sem a própria utilização e avaliação de métodos mais simples como os tradicionais. A segunda surge pelo facto de existir informação contraditória sobre a influência do pré-processamento dos dados no desempenho das previsões. A terceira foi definida com o sentido de explorar a combinação dos resultados de previsão dos métodos analisados, permitindo, desta forma, o retirar de conclusões para uma investigação futura mais detalhada.

3. Dados e Metodologia

O trabalho desenvolvido pretendeu dar resposta às questões anteriormente apresentadas. Para isso, foi realizado um estudo de caso que permitiu avaliar o desempenho dos modelos de previsão, tendo por base dados da área de retalho. A metodologia de estudo segue uma abordagem quantitativa, com a utilização de dados passados para a previsão de dados futuros.

Do ponto de vista do objetivo da investigação, o estudo assume o perfil de pesquisa exploratória, tratando de problemas pouco conhecidos, permitindo gerar proposições para futuras investigações.

3.1. Dados do estudo

A empresa de retalho aqui analisada trata-se de uma papelaria que se encontra aberta ao público desde 2006, localizando-se no centro da vila de Condeixa-a-Nova, distrito de Coimbra. É caracterizada pela oferta de uma variada gama de produtos com possibilidade de consumo durante o ano todo.

Neste estudo, foram analisadas as vendas mensais de cinco grupos de produtos, Jornais e Revistas; Guloseimas; Serviços; Papelaria e Tabaco, desde janeiro de 2014 a dezembro de 2018 (60 observações). A Figura 3 apresenta a evolução gráfica das vendas dos diversos produtos, considerando o período temporal em análise. Como se pode observar, a evolução das vendas é de natureza irregular, sendo visível um padrão de tendência decrescente para a totalidade dos produtos. Contudo, não se torna perceptível a existência de padrões de sazonalidade, com exceção dos artigos de Papelaria, onde é visível um crescimento das vendas no último mês de cada ano civil.

A série temporal Jornais e Revistas é caracterizada por uma tendência decrescente para os 3 primeiros períodos, recuperando no ano de 2017 e permanecendo de certa forma constante no último período. Quanto às séries Guloseimas e Tabaco, estas apresentam uma tendência decrescente durante os 5 períodos em análise. Os níveis de vendas registados no último período temporal são cerca de duas vezes inferiores aos registados para o primeiro período. Os Serviços seguem o padrão de tendência anterior, com exceção do último período, onde é possível verificar uma ligeira tendência crescente. Já a

série temporal Papelaria, permanece constante durante os 3 primeiros períodos, apresentando uma ligeira tendência decrescente para os restantes períodos, e ainda maior sensibilidade às flutuações sazonais.

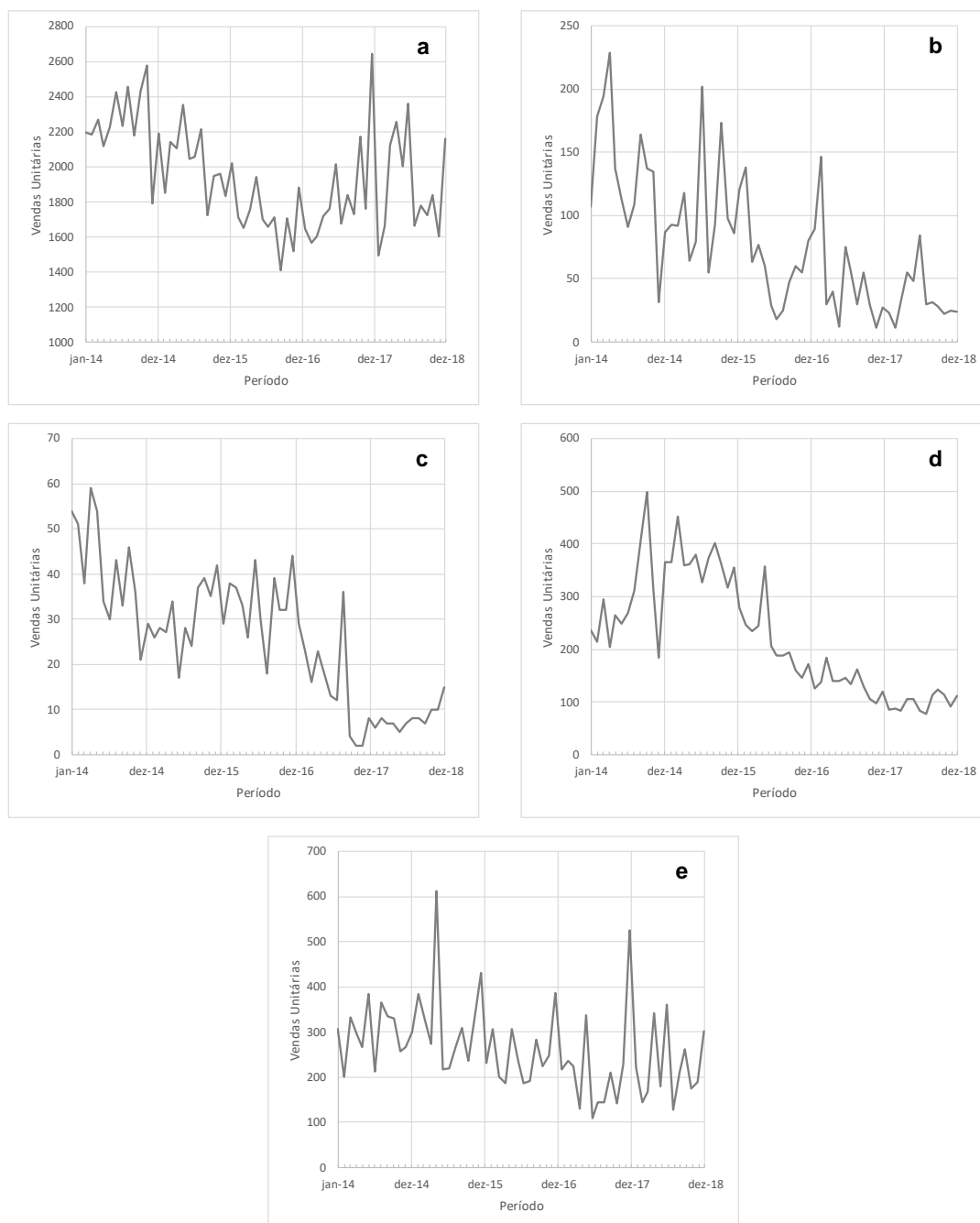


Figura 3 – Vendas mensais dos cinco grupos de produtos entre janeiro de 2014 e dezembro de 2018: (a) Jornais e Revistas, (b) Guloseimas, (c) Serviços, (d) Tabaco e (e) Papelaria.

Na comparação entre diferentes modelos, foi utilizada uma parcela dos dados para o ajuste dos modelos. Esta parcela designa-se de dados de Treino. Os restantes, são usados na avaliação dos modelos quanto à previsão de novos dados, sendo designados de dados de Teste. Para os cinco grupos em análise, o período de Treino foi definido como os primeiros quatro anos de vendas (48 observações). O último período (12 observações), foi considerado como o período de Teste. Todas as comparações entre os modelos, foram baseadas no período de Teste definido.

3.2. Erros de previsão

A análise e comparação do desempenho dos modelos de previsão foi realizada tendo em consideração a avaliação dos erros de previsão. Estes permitem entender a eficácia do modelo na previsão de um determinado horizonte temporal, fazendo, para isso, uma comparação entre os dados previstos e os dados observados. Neste estudo, apenas foram considerados os dados de Teste para a obtenção dos erros de previsão. O interesse da investigação residia na verificação do desempenho dos modelos na previsão de dados futuros e não na forma como estes se ajustaram aos dados de Treino.

Segundo Caiado (2016), e considerando Y_t como o valor observado no momento t , e P_t o valor previsto para o mesmo instante, os critérios dos erros de previsão mais utilizados baseiam-se nas seguintes funções:

- Erro Absoluto Médio (EAM)

Traduz o valor médio das diferenças absolutas entre os valores observados e os valores previstos para os instantes 1 até m :

$$EAM = \frac{1}{m} \sum_{t=1}^m |Y_t - P_t| \quad (3)$$

- Erro Quadrático Médio (EQM)

Traduz o valor médio das diferenças ao quadrado entre os valores observados e os valores previstos para os instantes 1 até m:

$$EQM = \frac{1}{m} \sum_{t=1}^m (Y_t - P_t)^2 \quad (4)$$

É comum a apresentação deste erro como a raiz quadrada do erro quadrático médio (REQM) por forma a reduzir a grandeza dos valores:

$$REQM = \sqrt{EQM} \quad (5)$$

- Erro Percentual Absoluto Médio (EPAM)

Traduz o valor percentual absoluto médio da diferença entre os valores observados e os valores previstos para os instantes 1 até m:

$$EPAM = \frac{1}{m} \sum_{t=1}^m \left| \frac{Y_t - P_t}{Y_t} \right| \times 100 \quad (6)$$

Os erros de previsão apresentados têm sido amplamente utilizados na avaliação do desempenho dos modelos de previsão ao longo dos últimos anos. Contudo, não existe um consenso na literatura sobre a métrica de medição mais apropriada.

Na comparação do desempenho de modelos de previsão, o erro EAM apresenta-se como um dos mais interessantes, uma vez que traduz a diferença absoluta dos erros, sendo por isso de fácil interpretação. Por sua vez, o REQM tem o inconveniente de ser muito sensível aos erros aberrantes, uma vez que estes são elevados ao quadrado. Não existe por isso um critério absoluto para a escolha entre o EAM e o REQM. Esta depende das unidades na qual a variável é medida e no grau de precisão da previsão, que depende da aplicação particular. A avaliação da percentagem de erro tem a vantagem de ser independente da escala, sendo frequentemente utilizada na comparação do desempenho de modelos de previsão entre séries temporais distintas. A desvantagem desta avaliação assenta no facto do erro se tornar infinito ou não definido quando os valores observados se encontram perto de zero.

Neste estudo, a seleção e comparação dos modelos mais precisos teve por base o erro de previsão EAM. Foi ainda considerado o horizonte temporal de 12 meses, correspondendo a 12 instantes de previsão. Para permitir uma compreensão mais alargada sobre a medição da previsão, os erros REQM e EPAM são também apresentados.

3.3. Pré-processamento dos dados

O pré-processamento de uma série temporal poderá ter um grande impacto no subsequente desempenho de uma previsão. Segundo Makridakis *et al.* (2018), o pré-processamento dos dados permite que as séries temporais, utilizadas como dados de input para os modelos previsão, atinjam uma condição estacionária tanto na média como na variância. Esta condição apresenta-se fundamental para que os modelos tradicionais, como por exemplo o modelo ARIMA, sejam capazes de realizar previsões mais eficazes. Modelos mais sofisticados, como o caso dos modelos de redes neuronais, conseguem processar e analisar séries temporais não estacionárias.

Conforme descrito por Makridakis *et al.* (2018), o pré-processamento dos dados pode ser realizado a partir de três formas: transformação logarítmica ou de potência, ajuste sazonal e remoção da tendência. São apresentadas abaixo as diversas alternativas de pré-processamento utilizadas neste estudo:

- Dados originais: Nenhum pré-processamento aplicado.
- Transformação dos dados: Transformação logarítmica e transformação de *Box-Cox* (Box e Cox, 1964) aplicada aos dados originais, por forma a ser possível atingir uma condição estacionária na variância.
- Remoção da sazonalidade: Utilização do modelo de decomposição aditiva de uma série temporal para remoção da componente sazonal da série. Processamento não usado para o modelo ARIMA, uma vez que este é formado por termos sazonais capazes de processar a informação sazonal contida nas séries temporais.
- Remoção da tendência: Transformação dos dados originais através do uso de diferenciação, por forma a atingir uma condição estacionária na média.

- **Combinações:** São realizadas combinações entre as várias técnicas de pré-processamento, procurando utilizar o benefício individual no ajustamento combinado dos dados originais.

Neste estudo, foram testadas as capacidades de previsão dos modelos ARIMA e MLP, tendo em consideração os dados originais e os dados pré-processados, utilizando de forma individual ou combinada os diferentes tipos de tratamento dos dados.

3.4. Análise dos dados, modelação e previsão (*software* utilizado)

A análise dos dados, modelação e previsão das diversas séries temporais foi realizada com recurso ao *software* de programação estatística R (<https://www.r-project.org/>). Este *software* é composto por um conjunto de pacotes que agregam diversas funções de cálculo estatístico. Neste estudo, foram utilizadas as funções integrantes dos pacotes *forecast*, *tseries*, *MASS*, *nnfor* e *neuralnet*. A capacidade do *software* R para a previsão de séries temporais é documentada por Hyndman e Athanasopoulos (2018), Trapletti e Hornik (2019) e Ripley (2019).

3.5. Metodologia ARIMA

A principal tarefa na previsão de dados utilizando o modelo ARIMA, consiste na identificação das ordens de diferenciação (d, D) e das ordens das partes autorregressiva (p, P) e de média móvel (q, Q), ver Figura 4.

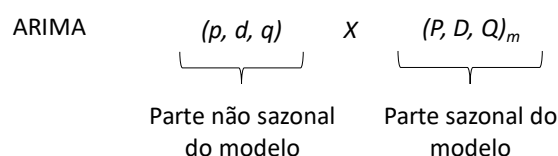


Figura 4 – Representação do modelo multiplicativo sazonal ARIMA.

A construção do modelo de previsão apropriado foi feita de forma iterativa, iniciando com a identificação do estado das séries, seguindo-se a identificação do tipo de modelo (sazonal ou não sazonal), o estimar das ordens apropriadas e

a análise diagnóstica da informação obtida. São apresentados abaixo, de forma detalhada, os passos seguidos na identificação do modelo a ser utilizado:

1. Identificação do estado da série:

A representação gráfica da evolução das séries temporais, permitiu a observação de padrões de tendência, variância não constante e irregularidades nos dados. De modo complementar, foi ainda investigado o estado estacionário da série utilizando-se o teste ADF. A hipótese nula deste teste assume que a série tem uma raiz unitária. Convencionou-se que não se rejeita esta hipótese se o valor de probabilidade (*p-value*), associado à componente estatística do teste, for superior a 0.05.

2. Identificação do tipo de modelo (sazonal ou não sazonal):

A identificação do comportamento sazonal das séries foi realizada com recurso à função *auto.arima*, proveniente do pacote *forecast*. Esta permite uma identificação rápida do modelo que melhor se ajusta aos dados, tendo como critério de seleção a combinação de ordens que geram o menor valor de AIC, AICc e BIC. Segundo Ramos *et al.* (2015), o modelo com menor valor de AIC e AICc é muitas vezes o melhor modelo para previsão.

Para o caso onde não foi verificado um padrão de sazonalidade, o modelo foi descrito na forma (p, d, q) . Para o caso sazonal, o modelo foi descrito na forma $(p, d, q) \times (P, D, Q)_m$.

3. Seleção das ordens (p, d, q) e $(p, d, q) \times (P, D, Q)_m$:

A seleção inicial das ordens p e q , teve em consideração a análise da Função de Auto Correlação Parcial (FACP) e da Função de Auto Correlação (FAC), respetivamente. O gráfico FACP ilustra a auto correlação parcial dos valores desfasados da série, enquanto o gráfico FAC ilustra a auto correlação dos valores desfasados da séries. O grau de diferenciação d foi obtido a partir da função *ndiffs* existente no pacote *forecast*.

Dado não ser de fácil interpretação visual, a definição inicial das ordens da parte sazonal $(P, D, Q)_m$ foi realizada com recurso à função *auto.arima*. Esta definição funcionou como ponto de partida para a avaliação dos modelos, tendo estes sido ajustados conforme a análise dos resíduos explicada no passo (4).

4. Análise dos resíduos:

Após a identificação do modelo inicial, torna-se necessário verificar se os pressupostos estabelecidos foram satisfatórios. A avaliação desta condição é realizada através da análise dos resíduos. Se o modelo não se encontrar devidamente discriminado, este terá reflexo nos resíduos sob a forma de tendências e outros padrões não capturados pelo modelo. Idealmente, os resíduos devem assumir a forma de um ruído branco, significado que não se encontram auto correlacionados. Esta análise, foi executada considerando os gráficos FAC e FACP dos resíduos de cada modelo, por forma a identificar a existência de padrões com elevado valor de probabilidade estatística. Foi considerado que os resíduos não exibem auto correlação analisando os intervalos de aceitação do nível representado pelas linhas a azul nos gráficos FAC e FACP.

De forma complementar, foi considerado ainda o teste de *Ljung-Box* referenciado por Ramos *et al.* (2015). Este permite testar se as primeiras k auto correlações dos resíduos são diferentes ao que seria esperado num processo de ruído branco. A hipótese nula assume que as primeiras k auto correlações são nulas. Convencionou-se que não se rejeita esta hipótese se o valor de probabilidade (*p-value*), associado à componente estatística do teste, for superior a 0.05.

Se existirem picos nos gráficos FAC e FACP dos resíduos, fora dos limites ($\pm 2/\sqrt{n}$) estabelecidos pelas linhas a azul, ou se o modelo “falhar” o teste de *Ljung-Box*, outro modelo deve ser testado e uma nova análise dos resíduos realizada. O processo iterativo finaliza quando os pressupostos forem satisfatórios.

3.5.1 Implementação da metodologia ARIMA

A seleção do modelo de previsão apropriado teve início na identificação do estado da série. Da análise conjunta entre os gráficos de evolução temporal das séries e o valor de probabilidade obtido pelo teste ADF (ver Tabela A1), foi possível constatar que as séries são de natureza não estacionária. A investigação sugere que todas as séries sejam transformadas para estabilização da média e da variância dos dados. Para a média, foi realizada diferenciação dos

dados, enquanto que para a variância, duas transformações foram realizadas, a saber: Logarítmica e de *Box-Cox*. Assim sendo, para cada categoria de produto, foram consideradas três séries temporais, uma representando os dados originais (“Original”), outra representando os dados transformados logaritmicamente (“Logarítmica”) e outra representando os dados resultantes da transformação de *Box-Cox* (“Box-Cox”).

A análise da função *auto.arima* permitiu concluir que os grupos Jornais e Revistas, Serviços e Papelaria apresentam componente sazonal evidente, pelo que, o modelo selecionado para estas séries foi do tipo $(p, d, q) \times (P, D, Q)_m$. As restantes séries foram estudadas tendo em consideração apenas a parte não sazonal do modelo ARIMA.

Para a seleção das ordens a aplicar aos modelos, foram analisadas as funções FAC e FACP das diferentes séries temporais. A título exemplificativo, são apresentados na Figura 5 os gráficos correspondentes às funções FAC e FACP para os dados originais da série Papelaria. Pretende-se assim descrever, de forma detalhada, a seleção do modelo apropriado tendo em conta os passos (3) e (4) apresentados no subcapítulo 3.5. Da observação da figura, constata-se que não existem picos relevantes fora da zona de insignificância estatística dos gráficos FAC e FACP. Conclui-se que os dados não têm auto correlações estatisticamente significativas entre os diversos *lags*. A análise inicial sugere que um possível modelo ARIMA seja representado por $(0, 0, 0) \times (1, 0, 0)_{12}$. Realizou-se de seguida uma verificação dos gráficos FAC e FACP dos resíduos deste modelo. Da análise da Figura 6, é possível observar um pico no *lag* 7 para o gráfico FACP. Este pico evidencia que a parte autorregressiva (p) do modelo deve ser corrigida considerando a ordem 7. O modelo é alterado passando a apresentar a seguinte forma: $(7, 0, 0) \times (1, 0, 0)_{12}$. Uma nova análise dos resíduos é realizada, concluindo-se que os mesmos apresentam evidências de serem um processo de ruído branco, não existindo necessidade de nova correção do modelo. Como suporte a esta conclusão, foi também realizado o teste de *Ljung-Box*, sendo os resultados apresentados na Tabela A1. O valor obtido comprova que não existem auto correlações significativas nos resíduos, dando suporte à conclusão alcançada pela análise dos gráficos FAC e FACP. O

modelo selecionado é mantido e o processo iterativo finaliza, iniciando-se o cálculo das previsões.

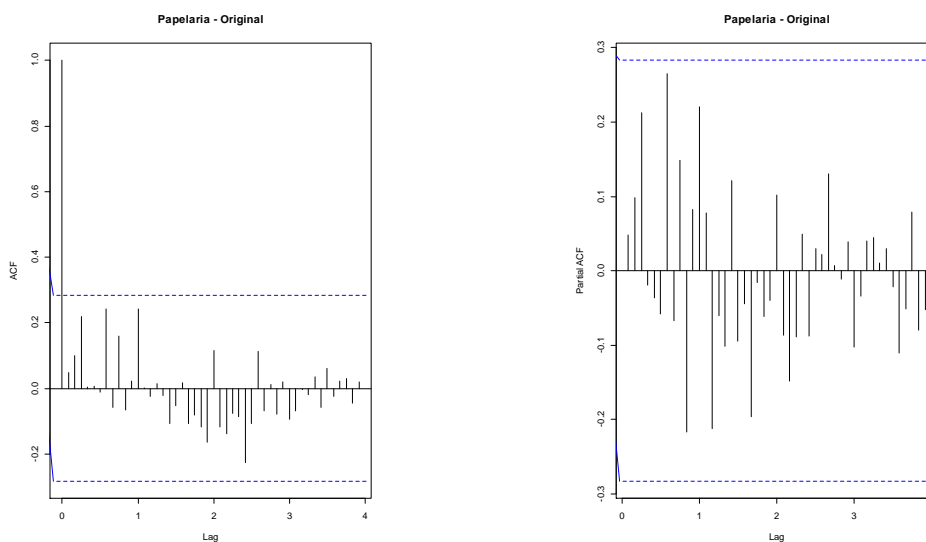


Figura 5 – Gráficos FAC (à esquerda) e FACP (à direita) para os dados originais da série temporal Papelaria.

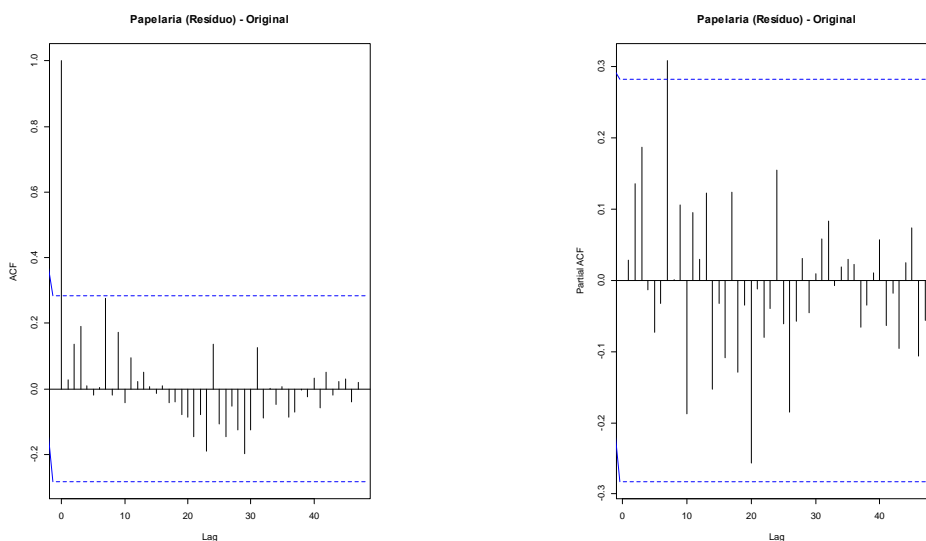


Figura 6 – Gráficos FAC (à esquerda) e FACP (à direita) dos resíduos dos dados originais da série temporal Papelaria.

3.6. Metodologia MLP

Neste estudo, foi utilizada a função *mlp* presente no pacote *nnfor* do *software* R. Este pacote é usado em paralelo com o pacote *neuralnet*, necessário para o treino e aprendizagem da rede. O algoritmo apresenta uma função de ativação do tipo sigmoide, sendo representada por:

$$f(z) = \frac{1}{1 + e^{-z}} \quad (7)$$

onde Z traduz o valor da função de soma.

O treino e aprendizagem da rede foi realizado com recurso ao algoritmo de retro propagação, mais especificamente retro propagação resiliente *rprop+* desenvolvido por Riedmiller e Braun (1993). Este algoritmo permite o ajuste dos pesos inicialmente atribuídos de forma aleatória, tendo como critério o erro gerado entre o resultado obtido e o resultado esperado. Este erro é propagado para a camada anterior onde os pesos são ajustados e assim sucessivamente até à camada de entrada. Este processo é repetido até que o erro seja inferior a um valor limite predeterminado. A Figura 7 apresenta uma ilustração básica da propagação de sinal no modelo MLP.

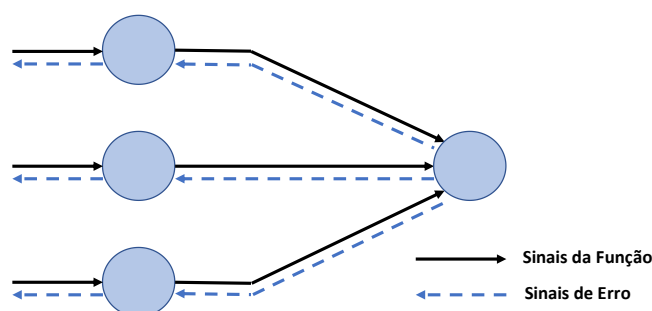


Figura 7 – Propagação de sinal num modelo MLP, adaptado de Haykin (2009).

3.6.1 Arquitetura da rede

Criar uma arquitetura de rede neuronal significa estabelecer valores para o número de camadas de cada tipo e o número de neurónios em cada camada. Para definir o número de camadas e neurónios numa rede MLP, algumas abordagens podem ser consideradas, como a experimentação sistemática e a intuição. Contudo, quer o tempo requerido quer a experiência do utilizador

podem comprometer a seleção da melhor arquitetura de rede para o caso específico.

Uma análise abrangente da literatura permitiu auferir que não existe uma regra fixa para construir a arquitetura de rede mais eficaz para cada problema em análise. A maioria dos autores afirmam que é no controlo das camadas escondidas e dos neurónios destas que reside a eficácia de uma rede neuronal. Para Ahmed *et al.* (2010), a definição do número de camadas escondidas permite o controlo da complexidade do modelo, sendo crítico evitar a sobreparametrização do mesmo. Makridakis *et al.* (2018), são da mesma opinião, seguindo algumas práticas para a seleção do número de neurónios destas camadas.

Segundo Heaton (2008), existem duas decisões a serem tomadas no que diz respeito às camadas escondidas: quantas camadas devem existir na rede e quantos neurónios devem estar presentes em cada uma dessas camadas. Este mesmo autor afirma que problemas que requerem duas camadas escondidas são raramente encontrados no dia-a-dia. A mesma opinião foi partilhada por Hornik *et al.* (1989) no seu estudo sobre as redes neuronais multicamada, onde demonstraram que uma camada é suficiente para modelar qualquer função contínua.

Neste estudo, a construção das redes neuronais teve em consideração uma camada de entrada, uma camada escondida e uma camada de saída. Quanto à camada de entrada, foi decidido manter o número de neurónios / *lags* identificados de forma automática pela função *mip*. Para a camada de saída, apenas um neurónio foi considerado, sendo o resultado deste neurónio o resultado global do modelo. Quanto à camada escondida, a decisão do número de neurónios é a parte mais importante na escolha da arquitetura da rede. Poucos neurónios significa que a rede é incapaz de detetar de forma adequada os sinais de um grupo de dados complicados. Demasiados neurónios, impõem à rede uma capacidade de processamento elevada, sendo os dados de entrada escassos para treinar os neurónios presentes nesta camada.

Tendo em conta a globalidade da literatura, quatro redes neuronais foram criadas para cada grupo de dados e para cada método de processamento dos mesmos. Cada rede apresenta as mesmas camadas de entrada e de saída,

sendo distinguida pelo número de neurónios que compõem a camada escondida, a saber:

- Rede 1: definida automaticamente pela função *mlp*;
- Rede 2: definida por um neurónio escondido;
- Rede 3: definida como: ARREDONDAR.PARA.CIMA $[1/2 (N_{\text{entrada}} + N_{\text{saída}})]$;
- Rede 4: definida como: ARREDONDAR.PARA.CIMA $[2/3 (N_{\text{entrada}} + N_{\text{saída}})]$ (Heaton, 2008).

3.7. Combinação de previsões dos métodos ARIMA e MLP

Segundo Caiado (2016), e considerando a grande maioria das situações, a combinação das previsões de diferentes métodos conduz a melhores previsões do que a utilização individual dos próprios métodos.

Neste estudo, a obtenção da previsão combinada dos métodos seguiu a metodologia apresentada por Caiado (2016), onde se atribuem pesos diferenciados às várias previsões. Estes pesos são inversamente proporcionais aos erros de previsão obtidos por estes métodos. Aqui, foi considerado o erro de previsão EAM (definido no subcapítulo 3.2), sendo a previsão combinada para o instante t (de janeiro de 2018 a dezembro de 2018) dada por:

$$P_t^{Comb} = \frac{(E_{total} - EAM^{ARI}) * P_t^{ARI} + (E_{total} - EAM^{MLP}) * P_t^{MLP}}{E_{total}} \quad (8)$$

onde EAM^{ARI} e EAM^{MLP} são os erros absolutos médios para a previsão ARIMA e MLP, respetivamente, sendo E_{total} o erro total dos dois métodos. P_t^{ARI} e P_t^{MLP} são as previsões para o instante t dos métodos ARIMA e MLP, respetivamente.

Para a obtenção dos erros de previsão a serem utilizados, um novo período de Treino foi definido, sendo constituído pelos primeiros três anos de vendas (36 observações). As seguintes 12 observações foram consideradas como o período de Ajustamento e as restantes como o período de Teste. Para este período de Ajustamento foram obtidas as previsões dos métodos ARIMA e MLP e calculados os respetivos erros. Estes foram então considerados na previsão combinada para o instante final (período de Teste). De referir que

apenas os modelos ARIMA e MLP de maior precisão (menor valor de EAM), para cada categoria de produto, são considerados nesta análise. A fase seguinte é a obtenção dos erros da previsão combinada e a sua comparação com os das demais metodologias.

4. Análise de Resultados

O desempenho dos modelos de previsão ARIMA e MLP, avaliado a partir dos erros de previsão EAM, REQM e EPAM e considerando os diferentes métodos de pré-processamento, pode ser encontrado na Tabela A2 e Tabela A3, respetivamente.

A partir dos resultados da Tabela A2, é possível observar que o erro de previsão das diferentes séries é bastante inferior quando uma transformação dos dados e remoção da tendência é aplicada. Existindo, em alguns casos, diferenças de erro superior a 100% quando comparados com os dados originais. A exceção ocorre para a série Jornais e Revistas, onde o desempenho de previsão foi superior com a utilização dos dados originais. A realização de um teste de normalidade de *Shapiro-Wilk* aos dados transformados da série Jornais e Revistas, demonstrou que os resíduos não se encontram normalmente distribuídos, refletindo um padrão heterocedástico não constante que afeta a precisão com a qual o erro padrão da média é estimado (ver Cohen *et al.*, 2003). Este poderá ser o motivo pelo qual a previsão utilizando os dados transformados obteve um tão baixo nível de precisão. Da análise dos resultados da Tabela A2 é também observável que todos os erros de previsão, salvo exceções pouco significativas, apontam para o mesmo modelo como sendo o mais preciso. Observa-se ainda que os resultados obtidos pelas transformações logarítmica e de Box-Cox são bastante similares, não existindo um claro indicador da melhor opção a aplicar.

A Tabela A3 sumariza o desempenho do método MLP para as várias possibilidades de pré-processamento assim como as suas combinações. Aqui, é possível observar que a combinação de diferentes métodos de pré-processamento traduz-se numa previsão mais eficiente. A exceção ocorre para a série temporal Serviços, onde a previsão utilizando os dados originais apresentou um erro de previsão inferior às restantes. Tal poderá ser explicado pela tendência decrescente acentuada da série, apresentando valores de vendas reduzidos para o período de Teste, fazendo com que previsão a partir dos dados pré-processados seja pouco eficiente. Para as séries temporais Guloseimas e Tabaco, e considerando o erro de previsão EAM, a combinação

que apresenta um erro inferior é a realização de transformação logarítmica seguida de remoção da sazonalidade. A série Jornais e Revistas adiciona mais um método de processamento à combinação ideal, que é a diferenciação dos dados. Já a série Papelaria apresenta como melhor combinação de tratamento dos dados a transformação de Box-Cox seguida da remoção da sazonalidade e da diferenciação.

Algumas observações interessantes podem ainda ser retiradas da análise dos resultados da Tabela A3, a saber:

- O ajustamento sazonal, a partir da remoção da componente de sazonalidade dos dados, e quando combinado com outros métodos de pré-processamento, fornece resultados significativamente superiores.
- A transformação dos dados utilizando o método de Box-Cox é ligeiramente uma melhor opção que a transformação logarítmica. Ambas permitem ainda obter resultados de previsão mais precisos.
- Não existem diferenças significativas considerando os diferentes erros de previsão apresentados. Com apenas ligeiras exceções para as séries Guloseimas e Papelaria, todos apontam para o mesmo tipo de pré-processamento como o mais eficaz na previsão.
- A realização de diferenciação de forma individual, induz um efeito adverso na previsão dos dados, fazendo com que estas se tornem bastante ineficazes. Resultado idêntico foi observado por Makridakis *et al.* (2018), onde a diferenciação de primeira ordem provocou um aumento significativo do erro de previsão.

A Tabela I apresenta um resumo do desempenho de previsão dos modelos mais precisos apresentados na Tabela A2 e Tabela A3. A seleção destes modelos foi realizada considerando o erro EAM. Esta apresenta ainda os resultados da combinação ARIMA e MLP descrita no subcapítulo 3.7. Os resultados demonstram que o método MLP é aquele que apresenta melhor desempenho para a maioria das séries temporais. A exceção ocorre para as séries Tabaco e Papelaria, onde a combinação ARIMA e MLP apresentou maior precisão. Para as séries Jornais e Revistas e Guloseimas, a diferença do erro EAM entre o método MLP e ARIMA foi inferior a 5%. Já para as séries Tabaco e Papelaria, esta diferença situou-se na ordem dos 10%. De salientar ainda, que

a diferença de erro encontrada na série Serviços foi superior a 100%, indicando um desempenho muito pouco eficiente do método ARIMA. Tal poderá ser explicado pelo facto desta série apresentar valores reais bastante reduzidos para o período de Teste, fazendo com que pequenas diferenças de previsão se transformem em grandes erros.

Tabela I – Comparação do desempenho da previsão dos métodos ARIMA, MLP e combinação ARIMA e MLP (janeiro 2018 a dezembro 2018).

Série Temporal	Modelo	EAM
Jornais e Revistas	ARIMA – Original (0, 0, 2) X (0, 0, 2) ₁₂	228.27
	MLP – Comb. [2] e [5] e [4] (5 nós)	217.25
	Combinação ARIMA e MLP	233.39
Guloseimas	ARIMA – Box-Cox (0, 1, 0)	12.26
	MLP – Comb. [2] e [5] (1 nó)	12.16
	Combinação ARIMA e MLP	12.16
Serviços	ARIMA – Log. (3, 1, 0) X (0, 0, 1) ₁₂	4.55
	MLP – Original (1 nó)	1.83
	Combinação ARIMA e MLP	2.58
Tabaco	ARIMA – Box-Cox (0, 1, 1)	17.16
	MLP – Comb. [2] e [5] (4 nós)	15.56
	Combinação ARIMA e MLP	15.54
Papelaria	ARIMA – Log. (0, 1, 1) X (1, 0, 2) ₁₂	59.46
	MLP – Comb. [3] e [5] e [4] (3 nós)	54.65
	Combinação ARIMA e MLP	54.55

De maneira a ser possível observar o comportamento da previsão mensal obtida para cada série temporal, foram construídos os gráficos de evolução de vendas para o período de Teste (ver Figura 8 e Tabela A4). Estes gráficos têm em consideração os modelos mais precisos apresentados na Tabela I. Para a série Jornais e Revistas, observa-se que os modelos ARIMA e MLP, e por consequência a combinação de ambos, têm boa capacidade em prever os movimentos de tendência e as flutuações sazonais existentes nos dados reais. Já para a série Guloseimas, os modelos não foram capazes de prever o pico de vendas observado entre abril e junho de 2018, produzindo uma situação de

previsão escassa. Para a série Serviços, os métodos de previsão foram capazes de captar a ligeira tendência crescente encontrada no período em análise. Observa-se ainda que o método ARIMA obteve uma previsão excessiva entre setembro e novembro de 2018, explicando assim o elevado erro de previsão encontrado na Tabela I. De forma similar ao observado na série Guloseimas, a série Tabaco, apresenta uma previsão de certa forma constante ao longo do ano, fazendo com que os modelos não sejam capazes de prever as flutuações de vendas existentes durante este período. Por sua vez, para a série papelaria, o modelo MLP e a combinação ARIMA e MLP foram capazes de captar as tendências e a sazonalidade existente nos dados reais. Contudo, estes não conseguiram prever os picos de vendas nos meses de abril e junho de 2018, apresentando previsões bastante reduzidas para estes meses.

Os modelos ARIMA e MLP, bem como a utilização combinada de ambos, mostraram ser abordagens complementares ao problema de previsão de séries temporais. Enquanto que o modelo tradicional baseia-se na descrição de auto correlações nos dados, o modelo MLP baseia-se no “aprender”, resolvendo problemas de otimização de maneira a encontrar um conjunto de parâmetros ou pesos, que minimizem a função de erro.

Os resultados encontrados reforçam a ideia da maior capacidade de previsão do método MLP quando comparada com a do método ARIMA. Ressalvam-se também os resultados encontrados pela combinação ARIMA e MLP. Esta combinação funciona como um sistema de ponderação dos erros produzidos pelos dois métodos, tendo permitindo obter resultados mais precisos para duas das cinco séries temporais alvo de estudo. Em adição, foi demonstrado que a capacidade de previsão dos métodos ARIMA e MLP é muito sensível ao tipo de dados. A remoção da tendência e da componente sazonal, mostraram ser os tipos de pré-processamento dos dados mais eficazes.

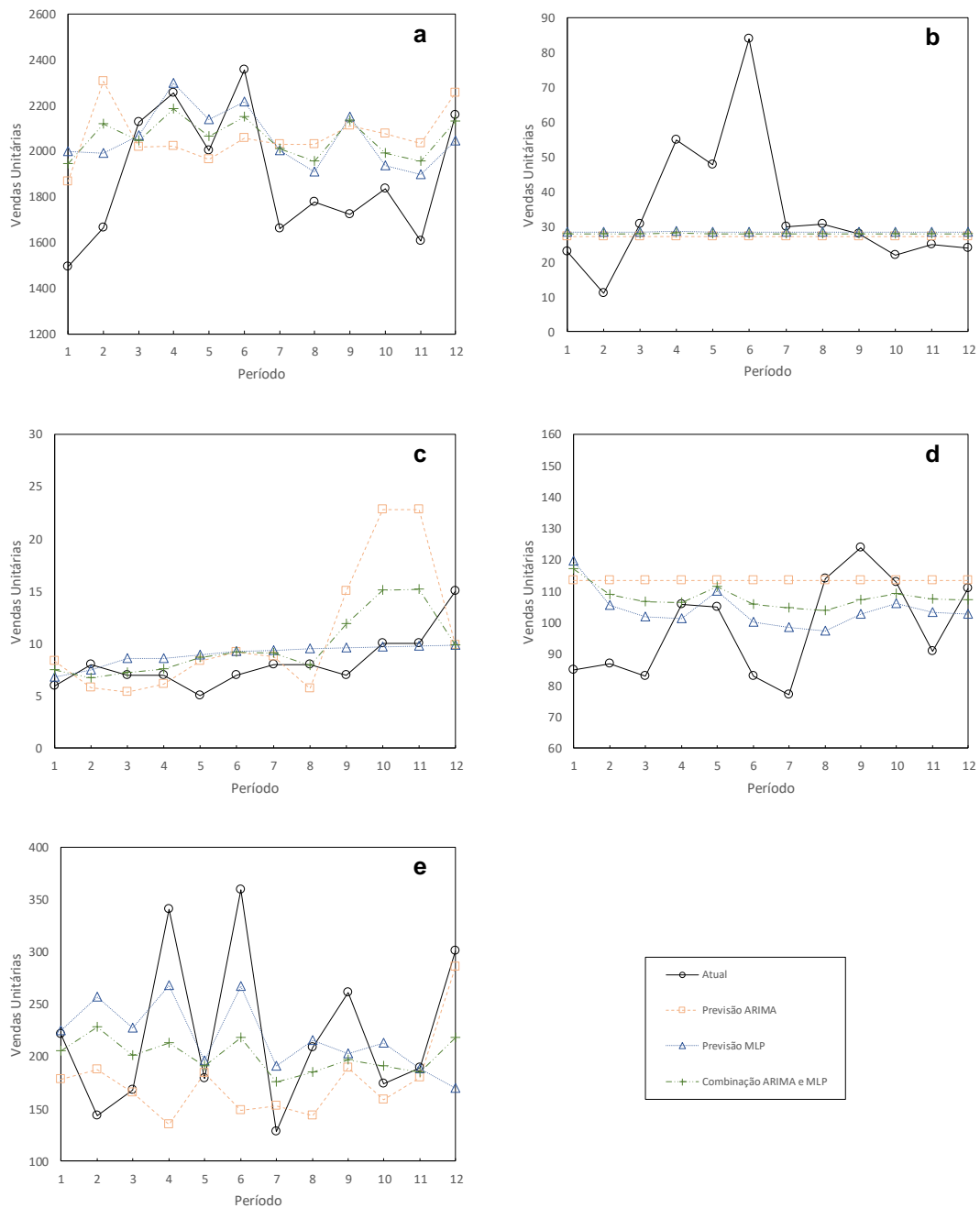


Figura 8 – Previsão de vendas mensais para as diferentes séries (entre janeiro 2018 a dezembro 2018): (a) Jornais e Revistas, (b) Guloseimas, (c) Serviços, (d) Tabaco e (e) Papelaria.

5. Conclusões, Limitações e Investigação Futura

O campo da previsão estatística tem obtido um grande desenvolvimento desde a década 40 do século passado, com o desenvolvimento de metodologias de alisamento exponencial, passando pela introdução da metodologia de *Box-Jenkins*, e mais recentemente pelo desenvolvimento da inteligência artificial com a criação das designadas redes neuronais (Makridakis *et al.*, 2018).

Previsões de vendas precisas podem ter um impacto elevado na gestão eficaz das operações de retalho. As séries temporais provenientes deste sector, exibem um forte padrão de tendência e variação sazonal, colocando desafios ao desenvolvimento de modelos de previsão eficazes. Como modelar de forma precisa estas séries e como melhorar as suas previsões são as grandes questões que se colocam. Apesar dos elevados esforços dos investigadores, os estudos existentes não levam a um consenso generalizado sobre o desempenho de previsão das metodologias ARIMA e MLP aplicadas ao sector do retalho.

O propósito deste trabalho residia na comparação do desempenho das metodologias ARIMA e MLP aplicadas ao caso de estudo de cinco grupos de produtos comercializados por uma empresa portuguesa do sector do retalho. O trabalho objetivava ainda dar resposta ao impacto do pré-processamento dos dados no desempenho da previsão, e à eficácia da combinação dos métodos de previsão.

A modelação das séries temporais utilizando o método ARIMA teve em consideração os dados de originais e os dados transformados com recurso aos vários tipos de pré-processamento. Foram criados diversos modelos, cada um representando a melhor modelação para o tipo de dados e para cada grupo de produtos analisados (ver Tabela A2). Dentro de cada grupo, a seleção do modelo mais preciso foi feita com recurso ao erro de previsão EAM (ver Tabela I). Para o método MLP, uma abordagem similar foi realizada. Aqui, foram adicionados mais tipos de processamento dos dados, resultantes da combinação de diferentes métodos de pré-processamento. Foram também criados diversos modelos representando o tipo de dados e os diferentes produtos analisados. A diferença residiu nas várias arquiteturas de rede aplicadas a estes modelos. A

seleção da arquitetura mais eficaz para cada tipo de dados foi realizada com recurso ao erro de previsão EAM.

As previsões de vendas para o instante de previsão considerado, foram produzidas utilizando os modelos selecionados. Os resultados indicam que o método MLP é aquele que permite modelar de forma mais eficaz o comportamento de uma série temporal com padrões de tendência e sazonalidade. A presença de componentes não lineares nos dados, ofereceu uma vantagem significativa ao método de *Machine Learning*. Contudo, o modelo ARIMA não pode ser descartado. Os resultados mostram que o seu desempenho foi muito próximo ao obtido pelo método MLP, indicando que este poderá ser um bom método para a realização de operações de *screening* ou seleção, ou até mesmo de previsões de natureza linear. O pré-processamento dos dados via transformação dos dados, remoção da tendência e ajustamento sazonal, permitiu um aumento do desempenho dos métodos de previsão analisados. Conclusão similar foi obtida por Zhang e Qi (2005), Ramos *et al.*, 2015, Makridakis *et al.*, 2018, entre outros. Outro facto a concluir é o bom desempenho obtido pela utilização combinada dos métodos ARIMA e MLP. Os resultados da Tabela I evidenciam que esta combinação foi a mais precisa para dois dos cinco grupos de produtos analisados.

Diversas limitações podem ser apontadas ao estudo. A primeira, diz respeito à elevada irregularidade encontrada nos dados de vendas, que poderá ter tido impacto no desempenho da previsão dos métodos aqui testados. A segunda, aponta ao facto de apenas um método tradicional e um método de rede neuronal terem sido testados. Por fim, a base de dados utilizada diz respeito a uma empresa que atua num segmento de negócio específico, comprometendo de certa forma a generalização dos resultados para todo o sector do retalho.

Como proposta de trabalho futuro, destaca-se a investigação de modelos de previsão híbridos, que permitam juntar a capacidade de previsão do modelo ARIMA com a capacidade das redes neuronais em reduzir os resíduos, diminuindo o erro na fase de treino. Este conceito não é novo, tendo sido investigado por Tseng *et al.* (2002), Zhang (2003), entre outros. Contudo, carece de uma investigação mais aprofundada, que permita o estabelecimento de uma metodologia padrão para o uso do sector.

Referências Bibliográficas

Adya, M., e Collopy, F. (1998). How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting*, 17 (5-6), 481-495.

Agrawal, D., e Schorling, C. (1996). Market share forecasting: An empirical comparison of artificial neural networks and multinomial logit model. *Journal of Retailing*, 72 (4), 383-407.

Ahmed, N. K., Atiya, A. F., Gayar, N. E., e El-Shishiny, H. (2010). An Empirical Comparison of Machine Learning Model for Time Series Forecasting. *Econometric Reviews*, 29 (5-6), 594-621.

Alon, I., Qi, M., e Sadowski, R. J. (2001). Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consume Services*, 8 (3), 147-156.

Barksdale, H. C., e Hilliard, J. E. (1975). A cross-spectral analysis of retail inventories and sales. *Journal of Business*, 48 (3), 365-382.

Box, G. E. P., e Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society Series B (Methodological)*, 26 (2), 211-252.

Box, G. E. P., e Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. *Holden Day*, San Francisco.

Caiado, J. (2016). *Métodos de Previsão em Gestão – Com Aplicações em Excel*, 2ª Ed. Lisboa: Edições Sílabo, Lda.

Chu, C.-W., e Zhang, G. P. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. *International Journal of Production Economics*, 86 (3), 217-231.

Cohen J., Cohen P., West S. G., Aiken L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3ª Ed. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Crone, S. F., Hibon, M., e Nikolopoulos, K. (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting*, 27 (3), 635-660.

Cui, G., Wong, M. L., e Lui, H.-K. (2006). Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming. *Management Science*, 52 (4), 597-612.

De Gooijer, J. G., e Franses, P. H. (1997). Forecasting and seasonality. *International Journal of Forecasting*, 13, 303-305.

Farway, J., e Chatfield, C. (1995). Time series forecasting with neural networks: A comparative study using the airline data. *Applied Statistics*, 47 (2), 231-250.

Franses, P. H., e Draisma, G. (1997). Recognizing changing seasonal patterns using artificial neural networks. *Journal of Econometrics*, 81 (1), 273-280.

Gardner, E., e McKenzie, E. (1989). Seasonal exponential smoothing with damped trends. *Management Science*, 35 (3), 372-376.

Ghysels, E., Granger, C. W. J., e Siklos, P. L. (1996). Is seasonal adjustment a linear or nonlinear data filtering process?. *Journal of Business and Economics Statistics*, 14 (3), 374-386.

Gorr, W. L. (1994). Research prospective on neural network forecasting. *International Journal of Forecasting*, 10 (1), 1-4.

Heaton, J. (2008). *Introduction to Neural Networks with Java*, 2ª Ed. Chesterfield-Missouri: Heaton Research, Inc.

Haykin, S. (2009). *Neural networks and learning machines*, 3ª Ed. Hamilton-Ontario: Pearson Education, Inc.

Hornik, K., Stinchcombe, M., e White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2 (5), 359-366.

Huang, T., Fildes, R., e Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research*, 237 (2), 738-748.

Hyndman, R. J., e Athanasopoulos, G. (2018). *Forecasting: principles and practice*, 2ª Ed. Melbourne: OTexts.

Ittig, P. T. (1997). A seasonal index for business. *Decision Sciences*, 28 (2), 335-355.

Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibdon, M., Lewandowski, R., Newton, J., Parzen, E., e Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, 1, 111-153.

Makridakis, S., e Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, 16 (4), 451-476.

Makridakis, S., Spiliotis, E., e Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13 (3).

Miller, D. M., e Williams, D. (2003). Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy. *International Journal of Forecasting*, 19 (4), 669-684.

Nelson, C. R., e Plosser, C. I. (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics*, 10 (2), 139-162.

Nelson, M., Hill, T., Remus, T., e O'Connor, M. (1999). Time series forecasting using NNs: Should the data be deseasonalized first?. *Journal of Forecasting*, 18 (5), 359-367.

Pierce, D. A. (1977). Relationships-and lack of thereof-between economic time series, with special reference to money and interest rates. *Journal of the American Statistical Association*, 72 (357), 11-26.

Pillo, G. D., Latorre, V., Lucidi, S., e Procacci, E. (2013). An application of learning machines to sales forecasting under promotions (Technical Reports, 2013-04). Roma, Itália: Sapienza Università Di Roma, Dipartimento Di Ingegneria Informatica Automatica e Gestionale Antonio Ruberti.

Ramos, P., Santos, N., e Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing*, 34, 151-163.

Reed, R. D., e Marks, R. J. (1999). *Neural Smithing: Supervised learning in feedforward artificial neural networks*, 6ª Ed. Cambridge: MIT Press.

Riedmiller, M. e Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *International Conference on Neural Networks*, 586-591.

Ripley, B. (2019). MASS: Support functions and datasets for Venables and Ripley's MASS. Software, R package.

Sharda, R. e Patil, R. B., (1992). Connectionist approach to time series prediction: An empirical test. *Journal of intelligent Manufacturing*, 3 (5), 317-323.

Thall, N. (1992). Neural forecasts: A retail sales booster. *Discount Merchandiser*, 23 (10), 41-42.

Trapletti, A. e Hornik, K. (2019). tseries: Time series analysis and computational finance. Software, R package.

Tseng, F.-M., Yu, H.-C. e Tzeng, G.-H. (2002). Combining neural network model with seasonal time series ARIMA model. *Technological Forecasting and Social Change*, 69 (1), 71-87.

Zhang, G. (1998). Linear and nonlinear time series forecasting with artificial neural networks. (Tese de Doutorado). Kent State University, Estados Unidos da América.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.

Zhang, G. P., e Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160 (2), 501-514.

Anexos

Tabela A1 – Modelos ARIMA utilizados na previsão de vendas. Teste ADF aos dados de entrada e teste de Ljung-Box aos resíduos dos modelos selecionados.

Série Temporal	Pré-process	Teste ADF (<i>p-value</i>)	Modelo inicial	Modelo selecionado	Teste Ljung-Box (<i>p-value</i>)
Jornais e Revistas	Original	0.70	(0, 0, 2) X (0, 0, 1) ₁₂	(0, 0, 2) X (0, 0, 2) ₁₂	0.52
	Log.	< 0.01	(1, 1, 1) X (0, 0, 1) ₁₂	(1, 1, 1) X (0, 0, 1) ₁₂	0.88
	Box-Cox	< 0.01	(1, 1, 1) X (1, 0, 1) ₁₂	(1, 1, 6) X (1, 0, 1) ₁₂	0.99
Guloseimas	Original	0.03 ⁽¹⁾	(1, 0, 3)	(1, 0, 3)	0.99
	Log.	< 0.01	(0, 1, 0)	(0, 1, 1)	0.94
	Box-Cox	< 0.01	(0, 1, 0)	(0, 1, 1)	0.96
Serviços	Original	0.68	(0, 0, 3) X (0, 0, 1) ₁₂	(0, 0, 3) X (0, 0, 1) ₁₂	0.99
	Log.	0.60 ⁽²⁾	(0, 1, 0) X (0, 0, 1) ₁₂	(3, 1, 0) X (0, 0, 1) ₁₂	0.99
	Box-Cox	0.01	(0, 1, 0) X (0, 0, 1) ₁₂	(2, 1, 0) X (0, 0, 1) ₁₂	0.85
Tabaco	Original	0.31	(0, 0, 2)	(0, 0, 7)	0.32
	Log.	< 0.01	(0, 1, 1)	(0, 1, 1)	0.67
	Box-Cox	< 0.01	(0, 1, 1)	(0, 1, 1)	0.79
Papeleria	Original	0.07	(0, 0, 0) X (1, 0, 0) ₁₂	(7, 0, 0) X (1, 0, 0) ₁₂	0.88
	Log.	< 0.01	(0, 1, 0) X (1, 0, 0) ₁₂	(0, 1, 1) X (1, 0, 2) ₁₂	0.93
	Box-Cox	< 0.01	(0, 1, 0) X (1, 0, 0) ₁₂	(0, 1, 1) X (1, 0, 2) ₁₂	0.93

Notas:

⁽¹⁾: Um teste de *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS) foi realizado de maneira a comprovar este valor. O teste apresentou um *p-value* > 0.1, indicando que a série é estacionária.

⁽²⁾: Após a transformação dos dados, um teste de *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS) foi realizado, apresentando um valor de *p-value* > 0.1, indicando que a série é estacionária.

Tabela A2 – Desempenho da previsão do método ARIMA para o período de janeiro 2018 a dezembro 2018.

Série Temporal	Modelo		EAM	REQM	EPAM (%)
	Pré-process.	ARIMA			
Jornais e Revistas	Original	$(0, 0, 2) \times (0, 0, 2)_{12}$	228.27	329.25	16.34
	Log.	$(1, 1, 1) \times (0, 0, 1)_{12}$	516.40	584.37	28.92
	Box-Cox	$(1, 1, 6) \times (1, 0, 1)_{12}$	770.81	892.43	42.09
Guloseimas	Original	$(1, 0, 3)$	26.27	28.93	108.69
	Log.	$(0, 1, 1)$	12.57	20.63	33.07
	Box-Cox	$(0, 1, 1)$	12.26	19.89	34.29
Serviços	Original	$(0, 0, 3) \times (0, 0, 1)_{12}$	30.35	31.12	399.22
	Log.	$(3, 1, 0) \times (0, 0, 1)_{12}$	4.55	6.17	53.81
	Box-Cox	$(2, 1, 0) \times (0, 0, 1)_{12}$	5.52	6.17	73.43
Tabaco	Original	$(0, 0, 7)$	119.01	122.58	124.19
	Log.	$(0, 1, 1)$	17.34	21.60	19.99
	Box-Cox	$(0, 1, 1)$	17.16	21.41	19.78
Papeleria	Original	$(7, 0, 0) \times (1, 0, 0)_{12}$	91.41	107.75	43.56
	Log.	$(0, 1, 1) \times (1, 0, 2)_{12}$	59.46	92.04	22.52
	Box-Cox	$(0, 1, 1) \times (1, 0, 2)_{12}$	60.58	93.07	22.94

Tabela A3 – Desempenho da previsão do método MLP para o período de janeiro 2018 a dezembro 2018 (considerando a arquitetura de rede com menor valor de erro EAM).

Série Temporal	Modelo	Rede		EAM	REQM	EPAM (%)
	Pré-process.	Lags	Nós escondidos			
Jornais e Revistas	[1] Original	[2, 6, 8, 11, 12]	1	269.91	348.42	14.83
	[2] Log.	[2, 6, 8, 11, 12]	1	308.28	383.33	16.63
	[3] Box-Cox	[2, 5, 6, 8, 11]	3	247.34	276.90	13.90
	[4] Difer.	[1, 6, 7, 8, 9, 10, 12]	4	559.92	622.04	31.00
	[5] Dessazo.	[1, 2, 3, 6, 7, 9, 11, 12]	1	366.74	433.27	21.49
	Comb. [2] e [5]	[1, 2, 3, 6, 7, 9, 11, 12]	5	334.25	389.77	19.34
	Comb. [3] e [5]	[2, 3, 6, 11, 12]	5	297.41	360.29	17.19
	Comb. [5] e [4]	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]	8	625.22	686.90	34.82
	Comb. [2] e [5] e [4]	[1, 3, 4, 6, 7, 9, 10]	5	217.25	262.80	12.67
	Comb. [3] e [5] e [4]	[1, 3, 4, 6, 7, 9, 10]	4	238.73	280.93	13.73
Guloseimas	[1] Original	[1, 3, 12]	2	17.29	20.53	72.02
	[2] Log.	Auto [1, 8]	1	12.51	20.44	35.03
	[3] Box-Cox	[1, 3, 12]	3	12.64	19.17	39.19
	[4] Difer.	Auto [1, 2]	1	19.89	23.04	71.08
	[5] Dessazo.	[1, 3]	1	20.18	22.29	78.23
	Comb. [2] e [5]	[1, 6, 10]	1	12.16	19.49	35.47
	Comb. [3] e [5]	[1, 10]	2	12.62	18.48	41.43
	Comb. [5] e [4]	[1, 2, 5, 12]	1	16.08	24.55	39.65
	Comb. [2] e [5] e [4]	[1, 5, 12]	2	13.73	21.54	41.47
	Comb. [3] e [5] e [4]	[1, 5, 12]	3	12.23	22.72	28.92

UTILIZAÇÃO DE MODELOS ESTATÍSTICOS E MACHINE LEARNING PARA A PREVISÃO
DE VENDAS NO SECTOR DO RETALHO – UM ESTUDO COMPARATIVO
Christian Raposo Lopes

Série Temporal	Modelo	Rede		EAM	REQM	EPAM (%)
	Pré-process.	Lags	Nós escondidos			
Serviços	[1] Original	[1, 2, 3]	1	1.83	2.32	24.19
	[2] Log.	[1, 4]	2	7.00	7.45	84.47
	[3] Box-Cox	[1, 2, 3]	1	1.90	2.39	25.63
	[4] Difer.	[1, 2, 10, 12]	3	5.61	6.13	71.30
	[5] Dessazo.	[1, 3, 6, 11, 12]	2	3.10	3.68	40.38
	Comb. [2] e [5]	[1, 4, 6, 12]	2	5.41	5.94	63.73
	Comb. [3] e [5]	[1, 3, 6, 11, 12]	3	3.45	3.94	44.69
	Comb. [5] e [4]	[1, 2, 10, 12]	1	6.52	7.91	75.68
	Comb. [2] e [5] e [4]	[1, 2, 12]	3	3.73	4.31	43.68
	Comb. [3] e [5] e [4]	[1, 2, 10, 12]	3	6.23	7.97	40.95
Tabaco	[1] Original	[1, 3, 5, 7, 10, 12]	4	18.26	22.74	20.79
	[2] Log.	[1, 3, 5, 9]	3	16.45	18.91	17.94
	[3] Box-Cox	[1, 3, 5, 11]	3	16.07	19.24	18.05
	[4] Difer.	[1, 2, 3, 4, 7, 8]	1	27.92	32.08	27.53
	[5] Dessazo.	[1, 3, 4, 5, 7, 10, 12]	5	16.72	19.21	18.36
	Comb. [2] e [5]	[1, 3, 4, 5, 7, 12]	4	15.56	17.62	16.87
	Comb. [3] e [5]	[1, 3, 5, 7, 12]	4	16.86	21.51	19.31
	Comb. [5] e [4]	[1, 2, 3, 4, 5, 7, 8, 9, 11, 12]	1	25.31	30.78	25.04
	Comb. [2] e [5] e [4]	[1, 2, 3, 5, 6, 7, 8, 12]	1	18.64	22.13	19.23
	Comb. [3] e [5] e [4]	[5, 7, 12]	3	19.05	22.13	20.34
Papelaria	[1] Original	[7, 12]	2	61.17	83.37	29.67
	[2] Log.	[7, 12]	1	57.72	77.76	24.12
	[3] Box-Cox	[2, 5, 7, 12]	1	56.06	80.52	23.52
	[4] Difer.	[1, 2, 5, 6, 10, 11]	5	82.85	102.37	47.57
	[5] Dessazo.	[1, 3, 6, 9, 12]	3	85.71	98.69	48.49
	Comb. [2] e [5]	[1, 2, 6, 7, 8, 12]	1	95.18	113.52	52.21
	Comb. [3] e [5]	[1, 2, 6, 7, 8, 12]	1	92.39	110.21	50.41
	Comb. [5] e [4]	[1, 2, 5, 9, 10, 12]	5	89.19	96.22	49.28
	Comb. [2] e [5] e [4]	[1, 8, 12]	5	55.95	71.63	25.74
	Comb. [3] e [5] e [4]	[1, 8, 12]	3	54.65	68.67	26.04

Tabela A4 – Previsão de vendas para as diversas metodologias em análise entre janeiro 2018 e dezembro 2018 (valores utilizados na Figura 8).

Série Temporal	Instante temporal	Dados Originais	Previsão ARIMA	Previsão MLP	Previsão Comb. ARIMA-MLP
Jornais e Revistas	jan. 2018	1493	1868	2000	1945
	fev. 2018	1667	2305	1990	2120
	mar. 2018	2126	2017	2070	2048
	abr. 2018	2257	2024	2298	2185
	mai. 2018	2003	1965	2139	2067
	jun. 2018	2358	2056	2217	2150
	jul. 2018	1662	2029	2001	2013
	ago. 2018	1778	2029	1909	1959
	set. 2018	1725	2112	2153	2136
	out. 2018	1837	2075	1936	1994
	nov. 2018	1606	2033	1900	1955
	dez. 2018	2158	2254	2048	2133
Guloseimas	jan. 2018	23	27	29	28
	fev. 2018	11	27	29	28
	mar. 2018	31	27	29	28
	abr. 2018	55	27	29	28
	mai. 2018	48	27	29	28
	jun. 2018	84	27	29	28
	jul. 2018	30	27	29	28
	ago. 2018	31	27	29	28
	set. 2018	28	27	29	28
	out. 2018	22	27	29	28
	nov. 2018	25	27	29	28
	dez. 2018	24	27	29	28

UTILIZAÇÃO DE MODELOS ESTATÍSTICOS E MACHINE LEARNING PARA A PREVISÃO
DE VENDAS NO SECTOR DO RETALHO – UM ESTUDO COMPARATIVO
Christian Raposo Lopes

Série Temporal	Instante temporal	Dados Originais	Previsão ARIMA	Previsão MLP	Previsão Comb. ARIMA-MLP
Serviços	jan. 2018	6	8	7	7
	fev. 2018	8	6	7	7
	mar. 2018	7	5	9	7
	abr.2018	7	6	9	8
	mai. 2018	5	8	9	9
	jun. 2018	7	9	9	9
	jul. 2018	8	9	9	9
	ago. 2018	8	6	10	8
	set. 2018	7	15	10	12
	out. 2018	10	23	10	15
	nov. 2018	10	23	10	15
	dez. 2018	15	10	10	10
Tabaco	jan. 2018	85	114	120	117
	fev. 2018	87	114	106	109
	mar. 2018	83	114	102	107
	abr.2018	106	114	101	106
	mai. 2018	105	114	110	112
	jun. 2018	83	114	100	106
	jul. 2018	77	114	98	105
	ago. 2018	114	114	97	104
	set. 2018	124	114	103	107
	out. 2018	113	114	106	109
	nov. 2018	91	114	103	108
	dez. 2018	111	114	103	107

UTILIZAÇÃO DE MODELOS ESTATÍSTICOS E MACHINE LEARNING PARA A PREVISÃO
DE VENDAS NO SECTOR DO RETALHO – UM ESTUDO COMPARATIVO
Christian Raposo Lopes

Série Temporal	Instante temporal	Dados Originais	Previsão ARIMA	Previsão MLP	Previsão Comb. ARIMA- MLP
Papelaria	jan. 2018	222	179	225	206
	fev. 2018	144	188	257	228
	mar. 2018	168	165	227	202
	abr.2018	341	135	268	213
	mai. 2018	179	184	196	191
	jun. 2018	360	149	267	218
	jul. 2018	128	153	191	175
	ago. 2018	209	143	215	185
	set. 2018	261	190	203	197
	out. 2018	174	159	213	191
	nov. 2018	189	180	189	185
	dez. 2018	301	286	170	218

Exemplo do código utilizado no *software* R para a Metodologia ARIMA:



Guloseimas.R

Exemplo do código utilizado no *software* R para a Metodologia MLP:



MLP_Combi_Gulosei
mas.R