

iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Potenciar a gestão de lojas de retalho através de previsões de tráfego com técnicas de *machine learning*

Tiago André Ramos Casimiro

Mestrado em Ciência de Dados

Orientadora:

Doutora Diana Elisabeta Aldea Mendes, Professora Associada,
ISCTE - Instituto Universitário de Lisboa

Setembro, 2023

Departamento de Métodos Quantitativos para Gestão e
Economia

Departamento de Ciências e Tecnologia da Informação

**Potenciar a gestão de lojas de retalho através de previsões
de tráfego com técnicas de *machine learning***

Tiago André Ramos Casimiro

Mestrado em Ciência de Dados

Orientadora:

Doutora Diana Elisabeta Aldea Mendes, Professora Associada,
ISCTE - Instituto Universitário de Lisboa

Setembro, 2023

*Mestres são os que,
mais do que conseguir reter conhecimento,
são capazes de o partilhar.*

Esta dissertação é dedicada a todos os que, por ela, se queiram sentir inspirados.

Agradecimento

A presente dissertação é o resultado de muitos anos de aprendizagem, mais do que os que eu próprio posso lembrar. A curiosidade que me levou a estudar este tema é a mesma que me acompanha em tantos outros momentos do dia. A curiosidade que faz com que olhe para os problemas como desafios à espera de solução, mais do que como obstáculos, e esse espírito, aos poucos, foi-me passado pelas inúmeras pessoas que direta ou indiretamente se cruzaram comigo no bolinar da vida e a quem não tenho forma justa de agradecer.

À professora Diana Aldea Mendes, orientadora desta dissertação, o meu muito obrigado pelo acompanhamento durante todo este processo, mas acima de tudo pela forma inspiradora como lecionou as unidades curriculares que tive o prazer de frequentar. Destaco, em particular, a unidade curricular de Análise de Séries Temporais e Previsão, que desde a primeira aula, há cerca de um ano e meio, me colocou a pensar que este estudo seria um dia uma possibilidade.

A todos os docentes que, em maior ou menor período de tempo, tiveram contacto comigo, o meu obrigado pelos incontáveis ensinamentos e paciência. Tenho como base olhar para a posição de docente como uma pessoa que, mais do que passar conhecimento, deve inspirar os seus alunos a querer ser mais, a querer ser melhor e, felizmente, múltiplas vezes tive a sorte de me deparar com isso mesmo. Este trabalho é também fruto de todas essas pequenas grandes influências que tiveram na minha pessoa.

À Altice Portugal, aqui representada, e bem, pelas pessoas de Romeu Coelho, Matias Catarino e Inês Vaz, o o meu muito obrigado pela disponibilização dos dados, pelo acompanhamento, muitas vezes informal, pela preocupação e atenção que tiveram comigo e com este estudo. Todos os debates de ideias que aconteceram ao longo deste período foram de altíssima importância para o resultado que aqui se apresenta.

Ao Marcelo dos Santos e à Madalena Guiomar, líderes das equipas de que fiz parte na Decathlon Sintra, quando comecei a trabalhar há uns já longínquos 5 anos, o meu muito obrigado pela fonte de inspiração que se tornaram, pelos vários conselhos e pela disponibilidade para me ajudar a entrar neste Mestrado, quando poucos o fariam. A forma como hoje olho para uma loja de retalho é altamente influenciada pelas partilhas que foram tendo comigo ao longo dos tempos e, estou certo, que o bichinho que me levou a abraçar este projeto começou a crescer nos tempos em que tinha o prazer de vos ouvir debater a gestão de uma loja de retalho. Para sempre ficará a expressão "só falta mesmo aqui um papel a dizer que temos caracóis".

A todos os meus grandes amigos, Miguel, Filipe, Bruno, Nuno, o meu muito obrigado por serem quem são e por me empurrarem para a frente, terem as palavras certas (e erradas) nos momentos certos, por terem sempre um sorriso e um abraço de companheirismo pronto e, acima de tudo, por sempre terem sido figuras para as quais olhei com muita, mas mesmo muita, admiração. Se hoje sou a pessoa que sou, capaz de apresentar este trabalho, muito se deve aos momentos que partilhámos.

À minha querida Inês Miranda, o meu muito obrigado nunca será suficiente para agradecer o porto de abrigo que és, e que permite que o melhor de mim se dê a conhecer. Muitas foram as horas em que te sentaste a meu lado enquanto escrevia este documento, apenas para o caso de eu precisar de alguma coisa. O teu desejo de saber mais sobre Ciência de Dados para me poder ajudar sentiu-se a quilómetros de distância, e isso vale por tudo. Um obrigado gigante também à família Miranda pelo apoio e palavras de encorajamento que tantas vezes me dão e pelos dias veranis de guitarra e churrasco que tantas vezes foram interrompidos para verificar os modelos que hoje estão nestas páginas

Por fim, termino com o que começou verdadeiramente tudo isto, a minha família. Tudo o que faço é para vos deixar orgulhosos. Aos meus pais o obrigado por sempre me darem a possibilidade de seguir os meus sonhos, de me aconselharem e apoiarem, mesmo quando quis tomar as decisões mais arriscadas. Sem o vosso apoio nada disto teria sido possível. Ao meu irmão e irmã, o meu obrigado por me darem uma fonte de energia inesgotável. Espero que um dia consiga estar à altura da imagem que vocês têm de mim. Aos meus tios e avós, o meu obrigado por serem as referências que são para mim.

A todos, o meu muito obrigado.

Resumo

A presente dissertação pretende fornecer duas ferramentas ligadas ao tráfego de lojas de retalho físico, que possam auxiliar as equipas da gestão da rede MEO. A primeira, uma previsão de valores mensais de tráfego para um horizonte temporal de vinte e quatro meses, e a segunda, uma previsão de valores horários de tráfego para uma semana. Este é um caso de escassa investigação, facto descrito na Revisão de Literatura. Começou por se recolher dados mensais de dez lojas, seleccionadas através de análise *cluster*, e dados horários de uma das maiores lojas da rede MEO. Foram também utilizadas as condições climatéricas, eventos externos e feriados, de forma a entender melhor o fluxo de clientes. Com a utilização de vários modelos foi possível chegar a um vasto leque de abordagens, experimentadas e avaliadas face ao que existia como ferramenta na rede MEO. As previsões mensais produzidas para as lojas obtiveram resultados positivos, entre os 5,640% e 20,504% de MAPE, e a previsão mensal para a rede agregada obteve 4,489% de MAPE para 2023, sendo agora aplicada na prática na empresa. No tráfego horário, o melhor modelo alcançado seguiu uma abordagem *ensemble*, com *ExpSmoothing*, *XGBoost* e LSTM, produzindo previsões com 11,493 de MAE, que constituem uma melhoria de 35,7%, face à ferramenta usada pela equipa de gestão. Com estes testes comprovou-se que o modelo *ensemble* apresenta vantagens face à utilização individual dos modelos e que as duas ferramentas conseguiram melhorar os inputs disponíveis às equipas de gestão da rede de lojas MEO.

Abstract

This dissertation aims to provide two tools linked to physical retail store traffic that can help MEO network management teams. The first is a forecast of monthly traffic for a time horizon of twenty-four months, and the second is a forecast of hourly traffic for one week. This is a case with little research, as described in the Literature Review. The first step was to collect monthly data from ten stores, selected through cluster analysis, and hourly data from one of the largest stores in the MEO network. Weather conditions, external events and public holidays were also used to better understand the flow of customers. By experimenting with multiple models, it was possible to come up with a wide range of approaches, which were tested and evaluated against what existed as a tool in the MEO network. The monthly forecasts produced for the stores obtained positive results, ranging from 5.640% to 20.504% MAPE, and the monthly forecast for the aggregate network obtained 4.489% MAPE for 2023, and is now being applied at the company. In hourly traffic, the best model achieved followed an ensemble approach, with ExpSmoothing, XGBoost and LSTM, producing forecasts with 11.493 MAPE, which is a 35.7% improvement on the tool used by the management team. These tests proved that the ensemble model has advantages over using the models individually and that the two tools managed to improve the inputs delivered to the MEO network management teams.

Índice

Agradecimento	iii
Resumo	v
Abstract	vii
Lista de Tabelas	xi
Lista de Figuras	xiii
Capítulo 1. Introdução	1
Capítulo 2. Revisão da Literatura	3
2.1. Critérios de Pesquisa	3
2.2. Artigos Analisados	4
2.3. Áreas de aplicação	6
2.4. Técnicas	9
Capítulo 3. Metodologia	13
3.1. <i>Business Understanding</i>	13
3.2. <i>Data Understanding</i>	15
3.2.1. Seleção de lojas	15
3.2.2. Recolha de dados	17
3.3. <i>Data Preparation</i>	18
3.3.1. Descrição do <i>dataset</i>	18
3.3.2. Tratamento do período da pandemia COVID-19	20
3.4. <i>Modeling</i>	21
3.4.1. Conjuntos de treino, conjuntos de teste e critérios de avaliação	22
3.4.2. Previsão de tráfego mensal	24
3.4.3. Previsão de tráfego horário	25
3.5. <i>Evaluation</i>	29
3.6. <i>Deployment</i>	29
Capítulo 4. Resultados e Discussão	31
4.1. Tráfego mensal	31
4.2. Tráfego horário	34
4.3. Discussão	46
Capítulo 5. Conclusões	47
Referências Bibliográficas	51

Lista de Tabelas

3.1	Grupo de lojas selecionadas para o estudo	17
3.2	Variáveis do <i>dataset</i> de tráfego mensal	19
3.3	Variáveis do <i>dataset</i> de tráfego horário	19
4.1	Resultados da previsão de tráfego mensal obtida para cada loja	32
4.2	Resultados das experiências SARIMA	37
4.3	Resultados das experiências <i>Random Forest</i>	38
4.4	Resultados das experiências <i>Exponential Smoothing</i>	39
4.5	Resultados das experiências Prophet	40
4.6	Resultados das experiências XGBoost	41
4.7	Resultados das experiências LSTM (direta)	42
4.8	Resultados das experiências LSTM (direta)	43
4.9	Resultados das experiências <i>Ensemble</i>	45
4.10	Modelos e resultados finais das várias experiências	46

Lista de Figuras

2.1	Número de documentos submetidos em cada década	5
2.2	Países com mais de 3 documentos submetidos na pesquisa	5
3.1	Organização lógica do processo de estudo	13
3.2	Correlação entre as três componentes da solução PCA	16
3.3	<i>Clusters</i> resultantes da análise de indicadores de lojas	16
3.4	Tráfego total na rede de lojas desde janeiro de 2019	20
3.5	Diferentes tipos de ajuste testados no tráfego (Loja E)	21
3.6	Método de validação cruzada para séries temporais	22
4.1	Tráfego mensal da loja C com ajuste aplicado	31
4.2	Tráfego mensal da loja E com ajuste aplicado	32
4.3	Performance do modelo SARIMA no conjunto de teste, para valores agregados da rede MEO	33
4.4	Previsão do modelo Naive Forecast na semana 8 do conjunto de teste	34
4.5	Previsão do modelo SeasWA (diário) na semana 2 do conjunto de teste	35
4.6	Previsão do modelo SeasWA (semanal) na semana 2 do conjunto de teste	35
4.7	Rolling mean dos dados da Loja Alfa	36
4.8	Autocorrelação dos dados da Loja Alfa	36
4.9	Previsão do modelo SARIMA na semana 1 do conjunto de teste	37
4.10	Previsão do modelo Random Forest na semana 6 do conjunto de teste	38
4.11	Previsão do modelo Exponential Smoothing na semana 4 do conjunto de teste	39
4.12	Previsão do modelo Prophet na semana 7 do conjunto de teste	40
4.13	Previsão do modelo <i>XGBoost</i> na semana 5 do conjunto de teste	41
4.14	<i>Loss function de treino e validação de uma rede LSTM (direta)</i>	43
4.15	<i>Loss function de treino e validação de uma rede LSTM (iterativa)</i>	44
4.16	Previsão do modelo <i>LSTM</i> (iterativa) na semana 8 do conjunto de teste	44
4.17	Previsão do modelo <i>Ensemble</i> na semana 3 do conjunto de teste	45
5.1	Previsão de tráfego mensal da rede de lojas MEO para 2023	48

CAPÍTULO 1

Introdução

Nos últimos anos, o canal de vendas online tem ganho bastante preponderância (Fisher & Raman, 2018), face ao canal físico. Esta é uma realidade agravada pela recente situação pandémica, que acelerou em muito o desenvolvimento e adaptação do cliente típico a este meio de contacto com o vendedor. No entanto, o que continua como um dos pontos de diferenciação entre o canal digital e o canal físico, aqui entendido como lojas de retalho físico, é precisamente o contacto humano, e toda a confiança e personalização que advém desse contacto (Lee et al., 2021). É, portanto, imperativo que as empresas conheçam e controlem cada vez melhor os fatores que fazem com que esta vantagem possa ser inteiramente potenciada, entenda-se esse grupo de fatores como o espaço de loja, o número de pessoas que nela entra, a equipa de loja que está disponível para as receber, as ofertas comerciais disponíveis, entre outras.

De entre os vários fatores, a presente dissertação vai focar-se no tráfego das lojas de retalho físico. Este é um dos indicadores mais importantes para qualquer empresa que tenha lojas físicas, uma vez que é a presença de pessoas na loja que, à partida, gera mais vendas, adesões a serviços e, conseqüentemente, uma melhor taxa de conversão e fidelização dos clientes (Perdikaki et al., 2012). Quando falamos de tráfego de loja, entendemos todas as pessoas que entram em loja com um qualquer intuito relacionado com o negócio da empresa, tenham ou não intenção de comprar um produto ou aderir a um serviço.

Um dos pontos importantes a ter em conta no começo deste estudo, é precisamente o contexto de retalho que o dirige. Este trabalho é levado a cabo com o contributo da Altice Portugal, e todos estes desenvolvimentos são pensados com o intuito de ser aplicados na rede de lojas MEO, ou seja, o contexto que abraça todo este estudo não é só o de retalho físico mas, em específico, o retalho físico do setor das telecomunicações, contendo, no entanto, todas as vertentes comerciais disponíveis nas suas lojas, incluindo a venda de serviços e equipamentos, não deixando para trás a vertente de *customer care*. Posto isto, apesar do presente estudo se enquadrar de forma geral no mercado das telecomunicações, o seu foco são as lojas MEO.

É precisamente devido à presença desta vertente de *customer care* que esta dissertação se foca no tráfego e não, por exemplo, em vendas como grande parte da investigação na área. Em média, segundo o relatório de 2022 da rede de lojas MEO, perto de 70% das interações dos clientes numa loja MEO são de *customer care* e pós-venda, ou seja, ainda que seja de alta relevância trabalhar a previsão de vendas, o contexto explicado torna ainda mais relevante trabalhar a previsão de tráfego.

É esse mesmo o propósito da investigação levada a cabo nesta dissertação, a previsão de tráfego das lojas MEO. Monitorizar o tráfego de uma loja e entender o que o faz variar é fundamental para as equipas de gestão de uma empresa. Entender quais os períodos de maior afluência, quais os períodos com maior potencial de vendas, bem como os períodos de menor tráfego, leva a que possa existir uma melhor gestão dos recursos de loja. Colocar os melhores vendedores nas horas mais prometedoras, colocar mais colaboradores em horas de maior afluência de forma a poder efetuar um melhor atendimento e, conseqüentemente, melhorar a experiência para todos os clientes.

Há vários anos que as variáveis que impactam no retalho são estudadas com o intuito de fornecer melhores indicadores aos vários níveis de gestão. A partir de certo momento, estas variáveis passaram a ser estudadas com todo um novo nível de complexidade, a variável tempo (Bradlow et al., 2017). Esta veio permitir que as análises por intervalos passassem a análises de dados contínuos e com essa nova realidade vieram também novas ferramentas como séries temporais. Os modelos de séries temporais permitem o estudo dos valores de histórico de forma a conseguir prever os valores do futuro.

O presente estudo utiliza o que é feito atualmente na rede de lojas MEO, no que toca a técnicas de previsão e organização de equipas, e utiliza-o como *baseline* sobre a qual trabalhar evolutivamente, com o objetivo de chegar ao fim das experiências com uma ferramenta mais robusta e capaz de munir as equipas de gestão. Pretende-se que sejam produzidas previsões para duas periodicidades de tráfego diferentes. A primeira pretende-se que produza previsões para os próximos 24 meses, ou seja, 2 anos de valores mês a mês, enquanto a segunda abordagem fica encarregue de produzir previsões hora a hora para a próxima semana, ou seja, um horizonte temporal de 168 períodos de 1 hora. Desta forma entende-se que o estudo fornece ferramentas para uma análise mais detalhada quer no ponto de vista macro quer no ponto de vista micro.

A estrutura da presente dissertação está dividida em quatro capítulos, organizados para que todo o processo seja explicado da forma mais organizada e clara possível. No próximo capítulo, Revisão de Literatura, é criada uma base de conhecimento teórico que retrate o estado-da-arte da área científica que abraça esta dissertação, no que toca a áreas de aplicação e técnicas utilizadas. De seguida, o capítulo Metodologia descreve todo o procedimento realizado durante o presente estudo, bem como as diferentes experiências realizadas, caracterizando os processos do estudo do tráfego mensal e tráfego horário em momentos diferentes. O capítulo Resultados e Discussão acompanha o capítulo Metodologia e apresenta todos os resultados das diferentes experiências, refletindo criticamente a forma como cada um deles contribuiu para o desenvolvimento e evolução da presente dissertação. Por fim, no capítulo Conclusões é feita uma revisão a todo o processo com o objetivo de apurar a importância dos resultados obtidos, pertinência e utilidade dos mesmos. É ainda feita uma retrospectiva do que foram as principais dificuldades e limitações sentidas durante todo o processo, bem como as áreas do estudo que deixam em aberto possíveis linhas de investigação para seguir no futuro.

CAPÍTULO 2

Revisão da Literatura

O capítulo de revisão de literatura serve de base ao presente estudo que se centra na previsão de tráfego das lojas MEO. Esta revisão de literatura está dividida em duas partes. A primeira parte, relacionada com as áreas de aplicação, que procura agregar as bases mais racionais daquilo que tem sido o desenvolvimento e integração das áreas das tecnologias e do retalho, nomeadamente, no sentido da aplicação das tecnologias de previsão naquilo que são os vários processos de gestão de uma loja de retalho. E a segunda parte, relacionada com as técnicas, procura focar-se essencialmente naquilo que é o estado de arte dos algoritmos de previsão e sua avaliação, neste caso, essencialmente aplicados a séries temporais.

Esta base de conhecimento procura dar resposta a cinco perguntas que podem resumir brevemente os vários pontos tocados pelo estudo que se pretende levar a cabo, colocando as três primeiras como bases da pesquisa para as áreas de aplicação e as restantes duas como base da pesquisa sobre as técnicas:

- (1) Como tem vindo o setor do retalho a evoluir no que diz respeito à implementação de técnicas de inteligência artificial?
- (2) Como é que a implementação de técnicas de inteligência artificial tem vindo a impactar a gestão de equipas no retalho?
- (3) Como vem a evoluir o racional da aplicação de técnicas de *time series forecasting*?
- (4) Quais os algoritmos mais utilizados?
- (5) Quais as métricas mais utilizadas para avaliação individual dos algoritmos e quais as melhores para comparação entre os mesmos?

2.1. Critérios de Pesquisa

O processo iniciou-se pela pesquisa de artigos de literatura científica que pudessem servir de base às várias nuances do estudo, salientando, em particular, duas direções distintas. A primeira direção conduziu uma pesquisa de cariz mais teórico, ou seja, que analisou artigos com uma abordagem mais teórica a temas como a evolução do setor do retalho físico, a utilização de inteligência artificial, nomeadamente *machine learning*, nos vários processos e indicadores de gestão de lojas físicas e ainda na evolução das utilizações de séries temporais, aqui com maior foco no seu contexto teórico, em termos de pré-processamento de dados, *features*, variáveis exógenas, entre outros. A segunda direção conduziu uma pesquisa bem mais focada na vertente das técnicas do tema abordado no presente estudo, isto é, artigos que se focassem mais na utilização prática dos vários algoritmos de *machine learning* para prever a procura de lojas físicas, tome esta procura

a forma de tráfego, ou de qualquer outro indicador que possa ter um comportamento semelhante ao tráfego. Este tipo de pesquisa permitiu ter uma ideia do panorama geral da evolução e forma de atuação dos diferentes algoritmos, tentando construir um padrão de contextos em que uns algoritmos apresentam melhores performances que outros. Foram utilizadas as ferramentas Google Scholar (GS) e, essencialmente, o Scopus. O GS foi utilizado numa ótica mais generalista, enquanto o Scopus foi utilizado de forma mais focada em artigos científicos, tendo sido a análise centrada essencialmente nos resultados das *queries* nele aplicadas.

Para a pesquisa no Scopus foram utilizadas as seguintes *queries*:

TITLE-ABS-KEY(("traffic" AND "retail") OR "store traffic" AND ("time series" OR "forecasting" OR "forecast" OR "prediction" OR "machine learning")), que, à data de 29 de dezembro de 2022, resultou num total de 147 documentos.

TITLE-ABS-KEY("retail" AND "management" AND ("machine learning" OR "time series" OR "forecast" OR "forecasting")), que depois de filtrada para manter apenas journal articles, das áreas de "Business, Management and Accounting", "Engineering", "Computer Science", "Decision Sciences", "Economics, Econometrics and Finance" e "Social Sciences", resultou, à data de 29 de dezembro de 2022, num total de 330 documentos.

Estes documentos foram analisados à luz do objetivo do presente estudo e enquadrados entre a vertente "áreas de aplicação" e vertente "técnicas" já antes mencionadas. É importante realçar que foi analisado o contexto e conteúdo de cada um dos documentos, para aferir a utilidade de uma análise mais detalhada ao seu contributo para a presente revisão de literatura.

De entre os documentos considerados úteis para a pesquisa das áreas de aplicação, foi feita uma análise segundo três prismas diferentes, Retalho, Gestão de Equipas e Séries Temporais, cada um deles focado em interesses diferentes para o presente estudo. Já nos documentos considerados úteis para a vertente das técnicas, foi importante identificar o use case a que o artigo apresentava resposta, os modelos utilizados e a forma como os mesmos foram utilizados e ainda as métricas, não só de avaliação individual dos modelos, mas também de eventual comparação entre eles.

Dar nota ainda de que alguns destes documentos deram acesso a outras fontes de informação consideradas de utilidade para a presente revisão de literatura, através das suas referências bibliográficas.

2.2. Artigos Analisados

Primeiramente, analisando a figura 2.1, foi possível verificar que a temática estudada está em claro ascendente no que toca a interesse e investigação. Ainda que tenham sido analisados apenas três anos na presente década, o número de artigos presentes nos resultados da *query* submetidos de 2020 em diante é superior ao de todas as décadas, com exceção da década de 2010, no entanto, extrapolando a sua dimensão, é de fácil conclusão que no fim da presente década, o número de artigos publicados será ainda mais significativo.

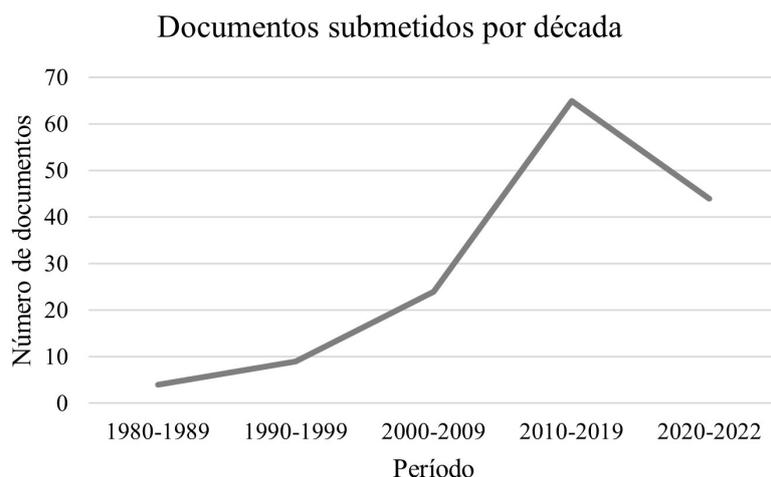


FIGURA 2.1. Número de documentos submetidos em cada década

No que diz respeito à origem dos documentos, foi interessante ver que grande parte dos documentos são originários dos Estados Unidos, país que contém também grande parte das gigantes empresas tecnológicas consideradas como as mais avançadas no mercado no que diz respeito a desenvolvimento e utilização de algoritmos de inteligência artificial. Observando a figura 2.2, é possível ver que China e Reino Unido ocupam o segundo e terceiro lugar, respetivamente, e somam em conjunto 20 documentos dos resultados obtidos.

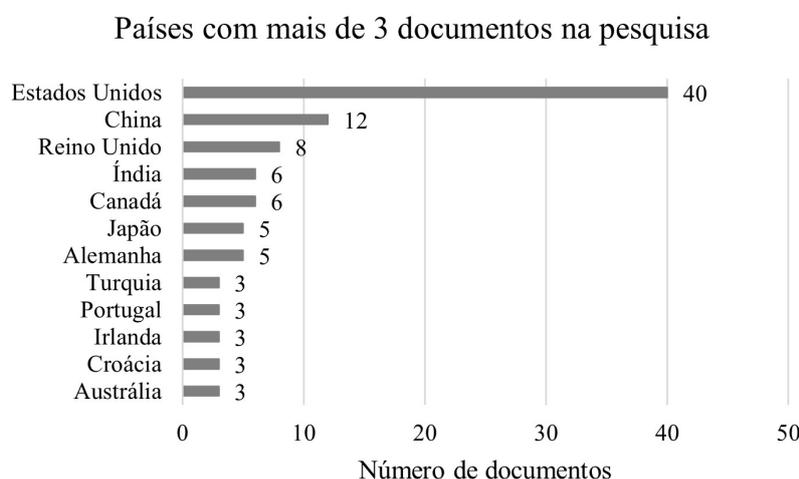


FIGURA 2.2. Países com mais de 3 documentos submetidos na pesquisa

Estes resultados fazem sentido uma vez que falamos de países com alto desenvolvimento em empresas ligadas ao retalho, ligadas às tecnologias de inteligência artificial e, acima de tudo, com uma visão integrada de ambos os campos de investigação. Esta é uma realidade que cada vez mais se espalha aos vários cantos do mundo e isso é refletido pela distribuição dos restantes documentos apresentados pela *query* do Scopus. Nota ainda

para o posicionamento de Portugal, com três artigos publicados, o que faz jus ao recente investimento tecnológico no país, no que toca à educação e investigação da área.

2.3. Áreas de aplicação

O tema de áreas de aplicação deste estudo pretende criar uma base sólida de conhecimento no que diz respeito às áreas abordadas no projeto a desenvolver, nomeadamente, Retalho, Gestão de Equipas e Séries Temporais. A próxima secção analisa as questões práticas do tema, mas esta secção pretende essencialmente entender como se têm vindo a desenvolver estas áreas, no que toca às abordagens utilizadas, ferramentas e boas práticas. Como é natural, esta análise foi feita à luz das aplicações de técnicas de inteligência artificial e desenvolvimento de métodos de previsão.

Para conduzir esta busca de conhecimento, foram levantadas três questões, já apresentadas anteriormente, que serão agora abordadas de forma individual.

- (1) *Como tem vindo o setor do retalho a evoluir no que diz respeito à implementação de técnicas de inteligência artificial?*

Cada vez mais, é dada relevância extrema ao entender do comportamento do consumidor. É este o ativo mais importante para qualquer empresa e, no contexto de loja, esta premissa ganha ainda mais significado. Por sua vez, entender o comportamento do consumidor envolve entender o comportamento do fluxo de clientes em loja. Este é um dos pontos fundamentais para o bom funcionamento de uma equipa de loja e, ainda mais, para um bom atendimento. Segundo (Perdikaki et al., 2012), o tráfego de loja está altamente correlacionado com a sua taxa de conversão e com o volume de vendas alcançado. Facilmente vemos estas relações como sendo retas crescentes, que traduzem em vendas o tráfego e conseqüentemente a taxa de conversão geral da loja. No entanto, isto não se verifica. Segundo esse mesmo estudo, existe uma relação crescente até dado ponto, ponto esse a partir do qual a relação começa a ser negativa, ou seja, a partir de determinada quantidade de clientes em loja, as vendas começam a não aumentar como antes e a taxa de conversão começa a reduzir. De acordo com (Lee et al., 2021), isto deve-se ao facto do maior número de pessoas acabar por ocupar por mais tempo os colaboradores de loja, fazendo com que a experiência para os restantes clientes, em média, acabe por piorar. Estes estudos afirmam ainda que o valor médio do “cesto de compras” que cada cliente acaba a comprar também é negativamente afetado pelo tráfego a partir de certo número. Este raciocínio faz sentido segundo (Perdikaki et al., 2012), uma vez que o cliente só está disposto a esperar o tempo proporcional ao valor que dá à compra que está prestes a fazer, ou seja, com mais clientes em loja, o tempo de espera tendencialmente aumenta, o que faz também aumentar a necessidade de valor dada ao produto/serviço que leva o cliente a estar à espera.

As empresas alinham-se então na necessidade de conseguir prever os maiores fluxos de clientes, mas as abordagens tendem a seguir o racional de vendas e não tráfego em si. (Lam et al., 1998) apresenta uma abordagem baseada exclusivamente em vendas para

tentar organizar as equipas de loja. Esta é uma teoria que pretende dar resposta ao dilema constante entre a quantidade de recursos humanos a disponibilizar em loja face à procura, neste caso quantificada em termos de vendas. Na teoria, o pretendido seria conseguir definir uma utilidade marginal para os colaboradores de loja que pudesse definir com clareza o ponto ótimo para a organização dos turnos. O maior desafio nesta teoria é a volatilidade da procura. O retalho é precisamente caracterizado pelas voláteis ondas de clientes que estão em loja, e segundo (Chuang et al., 2016), é precisamente essa volatilidade na procura que complica a organização dos turnos e afeta a qualidade do serviço prestado em última instância. Ainda assim, grande parte dos desenvolvimentos na área do retalho, no que toca ao encontro entre recursos e procura tem sido feito precisamente com base em vendas.

Um outro tema que há muito desperta interesse, é o das filas de espera. Uma questão altamente relacionada com o tráfego e que, com o crescimento do foco na experiência do cliente, tem vindo a ganhar cada vez mais preponderância. De acordo com (Larson, 1987), o sentimento de injustiça que paira na cabeça do consumidor aquando da sua espera para ser atendido ou finalizar a sua compra, faz com que este esteja num constante momento de tensão. Esta é a natureza das filas de espera. No entanto, seguindo a linha de investigação de Larson, este *mindset* pode ser alterado, uma vez que a experiência do consumidor fica bem mais positiva quando a espera é proporcionada com elementos de dinamização, tais como dinâmicas comerciais ou informação sobre a duração média da espera. O tempo ocupado passa, em termos de perceção, de forma mais rápida que o tempo desocupado.

(2) *Como é que a implementação de técnicas de inteligência artificial tem vindo a impactar a gestão de equipas no retalho?*

No seguimento da pergunta anterior, importa entender como tem vindo a mudar a gestão de equipas com a crescente preocupação com o serviço prestado de forma dirigida pelos dados. A forma mais eficaz de conseguir planear a gestão de turnos e adaptar os recursos alocados a cada turno é precisamente poder trabalhar com previsão de procura, seja ela materializada em termos de vendas ou tráfego. Estes indicadores, que com a complexidade da variável tempo passaram a ser abordados de melhor forma por séries temporais (Bradlow et al., 2017), são fundamentais para o bom funcionamento das equipas de gestão. De acordo com o estudo de (Fildes et al., 2022), as previsões devem ser vistas hoje, ao nível organizacional, como inputs importantíssimos para as melhores decisões não só da gestão de equipas de loja, mas também de atividades como marketing, produção e desenvolvimento e até finanças.

Estas previsões permitem então subir o nível de organização dos turnos das equipas. Esta alta capacidade de organização permite aos gestores criar horários mais estáveis e produtivos para as suas equipas. Esta matéria está por sua vez, altamente relacionada com a performance das próprias equipas. A consistência dos turnos de um colaborador, primeiro ao nível de dias da semana e depois ao nível de horas do dia (Lu et al., 2022), pode aumentar significativamente os seus níveis de produtividade. De acordo com o

mesmo estudo, esta consistência vai ter ainda mais impacto nos casos de colaboradores recentemente admitidos.

Uma outra questão a ter em conta na gestão dos turnos das equipas, são os fatores que podem influenciar a sua flutuação. A investigação na área tem sido levada por um caminho determinístico (Van Den Bergh et al., 2013), ou seja, os métodos de organização das equipas são guiados de forma tendencialmente rígida, não tendo agilidade suficiente para fazer frente aos picos de clientes que se fazem sentir segundo fatores mais “secundários”. A organização das equipas assenta muitas vezes em indicadores fixos, deixando de parte fatores como meteorologia e épocas festivas de menor dimensão (Abrishami et al., 2017).

(3) *Como vem a evoluir o racional da aplicação de técnicas de time series forecasting?*

Os métodos de previsão só fazem sentido assim que se perceber onde estes vão encaixar e impactar. Segundo o estudo (Seaman, 2018), apesar de ser importante ter todo o conhecimento necessário ao desenvolvimento do modelo de previsão, é fulcral que haja uma etapa de *business understanding*, que permita entender as reais necessidades do negócio.

Assim que se entenda a necessidade a que se pretende dar resposta com os modelos de séries temporais, é importante entender quais os fatores que mais influenciam o nosso *target*. Estes fatores podem ser diretamente deduzidos ou podem levar um pouco de tempo a ser revelados. O estudo (Abrishami et al., 2017), toma em atenção variáveis exógenas como a meteorologia, eventos especiais e feriados locais. Estes são fatores que, ainda que não sejam diretamente observados nos resultados do negócio, fazem flutuar e muito o tráfego de lojas. Faz, portanto, sentido que as séries temporais sejam pensadas com os mesmos em mente.

Outro ponto bastante importante na evolução das séries temporais aplicadas ao elemento de estudo deste trabalho, é a possibilidade de trabalhar várias séries temporais de forma conjunta. O estudo levado a cabo por (Abrishami E Kumar, 2019) utiliza dados recolhidos de 65 lojas e acaba por utilizar uma abordagem de *K-Nearest Neighbour*, de forma a utilizar dados de múltiplas lojas semelhantes entre si. Este raciocínio tinha sido referido também em (Seaman, 2018), mas no que toca a métricas de avaliação, levantando a possibilidade de analisar *forecasts* de várias lojas com uma única avaliação conjunta, baseada nos pesos da procura de cada uma das lojas.

Um dos aspetos mais importantes no projeto a desenvolver é o tema do *multi-step-ahead*. Por *multi-step-ahead* entende-se o número de “saltos” dados além do último valor real, ou seja, qual o horizonte temporal a prever pelo modelo. Este é sem dúvida um dos maiores temas de investigação em aberto na área das previsões de séries temporais. Em (An E Anh, 2015), são apresentadas cinco estratégias de *multi-step-ahead*; *iterated strategy*, *direct strategy*, *DirREC strategy*, que junta as duas anteriores, *Multi Input Multi Output* (MIMO) e ainda *DirMIMO*, que junta a MIMO *strategy* com a *direct strategy*. Nesse estudo, após comparações entre as cinco estratégias, a *DirREC* é apresentada como

a estratégia mais robusta. Já no estudo (Sousa et al., 2022), na secção de revisão de literatura, é abordado o tema do *multi-step-ahead* como sendo um tema sem vencedor, ou seja, apesar de haver vantagens nalgumas estratégias, estas acabam por depender bastante do conjunto de dados onde estão a ser aplicadas, acabando por não haver uma estratégia significativamente melhor que as restantes. O paradigma, no entanto, altera-se quando também no ano de 2022, o estudo (Livieris E Pintelas, 2022), apresenta uma nova ideia para a abordagem ao *multi-step-ahead*, introduzindo um novo modelo de *multi-step deep learning*. Este apresenta resultados tangivelmente melhores que os modelos anteriormente apresentados, que foram também utilizados neste estudo como método de comparação.

2.4. Técnicas

Fechado o tema das áreas de aplicação, importa agora fazer o apuramento das técnicas mais recorrentes na investigação internacional, no que toca aos modelos de *time series* e suas métricas de avaliação, individual e de comparação.

Para tal, serão abordadas individualmente as restantes duas questões apresentadas na introdução.

(4) *Quais os algoritmos mais utilizados?*

Em termos de modelos e abordagens analisadas nos vários documentos lidos, foi possível traçar alguns padrões de relevo.

O estudo (Abrishami et al., 2017) apresenta uma abordagem pelo *machine learning*, deixando de lado a análise de séries temporais, apostando nos algoritmos de *random forest regression* e *support vector regression*, como forma de combater os obstáculos do *multi-step-ahead*. Neste estudo foram analisadas várias lojas de vários tipos de setores comerciais, obtendo valores de MAPE entre os 13% (ginásios) e os 24% (restauração).

As opções mais triviais também podem apresentar resultados satisfatórios. Casos como *Seasonal Naïve*, que consiste em dar o valor ao *forecast* do último período de sazonalidade idêntico, ou o *Multiple Seasonal Naïve*, que segue a mesma linha de raciocínio, mas usando vários períodos passados, calculando depois a sua média. Este tipo de abordagens pode apresentar bons resultados em casos com alta sazonalidade e tendências nulas (Sousa et al., 2022). Estes modelos apresentam ainda características que os tornam muito úteis para modelos *baseline* de comparação.

No que diz respeito a modelos híbridos, o estudo (Abrishami & Kumar, 2019), apresentou os seus melhores resultados (MAPE 17,06%) com utilização de um modelo híbrido que junta *Nearest Neighbour* e *Random Forest* para previsão de tráfego hora a hora em lojas de vários setores comerciais. É uma abordagem interessante, uma vez que o modelo com base no *Nearest Neighbour* atua com o objetivo de encontrar a loja mais semelhante à que está a ser prevista, de forma a serem usados os dados agregados das duas lojas para treinar o modelo *Random Forest*. Esta é uma tendência também descrita nos principais destaques da competição M4, (Makridakis et al., 2020), onde foi dado realce aos benefícios do uso de informação de várias séries temporais para melhorar a performance individual

de cada uma delas. Outras abordagens híbridas foram ainda apresentadas em (Cortez et al., 2017), utilizando o modelo de alisamento Holt-Winters com SVM e ainda uma com dois modelos *Support Vector Machine* (SVM) a compor o modelo híbrido.

Dentro dos modelos desenvolvidos pelas equipas das grandes empresas tecnológicas, destaque para o modelo Prophet, que aparece referido em inúmeros artigos analisados, (Abrishami & Kumar, 2019; Junior et al., 2021; Sousa et al., 2022). Este modelo apresenta melhores resultados quando aplicado a séries temporais com alto nível sazonalidade e sazonalidades múltiplas (vários períodos temporais em simultâneo).

De entre os modelos mais tradicionais, destaque para a família dos modelos ARIMA, presente em grande parte dos estudos analisados, sendo como foco principal do estudo de tráfego pedestre na cidade de Melbourne, (Wang et al., 2017), ou apenas como comparação *baseline*.

No que toca a redes neuronais, destaque para os modelos *Long Short Term Memory* (LSTM), *Multi Layer Perceptron*, BiLSTM, que consiste numa variação bi direcionada do modelo LSTM, e ainda para os modelos de redes neuronais convulsionais (CNN) (Chandra et al., 2021). Dar nota ainda de que apesar de estes serem modelos que requerem alta capacidade computacional, segundo os destaques da competição M4 (Makridakis et al., 2020), os modelos de alta complexidade têm mostrado resultados crescentes, face aos modelos menos complexos.

De entre os vários artigos analisados, um dos pontos mais unânimes e interessantes são os modelos de *ensemble*. Esta abordagem consiste na utilização de vários *forecasts* e na atribuição de pesos a cada uma delas, tentando chegar dessa forma a um modelo mais completo e robusto (Petropoulos et al., 2022). Este tipo de abordagens já remonta ao século passado (Bates & Granger, 1969; Clemen, 1989), e trazem grandes vantagens aos modelos atuais. Todos os modelos top performance da competição M4 (Makridakis et al., 2020), foram resultados de combinações. O vencedor da competição M5 (Makridakis et al., 2022), foi resultado de uma combinação de 6 modelos. Também os estudos (Adhikari & Agrawal, 2012; Kumar & Anitha, 2022; Sousa et al., 2022), apresentam resultados bastante robustos utilizando abordagens de *ensemble*. Esta é uma abordagem que suscita grande interesse para o projeto a desenvolver uma vez que o objetivo passa por dar resposta à necessidade de previsão de tráfego para lojas com diferentes contextos, que poderão requerer abordagens diferentes.

(5) *Quais as métricas mais utilizadas para avaliação individual dos algoritmos e quais as melhores para comparação entre os mesmos?*

As métricas mais utilizadas nos artigos científicos analisados não fogem do comum. As métricas dependentes de escala englobam o *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE) e, para fugir à dependência de escala, tornando-se melhor para comparação entre avaliações de *forecasts*, o *Mean Absolute Percentage Error* (MAPE), que devolve valores em percentagem. No entanto, há que ter cuidado com estes valores, uma vez que também eles têm as suas desvantagens. Os erros em percentagem não

lidam bem quando a variável alvo assume valores pertos de 0 (Maçaira E Cyrino Oliveira, 2016). Segundo este mesmo artigo, para além das métricas dependentes de escala e das métricas percentuais, existem ainda as métricas *relative-error*, e as métricas independentes de escala, ou seja, métricas que são normalmente divididas por uma base referente ao caso particular a ser trabalhado. O estudo (Sousa et al., 2022) utiliza a métrica *Relative MAE*, que faz a relação entre o MAE obtido por um dado modelo e o MAE obtido por um modelo considerado baseline *para* o estudo. Esta é uma técnica interessante uma vez que cria um sentido evolutivo comparado a um modelo considerado inicial, apresentando ainda benefícios face às questões de dependência de escala e limitações das métricas de percentagem para valores perto de 0.

Para além destas métricas, o estudo (Seaman, 2018), refere ainda a métrica *Weighted MAPE*, que consiste num MAPE conjunto, neste caso utilizado para unir métricas de várias lojas, dando pesos relativos face ao peso das suas vendas. Ou seja, em vez ser utilizado um MAPE para cada *forecast* individual por loja, é obtida uma métrica geral de performance do modelo para o total das lojas. Esta é também uma abordagem interessante tendo em conta o pretendido no projeto a desenvolver com as lojas MEO.

CAPÍTULO 3

Metodologia

Por se tratar de um projeto na área de ciência de dados, as várias etapas deste estudo foram guiadas pelo conhecido *Cross-Industry Standard Process for Data Mining* (CRISP-DM). O CRISP-DM é hoje uma das metodologias mais utilizadas na indústria e ajuda a conduzir todo os diferentes tipos de projetos na área de ciência de dados. A sua estrutura incentiva o processo cíclico e está atualmente dividida em seis principais etapas (Shearer, 2000), que são seguidas no decorrer deste documento: *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation e Deployment*.

O fluxograma 3.1 resume os passos e etapas do presente estudo, que serão de seguida explorados.

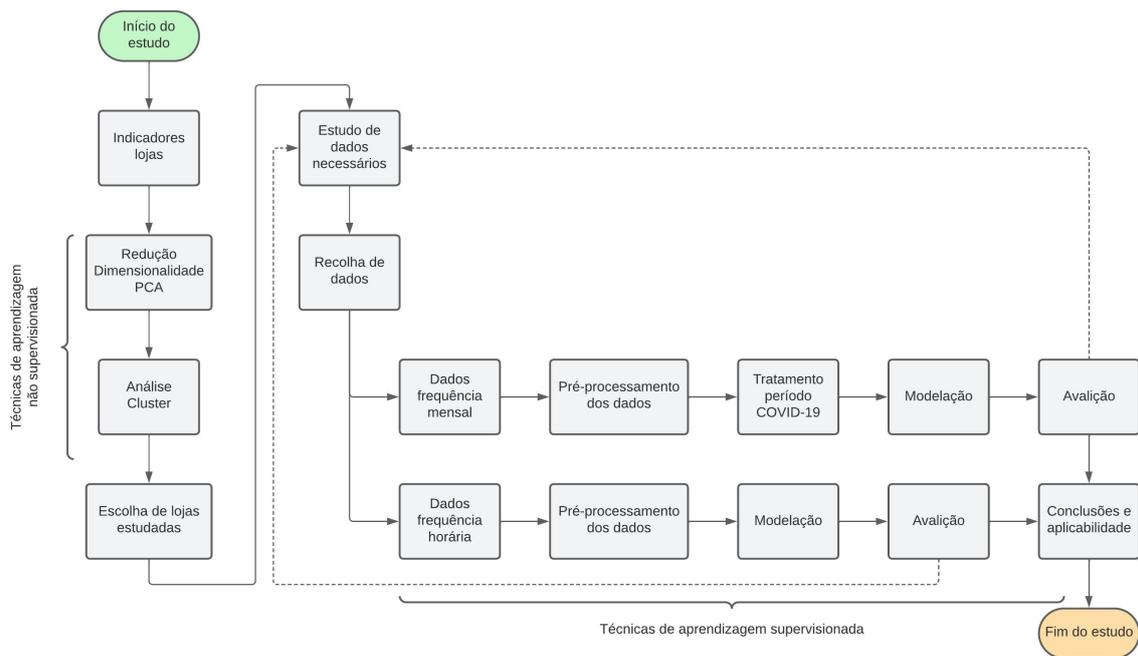


FIGURA 3.1. Organização lógica do processo de estudo

3.1. *Business Understanding*

Provavelmente a etapa mais importante do estudo, o momento de *Business Understanding* é fundamental para o bom decorrer do mesmo. Nesta fase são formuladas as questões a que se pretende obter resposta e são definidos os objetivos que determinam o sucesso da investigação.

No caso do presente estudo, a necessidade centra-se em fornecer ferramentas que possam alavancar a gestão das Lojas MEO através da previsão de tráfego. Este é um tema que pode tomar várias direções. Na revisão de literatura foi possível encontrar vários estudos que utilizam vários indicadores para formar o problema de previsão de procura. O contexto da rede de lojas MEO implica uma análise ligeiramente diferente. Ainda que possa fazer sentido olhar para a procura como vendas, os estudos (Lee et al., 2021) dizem que a relação entre o número de pessoas que entram em loja e o número de pessoas que formam os recursos humanos de uma loja está altamente ligado à taxa de conversão obtida, ou seja, uma má gestão desse rácio leva a que haja menos vendas, o que por sua vez nos leva a pensar que aprender a prever tráfego através de vendas pode ser uma visão em si já enviesada pelas boas ou más performances anteriores. A isto junta-se um facto muito característico do setor das telecomunicações, segundo o relatório de 2022 das Lojas MEO, perto de 70% do tráfego que entrou nas lojas MEO não foi com o intuito de comprar um produto ou subscrever um serviço, mas sim para tratar de questões de *customer care* e pós-venda. Situações como pagamento e revisão de faturas, alteração de condições contratuais, troca e reparação de equipamentos, são tudo temas recorrentes nas lojas MEO e que acabam por ocupar de forma significativa o tempo disponível das equipas para se centrar em vendas, o que vinca ainda mais o problema de utilizar as vendas como previsão de tráfego, como utilizado no estudo (Lam et al., 1998).

É então necessário olhar ao indicador que precede o momento de interação e eventual venda, o tráfego em bruto. Entenda-se por tráfego a quantidade de pessoas que entra numa qualquer loja, traduza-se esta pessoa numa compra ou interação com a equipa de loja ou não. Se uma pessoa entra na loja, deve ser considerado tráfego, uma vez que criou a oportunidade de abordagem e de proporcionar um serviço. É nesse indicador que, para este estudo, faz sentido alicerçar a análise.

O estudo do tráfego pode ser dividido em dois momentos. O primeiro momento que analisa o tráfego mensal das diferentes lojas, e da rede de lojas como um todo, permitindo traçar uma análise mais macro daquilo que são as tendências generalistas de tráfego. Esta análise permite gerir aquilo que são as necessidades em termos de recursos humanos para cada equipa. Um aumento da previsão de tráfego a seis, doze ou vinte quatro meses pode querer dizer que hoje é necessário repensar a dimensão de certas equipas de loja, bem como, no sentido inverso, uma diminuição dessas previsões pode levar a uma reestruturação de equipas de forma a alocar recursos em excesso a lojas que necessitem dessa capacidade. O segundo momento da análise centra-se na análise de tráfego horário, ou seja, uma ferramenta que permita uma análise mais micro ao tráfego mais específico. Com esta fase do estudo pretende-se aumentar a capacidade de gestão de equipas, a sua produtividade e satisfação (Lu et al., 2022), através de uma melhor gestão de turnos. O horizonte temporal destas previsões começa definido no valor de 24 meses, ou seja, dois anos para a previsão de tráfego mensal e nas 168 horas para a previsão de tráfego horário, que equivale à previsão de uma semana inteira hora a hora.

Estes dois momentos de estudo, entendem-se como inputs de alta importância para a capacidade das equipas de gestão, não só de loja, mas também de marketing, produto, entre outras (Fildes et al., 2022), pelo que é importante que exista também uma comparação entre o que é feito atualmente na rede de lojas no que toca a este tipo de indicadores, e seja criada uma base sobre a qual se deve trabalhar no sentido evolutivo.

3.2. Data Understanding

3.2.1. Seleção de lojas

Um dos aspetos de maior importância deste estudo foi o de entender quais as lojas que iam servir de teste para as diferentes etapas da investigação. A rede de lojas era composta, à data do início do estudo, por 179 lojas, distribuídas pelas várias zonas de Portugal continental, Madeira e Açores, todas elas com o seu enquadramento e tipo de público. Dado que o estudo envolve análise individual seria insustentável tentar olhar para todas da mesma forma. Para escolher que lojas seriam estudadas, começou por se recolher todos os indicadores de performance para todas as lojas, com o objetivo de proceder a uma análise *cluster*. O objetivo racional seria o de conseguir agrupar as lojas mediante os contextos e tipos de indicadores obtidos, para que se pudesse recolher um grupo de lojas mais reduzido, mas que, ao mesmo tempo, representasse de forma fiel todo o tipo de lojas existentes. Foi reunido um *dataset* com vinte e um indicadores para cada loja, ao qual foi feita uma análise de componentes principais (Hotelling, 1933), uma técnica amplamente utilizada para redução de dimensionalidade de bases de dados, recolhendo aquilo que são os principais padrões entre indicadores (Wold et al., 1987). Estes indicadores englobam não só valores de tráfego, como também valores de diferentes tipos de vendas (produtos e serviços), número de interações de vários tipos, trocas, assistência técnica, entre outros.

A análise de componentes principais devolveu resultados que permitiram o estudo do *dataset* com duas e três componentes. O método do cotovelo indicava uma possível solução de duas componentes, enquanto o critério de Kaiser indicava uma possível solução com três componentes. As duas hipóteses foram analisadas individualmente e contextualizadas mediante o objetivo do estudo. Ao analisar a solução de três componentes, ilustrada na figura 3.2, é possível verificar a forte correlação entre a primeira e terceira componente, o que levou a querer que a utilização de apenas duas componentes seria ideal para o estudo em causa, uma vez que a primeira e terceira componente trariam um pouco da mesma informação à análise.

Em termos de interpretação de cada uma das componentes, a Componente 1 foi interpretada como “Vendas Core”, ou seja, representando grande parte dos indicadores de vendas e prestação dos serviços considerados core de uma loja de telecomunicações, e a Componente 2 interpretada como “Troca e Recolha de Equipamentos”, agrupando os indicadores que dizem respeito a esse tipo de interações mais características de certas lojas viradas para o atendimento pós-venda.

Ao analisar visualmente o novo *dataset* das lojas composto por duas componentes, apresentado na figura 3.3, foi possível verificar a predominância de 4 grupos distintos,

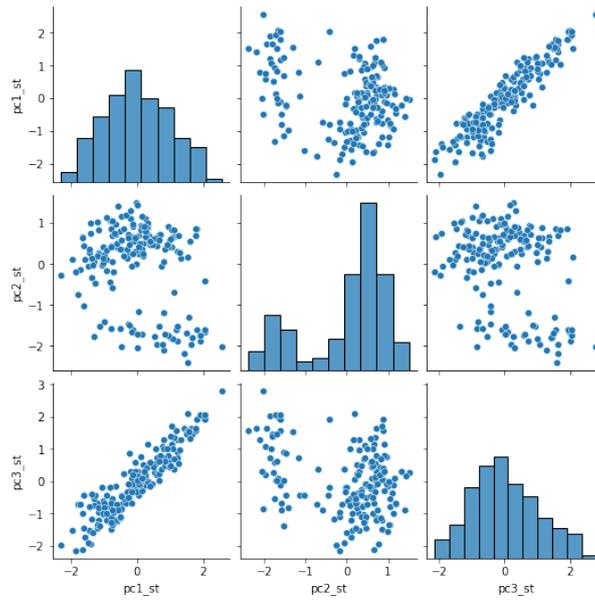


FIGURA 3.2. Correlação entre as três componentes da solução PCA

confirmados pelos métodos de *clusters* hierárquicos e *k-means*. Estes grupos reúnem um enquadramento lógico significativo uma vez que dividem as 179 lojas em quatro realidades caracterizadas por alto ou baixo valor de "Vendas Core" e alto ou baixo valor de "Recolha e Troca de Equipamentos", realidade que assenta bastante bem no conhecimento geral que existe do comportamento das Lojas MEO.

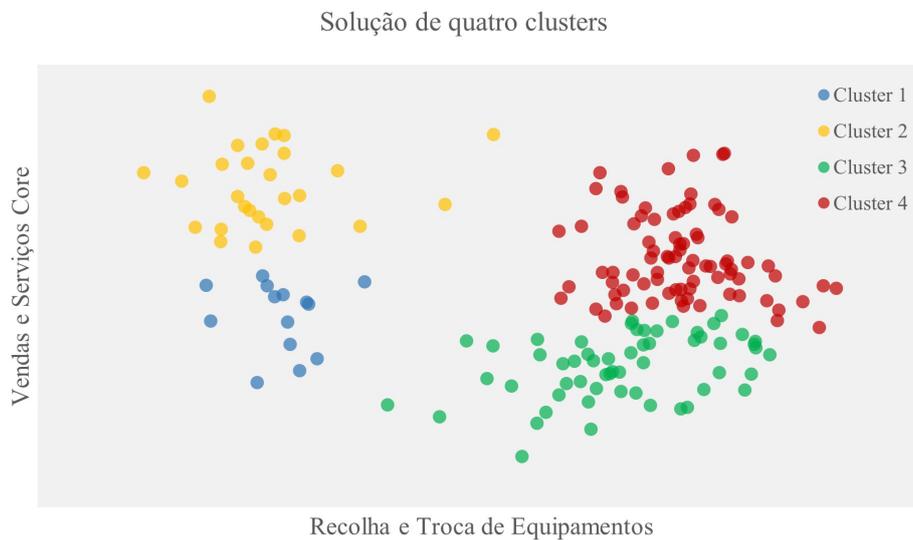


FIGURA 3.3. *Clusters* resultantes da análise de indicadores de lojas

De seguida, juntaram-se a este indicador as informações geográficas de cada loja, o tipo de *layout* das mesmas (dividido entre novo *layout* e antigo *layout*) e ainda o tipo de loja (loja de rua ou loja de centro comercial) e procedeu-se à seleção mais equilibrada entre grupos, região e tipo de loja resultando no grupo de lojas apresentado na tabela 3.1.

TABELA 3.1. Grupo de lojas selecionadas para o estudo

Loja	Município	Região	Tipo de loja	Tipo de <i>layout</i>
Loja A	Ponte de Lima	Norte	Rua	Antigo
Loja B	Matosinhos	Grande Porto	Centro Comercial	Novo
Loja C	Espinho	Centro	Rua	Antigo
Loja D	Aveiro	Centro	Centro Comercial	Novo
Loja E	Sintra	Grande Lisboa	Centro Comercial	Novo
Loja F	Lisboa	Grande Lisboa	Centro Comercial	Novo
Loja G	Lisboa	Grande Lisboa	Rua	Novo
Loja H	Beja	Sul	Rua	Novo
Loja I	Portimão	Algarve	Centro Comercial	Novo
Loja J	Tavira	Algarve	Centro Comercial	Antigo

Esta tabela de lojas serviu de base para o estudo do tráfego mensal, procurando obter previsão para cada uma das lojas, para além de procurar uma previsão de tráfego agregado da rede.

Em relação ao tráfego hora a hora, existe um ponto muito importante a ter em conta. A organização de equipas é essencial em qualquer loja, como já falado anteriormente, no entanto esta ganha outra relevância, na ótica da utilização deste tipo de ferramentas, quando trabalhamos com equipas maiores. Posto isto, este desenvolvimento será muito mais útil numa equipa de loja com 20/30 colaboradores, quando comparado com uma equipa de loja com 2 ou 3 colaboradores, pois esta última não tem praticamente margem de manobra no que toca a gestão de turnos face ao tráfego. Dado que estas duas realidades existem na rede de lojas MEO e, regra geral, são aplicadas a muitas redes de lojas de retalho de outras marcas, faz sentido que o estudo se centre numa loja teste com muitos recursos à disposição para trabalhar. Foi por essa mesma razão que o estudo do tráfego hora a hora foi desenvolvido com uso dos dados de uma das maiores lojas da rede MEO, daqui em diante, nomeada de Loja Alfa (por questões de confidencialidade dos dados disponibilizados ao longo do estudo). Esta loja tem perto de 30 colaboradores, situa-se num centro comercial da Grande Lisboa e é uma das lojas que mais tráfego recebe. Para além disso tem a particularidade de se situar perto de um local de grandes eventos desportivos e culturais, que podem impactar o decorrer normal das entradas e saídas de clientes em loja.

3.2.2. Recolha de dados

Sem dados de qualidade, a própria qualidade do estudo fica comprometida logo à partida. Foi, portanto, muito importante entender com que dados trabalhar e de que forma os mesmos estavam a ser recolhidos.

O tráfego das lojas foi recolhido com utilização de dispositivos colocados na entrada de cada uma das lojas. Estes dispositivos fazem a contagem de pessoas que entram através de um par de lasers e devolve valores para intervalos de 15 minutos durante o decorrer de todo o dia. Estes dados são depois agregados mediante a necessidade do estudo. No caso

do estudo do tráfego mensal, são agregados para períodos de 1 mês, e no caso do tráfego horário são agregados em períodos de 1 hora.

Como visto no artigo (Abrishami et al., 2017), as variáveis exógenas são um ponto importante numa análise de previsão de indicadores nas, denominadas pelo próprio autor, “*smart stores*”. É, portanto, relevante para o presente estudo o levantamento das variáveis exógenas que podem impactar o tráfego, em especial, da Loja Alfa. Esta tem um foco particular uma vez que o tráfego em períodos horários é muito mais suscetível a variáveis exógenas, como condições climáticas adversas ou eventos externos, do que propriamente o tráfego em períodos mensais, uma vez que este acaba por se diluir naquilo que é a sazonalidade natural das diferentes alturas do ano e que deverá ser mais simples de captar pelos modelos sem necessidade de inputs adicionais.

Posto isto, o *dataset* para o estudo do tráfego mensal utilizou apenas os dados dos dispositivos de contagem de tráfego das entradas de loja, para um período compreendido entre janeiro de 2016 e abril de 2023, totalizando um valor de 88 meses. Já o *dataset* para o estudo do tráfego para períodos horários utilizou esses mesmos dados de contagem de entradas e, através de recolha de *feedback* junto do responsável de loja da Loja Alfa, optou-se por usar dados relacionados com as condições meteorológicas, eventos externos e feriados. No caso das condições meteorológicas, foi utilizada a API da OpenWeather que fornece dados de temperatura (graus celsius), temperatura sentida (graus celsius), pressão atmosférica (hectopascal), humidade (percentagem), vento (metros por segundo), nebulosidade (percentagem), chuva (volume milímetros) e identificação do clima (ID de tipo de clima), com frequência horária para uma dada localização (par latitude/longitude). Os feriados foram recolhidos, neste caso para a zona da Grande Lisboa, e o mesmo aconteceu para os eventos externos, neste caso relacionados com a ocorrência de eventos de cariz desportivo e cultural junto à Loja Alfa. O *dataset* contém dados com frequência horária para o período entre as 00h00 de 1 de janeiro de 2022 e as 23h59 de 31 de março de 2023, totalizando um valor de 10920 horas.

3.3. Data Preparation

3.3.1. Descrição do *dataset*

Após reunidas todas as fontes de dados necessárias para as duas vertentes do estudo, mensal e horária, foi então feita a construção e preparação dos *dataset*. Nesta fase é aferida a necessidade e utilidade dos vários inputs, são processadas as eventuais limpezas ao *dataset* e criadas e reformatadas as variáveis que mostrarem essas necessidades.

No que diz respeito ao *dataset* do estudo de tráfego mensal, os dados estavam já razoavelmente formatados, sendo apenas necessário ajustar o formato em que a data era apresentada de forma que se pudesse fazer a agregação dos intervalos para períodos mensais. Em termos de limpeza, este *dataset* teve uma particularidade interessante, discutida em pormenor na subsecção seguinte, relativa ao tratamento utilizado para suplantar o impacto que o período da pandemia COVID-19 teve nos dados de tráfego e que alterou os padrões durante sensivelmente dois anos, 2020 e 2021. O *dataset* final seguiu com as

variáveis e formato apresentadas na tabela 3.2 e foi utilizando durante o decorrer de todo o processo de investigação relativo à previsão de tráfego mensal para as lojas escolhidas, quer quando estudadas individualmente, quer quando estudadas como um total de 179 lojas MEO, utilizando nesse caso os dados de todas as lojas somados.

TABELA 3.2. Variáveis do *dataset* de tráfego mensal

Variável	Descrição
Data	Mês e ano a que corresponde o tráfego
Value	Valor inteiro correspondente ao número de entradas do período

Em relação ao *dataset* para estudo do tráfego horário, começou por ser feita uma limpeza nas horas que estavam no *dataset*. Uma vez que a Loja Alfa só estava aberta das 9 horas às 24 horas de todos os dias da semana, filtrou-se o *dataset* para que apanhasse apenas as 15 horas de trabalho. Esta alteração em nada impactou os padrões, uma vez que nos restantes intervalos a loja está fechada. De seguida, optou-se por criar um conjunto de oito variáveis relacionadas com a data, desconstruindo a data em várias componentes que possibilitassem uma mais fácil perceção dos diferentes períodos sazonais e padrões de tráfego, sendo essas variáveis *Year*, *Month*, *Day*, *Hour*, *WeekDay*, *WeekEnd*, *Feriado* e ainda *Closed*, sendo que esta última variável foi necessária por haver períodos em que a loja está fechada por razões extraordinárias como feriado do Natal ou passagem de ano. A estas variáveis adicionou-se ainda uma outra variável binária relativa aos eventos externos, *Jogo*, e por fim adicionaram-se também as variáveis retiradas da API da OpenWeather, nomeadamente *Temp*, *Feels*, *Pressure*, *Humidity*, *Wind*, *Rain*, *Clouds* e ainda *Weather*. O *dataset* final, ficou constituído pelas variáveis e formatos apresentados na tabela 3.3, com um total de 6825 linhas e 18 colunas.

TABELA 3.3. Variáveis do *dataset* de tráfego horário

Variável	Descrição
<i>DateTime</i>	Data e hora do início do período horário de tráfego
<i>Value</i>	Valor inteiro correspondente ao número de entradas do período
<i>Year</i>	Ano do período, retirado da variável <i>DateTime</i>
<i>Month</i>	Mês do período, retirado da variável <i>DateTime</i>
<i>Hour</i>	Hora do período, retirado da variável <i>DateTime</i>
<i>WeekDay</i>	Dia da semana (1 a 7), retirado da variável <i>DateTime</i>
<i>WeekEnd</i>	<i>Boolean</i> que indica se o período se integra no fim de semana
<i>Closed</i>	<i>Boolean</i> que indica se nesse período a loja está fechada
<i>Feriado</i>	<i>Boolean</i> que indica se o período faz parte de um feriado nacional/regional
<i>Jogo</i>	<i>Boolean</i> que indica se durante o período ocorre um evento externo
<i>temp</i>	Temperatura (°C) medida durante o período
<i>feels</i>	Temperatura (°C) sentida durante o período
<i>pressure</i>	Pressão atmosférica (hPa) medida durante o período
<i>humidity</i>	Humidade (%) medida durante o período
<i>wind</i>	Velocidade do vento (m/seg) medida durante o período
<i>clouds</i>	Nebulosidade (%) medida durante o período
<i>rain</i>	Volume de precipitação (mm) medida durante o período
<i>weather</i>	Variável categórica com informação do clima durante o período

3.3.2. Tratamento do período da pandemia COVID-19

O mês de março de 2020 marcou o início da pandemia COVID-19 em Portugal. Um período sem paralelo que afetou todos os setores de atividade do país e que abalou de forma significativa tudo o que envolvesse visitas e atendimentos presenciais. A rede de lojas MEO não foi exceção e os dados do tráfego mostram bem esse impacto. A figura 3.4 mostra a evolução do tráfego mensal ao longo dos meses, desde 2019, no total das lojas MEO e permite verificar o quão significativa foi a descida de tráfego por um longo período.

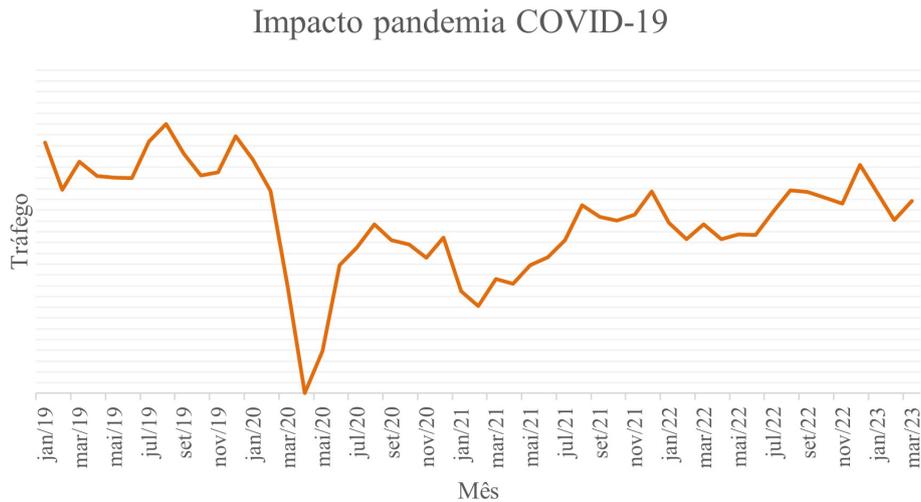


FIGURA 3.4. Tráfego total na rede de lojas desde janeiro de 2019

Este impacto tem muita influência naquilo que é a performance dos modelos de previsão, por duas razões em particular. A primeira razão diz respeito ao impacto nos dados disponíveis para aprendizagem. Os dados de 2020 e 2021 são dados pouco naturais e altamente influenciados pelo sentimento de insegurança que se sentia nos clientes e equipas de loja e que fazia com que o normal fluxo de clientes não ocorresse. A segunda razão consiste no regresso das tendências "normais" de consumo, facilmente identificadas no ano de 2022 (ver figura 3.4), ou seja, os padrões retirados do período mais próximo do que está a ser estudado e do que se pretende prever estão com uma tendência crescente que não diz respeito a um aumento natural de clientes, mas sim a um regresso de uma normalidade que o modelo pode não ter capacidade de conhecer, com base no histórico anómalo da pandemia

Para combater esta situação encontrada nos dados de treino (divisão de conjunto de treino e teste explicada na próxima sub-secção), usaram-se dados artificiais, recorrendo a técnicas de substituição de *outliers*, para o intervalo anómalo considerado. Cada uma das lojas teve os seus valores de tráfego analisado de duas formas diferentes. A primeira opção substitui períodos, considerados como negativos para o treino, por previsões do que teria sido esse período no decorrer normal do ano, ou seja, considerava-se o período anterior e produzia-se uma previsão simples do que seriam aqueles meses sem pandemia, com

utilização da ferramenta Prophet¹. Depois de feita uma previsão de valores, estes poderiam substituir o intervalo anômalo de três formas distintas. A primeira forma partiria do princípio que o tráfego se mantinha nos mesmos níveis pré-pandemia (*last observation carried forward*), a segunda opção considerava que o tráfego quebrava, mas mantinha o padrão, ou seja, utilizávamos a previsão mas colocada no ponto inicial depois da quebra de tráfego (*next observation carried backward*), e a terceira opção considerava uma média entre as duas primeiras, ou seja, colocava a previsão a meio das outras duas opções, de forma a não ter tanto impacto no padrão. A segunda opção passou por identificar os pontos considerados como *outliers* automaticamente, substituindo-os diretamente via interpolação. A figura 3.5 ilustra um exemplo dos diferentes ajustes feitos, neste caso, na loja E.

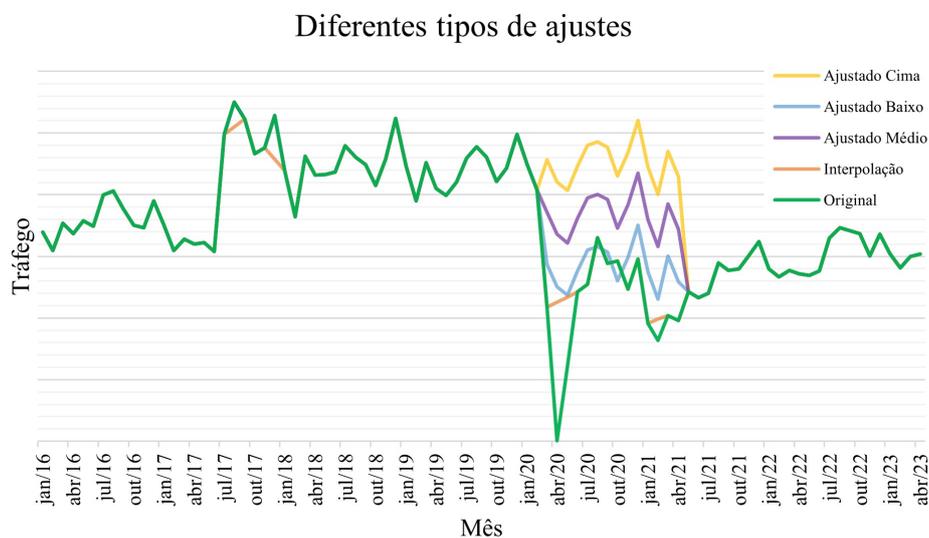


FIGURA 3.5. Diferentes tipos de ajuste testados no tráfego (Loja E)

Depois de criados os diferentes ajustes para cada loja, foi calculada a correlação entre cada um dos testes e o tráfego real, de forma que a decisão de qual usar tivesse em conta as diferenças para a realidade. Dar nota de que, devido ao baixo tráfego e, conseqüentemente, menor impacto da queda de tráfego, algumas lojas não justificaram ter as três hipóteses da previsão de tráfego utilizando o algoritmo base do Prophet.

3.4. Modeling

Nesta fase do estudo são escolhidos e testados os diferentes modelos de previsão. É uma fase caracterizada pela tentativa e erro, numa procura contínua de encontrar o modelo que melhor se ajusta aos objetivos pretendidos. Para que os modelos alcancem a sua melhor performance é também necessário que os dados sejam processados, quando assim necessário, de acordo com certas especificações de cada modelo e por essas razões pode ser

¹<https://facebook.github.io/prophet/>

necessário voltar atrás no processo de estudo, nomeadamente, até ao momento de recolha e seleção dos dados (Shearer, 2000).

Esta fase de modelação foi levada a cabo em dois momentos independentes, cada um deles de acordo com as duas necessidades a que o estudo pretende entregar resposta e que já foram mencionadas anteriormente. Um primeiro momento para previsão de tráfego mensal, utilizando dez lojas, e um segundo momento para tráfego horário usando apenas uma loja, a Loja Alfa.

3.4.1. Conjuntos de treino, conjuntos de teste e critérios de avaliação

Esta é também uma etapa de altíssima importância para o bom desenvolvimento do estudo. Nesta fase foram definidos os conjuntos de treino e teste e as métricas que serviram de base para avaliar cada um dos modelos individualmente e, em última instância, para os comparar entre si.

No que toca ao estudo do tráfego mensal, o *dataset* foi dividido com os dados compreendidos entre janeiro de 2016 e dezembro de 2021 para conjunto de treino e o restante conjunto, compreendido entre janeiro de 2022 e abril de 2023 para conjunto de teste, totalizando um valor de 72 valores para treino e 16 para teste.

No que toca ao estudo do tráfego horário, o método utilizado foi diferente. De forma a ter uma base de avaliação mais robusta, foi utilizada uma técnica de validação cruzada própria para estudo de séries temporais. O original método de validação cruzada segue a linha de pensamento de dividir um dado *dataset* em k partes iguais, sendo depois o modelo testado k vezes, sendo que a cada uma delas um diferente subconjunto é usado como teste, sendo os restantes usados para treino. Tendo em conta que lidamos com séries temporais neste estudo, este método não é útil, uma vez que uma série temporal tem em si intrínseco uma dependência entre valores que segue uma linha de tempo, sendo, portanto, impossível tentar prever um intervalo tendo como input o intervalo a seguir. O método usado foi o de validação cruzada, com a condicionante de que o conjunto de teste para prever um momento t é depois incorporado no conjunto de treino para prever o momento $t+1$ (ver figura 3.6), resultando a média de todos os momentos individuais como métrica final

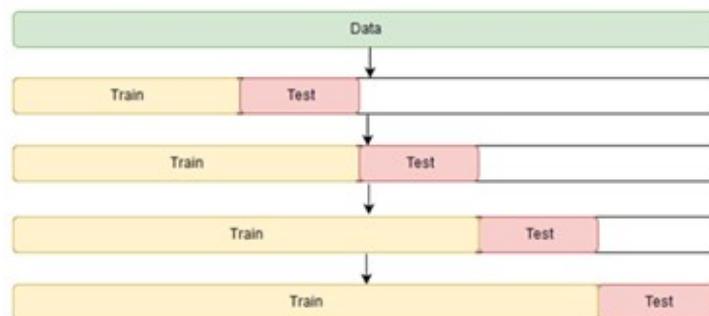


FIGURA 3.6. Método de validação cruzada para séries temporais

O *dataset* de tráfego horário foi então dividido em conjuntos semanais, sendo que, para teste foram usados oito conjuntos diferentes de 105 valores cada (105 horas correspondentes às 15 horas de trabalho de cada um dos sete dias da semana), ou seja, cada um dos modelos foi testado em oito semanas diferentes, de forma iterativa.

Esta é uma técnica que permite testar o modelo em várias realidades diferentes, tendo sido o valor de 8 semanas escolhido por abraçar 2 meses de dados (neste caso, fevereiro e março de 2023).

Em termos de métricas utilizadas para avaliação da capacidade de generalização das previsões também se optou por expressões diferentes. Para a previsão de tráfego mensal, foi escolhido o *Mean Absolute Percentage Error* (MAPE):

DEFINIÇÃO 3.1.

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{x_i} \right|$$

Uma vez que os modelos foram testados em diferentes lojas que, por sua vez, têm níveis diferentes de tráfego, foi necessário escolher uma métrica que não dependesse de escala. Esta é uma métrica que neste contexto não apresenta problemas de maior uma vez que estando a trabalhar valores mensais, é altamente improvável que existam valores perto de 0, o que afetaria de forma significativa o valor do MAPE (Maçaira & Cyrino Oliveira, 2016).

Para a previsão de tráfego horário, foi utilizada a métrica *Relative Mean Absolute Error*, com base no estudo (Sousa et al., 2022). Esta métrica utiliza um modelo *baseline* para obter um *Mean Absolute Error* (MAE) referência e todos os modelos que seguem são avaliados em relação a esse mesmo valor *baseline*.

DEFINIÇÃO 3.2.

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - \hat{x}_i|$$

DEFINIÇÃO 3.3.

$$RelMAE = \frac{MAE}{MAE_{baseline}}$$

Desta forma, inserimos na métrica uma intrínseca avaliação evolutiva face à realidade atual, cujo objetivo passa então por minimizar esse mesmo valor, tendo ainda uma noção percentual da melhoria das métricas face à *baseline*. A utilização desta métrica vem também resolver o problema que o MAPE tinha para esta vertente do estudo. Uma vez que o tráfego é analisado para períodos horários, é normal que nalguns intervalos (maioritariamente no início e fim do dia) existam valores mais reduzidos e eventualmente até possam ser valores nulos, o que comprometeria de forma significativa a utilização do MAPE (Maçaira & Cyrino Oliveira, 2016).

3.4.2. Previsão de tráfego mensal

O estudo da previsão de tráfego mensal englobou duas famílias de modelos. Os modelos da família ARIMA e os modelos Prophet. Esta foi uma escolha feita com base no reduzido número de observações para treino, esperando logo desde início que os modelos da família ARIMA prevalecessem sobre os Prophet.

Em relação ao horizonte temporal a prever, como mencionado na Introdução, os modelos foram pensados para prever 24 valores, ou seja, um horizonte temporal correspondente a dois anos de previsão.

De acordo com o estudo das substituições dos períodos anómalos, cada uma das lojas foi utilizada com a melhor opção de ajuste para o impacto do COVID-19.

Primeiramente, no que toca aos modelos SARIMA, foi verificado se as séries eram consideradas estacionárias, através do teste Augmented Dickey–Fuller (ADF)², que comprovou que todas as séries estudadas eram não-estacionárias, ou seja, a hipótese nula não foi rejeitada em nenhum dos testes. Com esta informação, e sabendo que os modelos ARIMA devem ser aplicados em séries temporais estacionárias, o atributo de diferenciação foi fixado em 1 ($d=1$), uma vez que esse comprovou ser o valor para o qual as séries passavam a verificar as condições de estacionariedade necessárias. De seguida, com utilização da função `auto-arima`³, procurou-se o modelo cujo critério de informação (Akaike Information Criterion, AIC) tinha o menor valor, utilizando como parâmetros de iniciação:

start p = 1, start q = 1, max p = 4, max q = 4, m = 12,

start P = 0, start Q = 0, seasonal = True, d=1, D = 1

O processo decorreu individualmente para cada loja, sendo que cada modelo foi testado num conjunto de teste, anotando posteriormente os resultados obtidos por cada um deles (apresentados na secção seguinte). As previsões dos modelos SARIMA foram programadas para abordar a técnica do *multistep-ahead* através da estratégia iterativa (An & Anh, 2015). Esta estratégia é a mais antiga e possivelmente a mais intuitiva, pois consiste em fazer previsão de um valor e inseri-lo de seguida no conjunto de dados utilizado para previsão do próximo valor.

O segundo conjunto de modelos testado foi a família de modelos Prophet, desenvolvidos pela equipa da Meta. Nesta abordagem, o *dataset* das várias lojas não precisou de ser diferenciado, mas precisou de ser logaritimizado para obter melhores resultados. Os modelos Prophet foram testados com sazonalidade anual aditiva e multiplicativa, apesar de à priori se ter considerado que a sazonalidade atuava de forma aditiva. Ao contrário dos modelos SARIMA, os modelos Prophet devolvem todas as previsões de uma só vez por predefinição, não sendo necessária a criação de uma função para lidar com o *multistep-ahead*.

²<https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.adfuller.html>

³https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html

3.4.3. Previsão de tráfego horário

O processo de estudo da previsão de tráfego horário, apesar de apenas para uma loja em vez de para dez, foi bastante mais complexo que o de previsão de tráfego mensal. Por ser uma frequência temporal mais alta, faz com que os valores sejam mais delicados, havendo um maior espaço para o fator aleatório atuar, para além da complexidade de trabalhar vários períodos sazonais (diário, semanal, mensal e anual).

Para este estudo foram testados os modelos *Naive Forecast*, *Seasonal Window Average*, ARIMA, *Random Forest*, *Exponential Smoothing*, Prophet, XGBoost e ainda *Long Short Term Memory* (LSTM). Esta vertente do estudo foi mais diversificada de forma a criar uma base sólida e vasta de modelos para poder também testar a criação de um modelo híbrido, usado em vários estudos referidos na revisão de literatura da presente dissertação.

Em relação ao horizonte temporal, este estudo foi desenvolvido com o objetivo de prever a próxima semana de tráfego horário, ou seja, 168 períodos, dos quais 105 dizem respeito a horas em que a loja está aberta (*multistep-ahead* fixado em 105 por se terem retirado as horas em que a loja está fora do horário normal).

Naive Forecast e Seasonal Window Average

Os modelos mais simples de todos os testados, correspondem à combinação de valores anteriores para definir o valor futuro. No caso de *Naive Forecast* a teoria é mesmo apontar que o próximo valor será igual ao último. Neste caso, consideramos o último valor do mesmo período sazonal. O *Naive Forecast* com utilização do último valor para o período sazonal pode ser considerado o modelo da atual realidade da MEO para a previsão de tráfego pelo que será usado como modelo *baseline* ao logo deste estudo. O valor obtido de MAE será usado para calcular o *RelMAE* de todos os modelos que se seguem.

Nos modelos *Seasonal Window Average*, o pensamento é maioritariamente o mesmo, mas com a diferença de que apontamos o próximo valor como sendo uma média aritmética simples de n valores anteriores para uma janela sazonal de tamanho m (m períodos formam um ciclo sazonal). Estas técnicas apresentam vantagens para séries temporais sem grande presença de tendências e com sazonalidades bem definidas (Sousa et al., 2022), daí terem sido testadas no presente estudo. De forma a encontrar a melhor combinação entre número de valores (n) e janela de ciclo sazonal (m), foram feitos testes para combinações de possibilidades de $m = 15$ (diário) e $m = 105$ (semanal), considerando janelas até 28 valores (no caso diário) e 10 valores (no caso semanal).

SARIMA

O processo de modelação SARIMA⁴ seguiu o mesmo raciocínio do teste de previsão de tráfego mensal. É um dos modelos mais utilizados na investigação académica, fornecendo muitas qualidades em termos de facilidade de compreensão do modelo e dispêndio computacional. É também um modelo usado de forma comum para estabelecer um *benchmark* (Panay et al., 2021). Começou por se verificar que a série não é

⁴<https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html>

estacionária através dos testes Augmented Dickey–Fuller, Phillips-Perron (PP)⁵ e Kwiatkowski–Phillips–Schmidt–Shin (KPSS)⁶. A série temporal de tráfego horário apresentou resultados nos testes de hipóteses não consistentes entre si, sendo então necessário a utilização dos três testes de raiz unitária e estacionariedade (ADF, PP, KPSS). Uma vez verificada a não-estacionariedade da série, pôde fixar-se o valor de diferenciação em 1, uma vez que este permitia confirmar estacionariedade em todos os testes anteriores. Foi ainda testado o *dataset* com logaritmização antes de ser diferenciado, para evitar grandes discrepâncias de valores, uma vez que o tráfego no início e fim de dia pode chegar a valores pertos de 0. Foi também necessário adicionar um valor fixo a todos os intervalos para permitir a logaritmização com sucesso. Com utilização da função auto-arima foi iniciada a procura pelo melhor modelo, que minimizasse o AIC. A função auto-arima foi utilizada no *dataset* com e sem logaritmização. Para estes testes foram apenas utilizados os valores de tráfego e todos os modelos foram definidos com uso da estratégia recursiva para lidar com o *multistep-ahead*.

Random Forest

*Random Forest*⁷ é um modelo cuja lógica consiste em combinar resultados individuais de várias árvores de decisão de forma a fornecer um resultado mais robusto. Este modelo foi testado com várias combinações do *dataset* e, nalguns casos, foram feitas experiências com variáveis criadas em específico para este modelo. Dado que o modelo *Random Forest* não assume os dados como uma série temporal, foi necessário transformar o *dataset* num problema de aprendizagem supervisionada comum, onde são fornecidos *inputs* e devolvido um *output*, não havendo entre observações qualquer dependência temporal, como acontece numa série temporal comum. Para combater este problema, foram criadas variáveis com referência a momentos passados. A título de exemplo, para prever um dado valor t , foram criadas as variáveis ‘-1’, ‘-2’, ‘-3’, ‘-4’, ‘-5’, ‘-15’ e ‘-105’, para poder fornecer ao modelo *Random Forest* informações não só de variáveis exógenas (utilizadas também nos restantes modelos), como também *inputs* relativos à evolução da série temporal, neste caso, dando valores das últimas 5 horas, do valor do último período sazonal diário (-15 horas) e do último período sazonal semanal (-105 horas). Em termos de número de unidades de regressão foram testados os valores de 500 e 1000 unidades. A estratégia utilizada para a previsão *multistep-ahead* foi também a iterativa, utilizada já anteriormente.

Exponential Smoothing O modelo de *Exponential Smoothing* teve também lugar no decorrer do processo de aprendizagem deste estudo. Os hiper parâmetros foram testados nas suas várias combinações lógicas, tipo de comportamento da sazonalidade e tendência (aditiva e multiplicativa), *damped trend*⁸ como True e False e ainda os períodos sazonais,

⁵<https://arch.readthedocs.io/en/latest/unitroot/generated/arch.unitroot.PhillipsPerron.html>

⁶<https://www.statsmodels.org/stable/generated/statsmodels.tsa.stattools.kpss.html>

⁷<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

⁸<https://www.statsmodels.org/dev/generated/statsmodels.tsa.holtwinters.ExponentialSmoothing.html>

variando entre o valor de 15 (diário) e 105 (semanal). Nota de que a *damped trend*, quando True, serve para amenizar o efeito constante da tendência (quando esta existe).

Prophet

O modelo Prophet, já apresentado na previsão de tráfego mensal, apresenta boa capacidade para lidar com vários períodos sazonais, daí ter sido também testado nesta vertente do estudo. No decorrer dos testes, foram utilizados diferentes conjuntos de hiper parâmetros. Recorreu-se a um *for loop* para testar as possíveis combinações para 'changepoint_prior_scale' (valores 0,05 até 0,5)⁹, 'seasonality_prior_scale' (valores 10 até 60), 'seasonality_mode' ('additive' e 'multiplicative'), 'weekly_seasonality' e 'daily_seasonality' (valores True e False para ambas)¹⁰. À partida, alguns destes valores podiam ser vistos como de clara resposta, como as questões de sazonalidade diária e semanal, ainda assim foram feitos testes com todas as possibilidades. Com os modelos mais claros, foi ainda testada a possibilidade de aplicar uma transformação de escala aos dados (scaler - min e max), bem como a utilização de variáveis exógenas como regressores, para apurar se melhorava a *performance* do modelo. Tal como no modelo de previsão de tráfego mensal, o Prophet devolve logo o número de previsões igual ao horizonte temporal pretendido, não sendo necessário recorrer a funções de *multistep-ahead*.

XGBoost

Semelhante ao raciocínio dos modelos *Random Forest*, os modelos XGBoost¹¹ calculam cada árvore de decisão como uma forma de evolução da última, num género de *gradient boosting* para procurar árvores com melhor performance, ao contrário dos *Random Forest* que processam as várias árvores em paralelo e de forma independente, aglomerando os seus resultados no fim. Os modelos XGBoost foram testados com o mesmo processo lógico das *Random Forest*, ou seja, transformando o problema num tema de aprendizagem supervisionada comum e criando variáveis temporais que introduzissem no modelo relações temporais (variáveis '-1', '-2', '-3', '-4', '-5', '-15' e '-105'). Foram testados modelos com 500 e 1000 unidades para estimadores, utilizando também diferentes tamanhos no conjunto de treino para testar como variava a capacidade de aprendizagem do modelo. Em termos de otimizador, foi usada a métrica de erro absoluto, com o objetivo de evoluir o modelo no sentido da sua minimização.

Long Short-Term Memory

Por último, foram testados os modelos *Long Short-Term Memory* (LSTM), utilizados cada vez com maior frequência na investigação. As LSTM são um tipo específico de redes neuronais recorrentes (RNN), mais robustos para séries temporais que permitem resolver com mais eficácia os problemas de *vanishing gradient*. Para além disto, as LSTM, em comparação com as originais RNN standard, aprendem com maior facilidade, e de

⁹https://facebook.github.io/prophet/docs/trend_changepoints.html

¹⁰https://facebook.github.io/prophet/docs/seasonality,_holiday_effects,_and_regressors.html

¹¹<https://xgboost.readthedocs.io/en/stable/>

forma muito mais rápida, os padrões dos dados de longas séries temporais (Hochreiter & Schmidhuber, 1997).

As LSTM tiveram um foco especial neste estudo, uma vez que foram os modelos cujo tempo investido na sua aprendizagem e avaliação foi maior. Estas necessitam de bastante atenção ao denominado *hyperparameter tuning* para que a sua *performance* possa alcançar os níveis pretendidos.

Estas experiências utilizaram a *framework* Keras¹² dividiram-se também em duas partes, que dizem respeito à técnica de abordagem ao problema do *multistep-ahead*. Um primeiro grupo de testes foi feito com uma estratégia direta, ou seja, a última camada da rede, *output layer* como é referida de forma comum, continha 105 valores, ou seja, as previsões para uma semana foram feitas todas de uma vez, retirando da equação a ligação entre os valores, uma vez que a rede os trata como se fossem *outputs* independentes. O segundo grupo de testes foi realizado com a estratégia recursiva, utilizada nos restantes modelos (com exceção do Prophet), onde a *output layer* tinha apenas um valor de saída e o processo era repetido 105 vezes, utilizando depois os valores previstos como valores input.

Foram testados modelos com diferentes números de camadas e diferentes números de neurónios em cada uma das camadas. Em relação à função de ativação, foi utilizada a função ReLU de forma que a rede pudesse aprender dependências não-lineares entre os dados. O *batch size* (tamanho do conjunto de dados entre cada atualização de parâmetros da rede) e o número de *epochs* (número de vezes que um *dataset* é passado pela rede para aprendizagem) foram testados com vários valores, de forma a apurar a melhor combinação de hiper parâmetros para aprendizagem da rede e evitar *underfitting* ou *overfitting* dos dados, ou seja, momentos em que a rede não tem capacidade de generalizar os padrões dos dados, quer no conjunto de treino quer no conjunto de teste, ou momentos em que a rede se adapta demasiado ao conjunto de treino perdendo depois capacidade de obter uma boa performance preditiva em dados não vistos (conjunto de teste).

De forma a correr as várias abordagens da rede neuronal LSTM foi criado ainda um conjunto de dados considerado conjunto de validação, pertencente inicialmente ao conjunto para treino, de forma que o modelo pudesse ser treinado e estivesse a ser recolhido *feedback* da sua evolução sem utilização do conjunto de teste. Em termos de otimizador da rede, foi testado o otimizador Adam e *stochastic gradient descent* (SGD). Foram ainda feitos testes com diferentes *loss function*, variando entre *mean absolute error* e *mean squared error*, conforme o modelo parecesse estar a adaptar-se melhor. Por último, o *learning rate* do otimizador da rede (valores entre 0,0001 e 0,05) e os *dropouts* entre camadas foram também testados com vários valores de forma a apurar o método mais estável para aprendizagem da rede. O *learning rate* é importante para diminuir a velocidade de aprendizagem e conseqüentemente tornar a aprendizagem um pouco mais estável em termos de *update* de valores em cada uma das unidades, e os *dropouts* foram testados com o objetivo

¹²<https://keras.io/>

de cortar neurónios das camadas de forma aleatória tentando assim que a rede ganhasse maior capacidade de generalização para todos os dados, evitando o *overfitting* à série.

Modelo *Ensemble*

Um dos grandes objetivos deste estudo era o de testar um modelo *ensemble*. Uma abordagem que aglomera vários resultados, atribuindo pesos a cada uma das previsões dos diferentes modelos com o objetivo de alcançar uma ainda melhor performance que os modelos individuais. Esta é uma técnica em crescente tendência, utilizada com sucesso no estudo (Sousa et al., 2022) e com ótimos resultados nas recentes M4 (Makridakis et al., 2020) e M5 *forecasting competitions* (Makridakis et al., 2022).

Após todos os modelos terem alcançado as suas melhores previsões, uns com maior capacidade de generalização e outros com mais sensibilidade aos picos e quedas de tráfego, procurou-se abordar o problema como uma questão de otimização, ou seja, utilizar combinações de modelos com pesos atribuídos de forma a minimizar o erro absoluto, num género de média ponderada de previsões.

3.5. *Evaluation*

Nesta fase do estudo foi feita a avaliação dos diferentes modelos e, acima de tudo, dos resultados que foram sendo obtidos para as necessidades e problemas inicialmente identificados. Esta fase foi maioritariamente composta por momentos de recolha de *feedback*, junto dos vários *stakeholders* do projeto (responsável de loja, gestor de lojas e equipa de gestão MEO). Segundo a metodologia CRISP-DM esta fase é definida por “*Evaluate Results*”, “*Review Process*” e “*Determine Next Steps*” (Shearer, 2000), e foi precisamente nesse sentido que estas reuniões tomaram lugar. Como referido anteriormente, a metodologia tem implícita uma linha de pensamento cíclica, sendo esta fase a última do ciclo, pois com as informações e *feedback* recolhidos é possível regressar a fases anteriores, nomeadamente a fase de *Data Preparation* e *Modeling*, para que o estudo possa ser alinhado de acordo com os novos pontos que são levantados.

3.6. *Deployment*

Esta é uma fase, que num projeto académico não é muitas vezes referida, ainda que faça parte da metodologia CRISP-DM dado que a mesma existe para guiar todo o tipo de projetos na área da ciência de dados (ou *data mining* com indica o nome). Ainda assim, é importante dar nota que o processo de estudo e previsão de tráfego mensal e horário foi desenvolvido com o intuito de ser aplicado na rede de lojas MEO, pelo que no caso de sucesso futuro deste projeto, a fase de *Deployment* entra em jogo para entregar valor à empresa e, em particular, às equipas de gestão das lojas MEO.

No caso particular da previsão de tráfego mensal, foi já feito o levantamento das áreas internas que podem tirar valor destes *inputs* e em breve os mesmos deverão ser disponibilizados de forma organizada na sua íntegra com base de frequência mensal, desde gestores de topo até equipas operacionais.

CAPÍTULO 4

Resultados e Discussão

Após ter sido descrita toda a metodologia utilizada nas várias fases do presente estudo, importa agora apresentar os resultados obtidos das várias experiências realizadas nas duas vertentes do trabalho. Neste capítulo serão também discutidos os resultados à luz dos objetivos propostos para dissertação, bem como relativamente às necessidades levantadas do ponto de vista do negócio.

Mais uma vez, este capítulo vai estar dividido mediante as duas vertentes do estudo de tráfego, mensal e horário.

4.1. Tráfego mensal

O *dataset* de cada uma das dez lojas estudadas foi composto por dados de tráfego mensal compreendidos entre janeiro de 2016 e abril de 2023. Os gráficos de linhas, ilustrados nas figuras 4.1 e 4.2, mostram a evolução do tráfego da Loja C e da Loja E, respetivamente, de onde podemos verificar o já referido impacto do período pandémico. Dar nota de que a Loja E tem muito mais tráfego que a Loja C, daí os padrões serem mais definidos e o impacto da pandemia ser mais significativo nos mesmos.

Nos gráficos podemos verificar ainda dois exemplos de ajustes feitos aos diferentes *datasets*. No caso da Loja C, com a opção ajustado baixo, ou seja, estratégia do género *next value carried backward*, onde a previsão foi ajustada para o nível médio entre o ajustado cima e o ajustado baixo (ver figura 4.1).

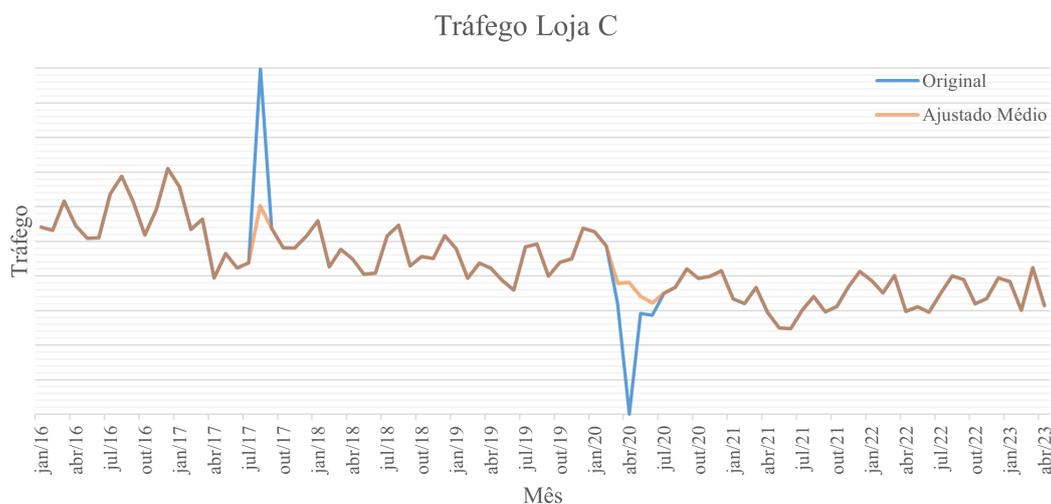


FIGURA 4.1. Tráfego mensal da loja C com ajuste aplicado

Já no caso da Loja E a opção recaiu no ajustado baixo, onde a previsão do Prophet para o período da pandemia foi ajustado para o primeiro valor a seguir ao impacto, num tipo de estratégia *next value carried backward* ao invés de *last value carried forward* (ver figura 4.2).

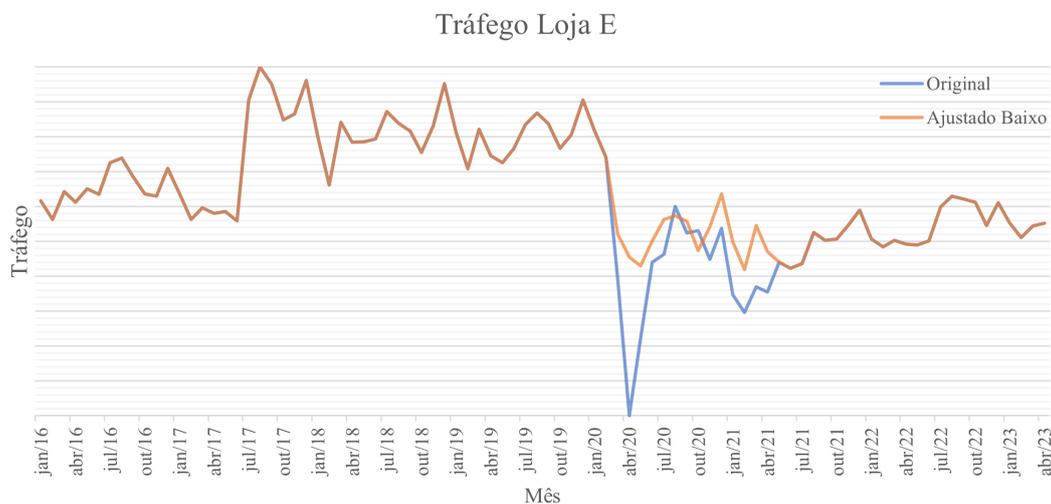


FIGURA 4.2. Tráfego mensal da loja E com ajuste aplicado

Como já referido na Metodologia, foi criada uma tabela com a correlação entre cada *dataset* ajustado e a sua versão original para que se pudesse definir quais os ajustes a ser testados nos modelos. Tabela A.XY em anexo.

De seguida, começou por ser aplicado modelo SARIMA. Já com informação do grau de diferenciação necessário ($d=1$), recorreu-se à função `auto_arima` para procurar a melhor combinação de hiper parâmetros. Uma vez devolvida uma solução, os hiper parâmetros (as ordens dos modelos) foram ajustados de acordo com a significância descrita no sumário do modelo e de seguida foram testados os modelos no conjunto de teste definido (dados de janeiro de 2022 até abril de 2023). A Tabela 4.1 ilustra os diferentes resultados obtidos para cada loja e respetiva métrica de avaliação da performance de previsão.

TABELA 4.1. Resultados da previsão de tráfego mensal obtida para cada loja

Loja	Modelo	Ajuste	MAPE
Loja A	SARIMA(1, 1, 1)x(0, 1, 1, 12) recursivo	Interpolação	7,169%
Loja B	SARIMA(0, 1, 0)x(0, 1, 1, 12) recursivo	Interpolação	7,553%
Loja C	SARIMA(0, 1, 0)x(0, 1, 1, 12) recursivo	Ajustado Médio	5,640%
Loja D	SARIMA(0, 1, 1)x(0, 1, 1, 12) recursivo	Interpolação	8,820%
Loja E	SARIMA(0, 1, 1)x(0, 1, 2, 12) recursivo	Ajustado Baixo	6,359%
Loja F	SARIMA(0, 1, 0)x(0, 1, 1, 12) recursivo	Interpolação	8,590%
Loja G	SARIMA(0, 1, 0)x(0, 1, 2, 12) recursivo	Interpolação	9,359%
Loja H	SARIMA(0, 1, 0)x(0, 1, 1, 12) recursivo	Interpolação	10,117%
Loja I	SARIMA(0, 1, 0)x(0, 1, 1, 12) recursivo	Interpolação	11,899%
Loja J	Prophet(yearly, seasonality_mode='additive')	Interpolação	20,504%

Os modelos SARIMA tiveram predominância em todas as lojas, com exceção da loja J, onde o tráfego é bastante baixo, o que faz com que seja também difícil encontrar padrões que possam guiar as previsões.

Do conjunto de testes, traçou-se um padrão. O modelo SARIMA(0,1,0)x(0,1,1, 12)¹ pareceu ajustar-se bem a grande parte dos *datasets*. Mesmo não sendo a melhor combinação de parâmetros para algumas lojas, os resultados obtidos quando testados não foram significativamente diferentes das melhores combinações, no que toca ao MAPE obtido. Com essa conclusão testou-se esse mesmo modelo no conjunto de dados da rede inteira, ou seja, o valor do tráfego de todas as 179 lojas somadas, com o objetivo de poder prever a evolução do tráfego da rede a 24 meses sem que fosse necessário um grande conjunto de dados. A experiência foi realizada no conjunto de dados a começar em janeiro de 2019, tendo-se usado o mesmo conjunto para teste das restantes experiências. Em termos de *outliers* da pandemia, os mesmos foram substituídos por interpolação. Os resultados da previsão e a série mensal no conjunto de teste são apresentados na figura 4.3. Neste caso, alcançaram um MAPE de 4,489%, que corresponde a um valor bastante positivo para o estudo, pois, a previsão da série global apresenta melhor *performance* do que as previsões individuais.

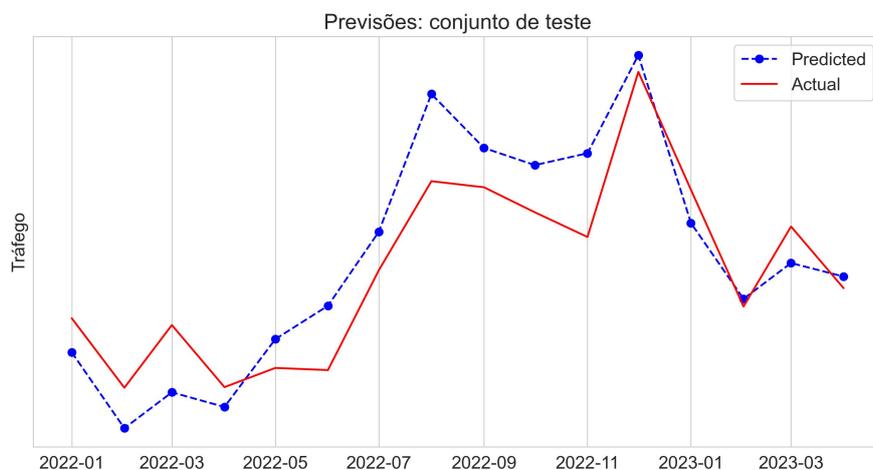


FIGURA 4.3. Performance do modelo SARIMA no conjunto de teste, para valores agregados da rede MEO

Em relação aos modelos Prophet, a única loja cujo resultado foi melhor que os modelos ARIMA, como já indicado, foi a loja J, mas ainda assim obteve um resultado pouco positivo face aos restantes, com a métrica MAPE de 20,504% pelo que não houve avanços significativos na utilização desse modelo.

¹<https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html>

4.2. Tráfego horário

Passamos agora à vertente do estudo do tráfego horário, parte da investigação que necessitou de mais atenção dado o grande conjunto de variáveis disponíveis e a complexidade dos vários tipos de sazonalidade presentes (diária, semanal, mensal e anual).

Para iniciar este estudo, como descrito na Metodologia, foi necessário apurar o modelo *baseline* sobre o qual iriam ser avaliados todos os restantes modelos. Dado que na empresa, a metodologia para tráfego está muito ligada a comparação com períodos anteriores, optou-se por definir o método *naive forecast* como modelo *baseline*.

Como nota, tendo em conta que os modelos foram testados com validação cruzada aplicada a séries temporais, ou seja, testados em oito períodos semanais diferentes, mas seguidos, os gráficos apresentados de seguida poderão não conter todos os testes feitos, pois os mesmos não seriam possíveis de apresentar de forma clara na mesmo gráfico.

Naive Forecast

As previsões do modelo *Naive Forecast* foram produzidas com sazonalidade semanal, ou seja, na teoria prevê-se que a próxima hora vai ter o mesmo tráfego que a mesma hora de há exatamente uma semana. Este modelo obteve um MAE de 17,864 e este valor será o valor *baseline* sobre qual todos os modelos serão comparados, de forma a constituir a métrica RelMAE. A Figura 4.4 ilustra a sua performance na semana 8 do conjunto de teste.

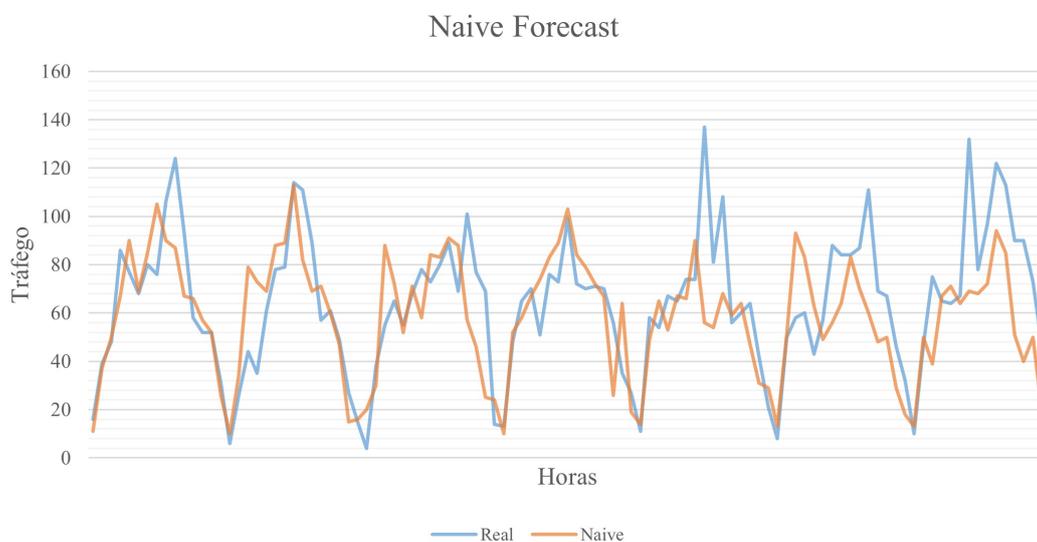


FIGURA 4.4. Previsão do modelo Naive Forecast na semana 8 do conjunto de teste

Seasonal Window Average

Os modelos *Seasonal Window Average* foram testados com duas sazonalidades. No primeiro teste foi utilizada a sazonalidade diária, ou seja, 15 horas. Recorreu-se a um *for loop* que testasse as várias possibilidades para o tamanho do grupo de valores apanhados para essa sazonalidade. Para sazonalidade diária, o modelo com melhor MAE foi o modelo com 13 valores, que obteve um MAE de 14,206, correspondendo a um RelMAE

de 0,795. No segundo teste foi utilizada a sazonalidade semanal, método mais próximo do *Naive Forecast* utilizado para *baseline*. De novo, foi criado um *for loop* para testar várias possibilidades. O modelo com melhor *performance* foi o modelo com 5 valores, obtendo um MAE de 14,687 e um correspondente RelMAE de 0,822. As Figuras 4.5 e 4.6 mostram as previsões produzidas pela abordagem de sazonalidade diária e semanal, respetivamente, para a semana 2 do conjunto de teste.

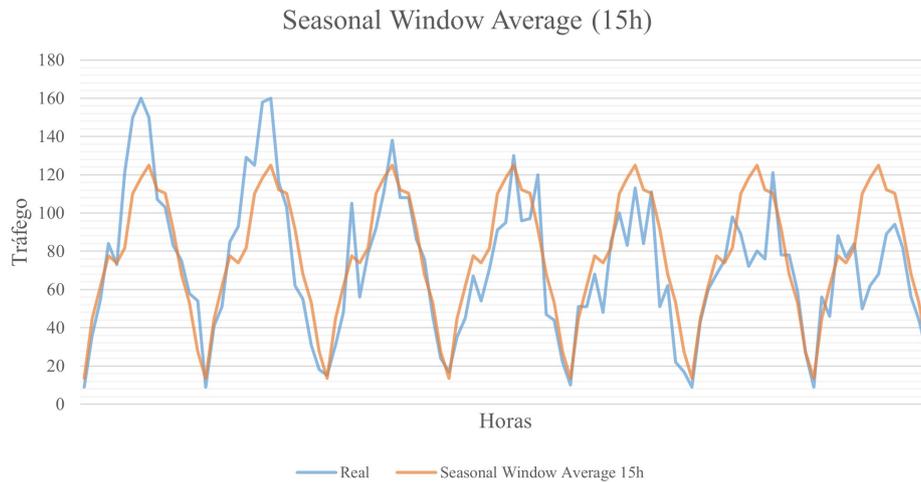


FIGURA 4.5. Previsão do modelo SeasWA (diário) na semana 2 do conjunto de teste

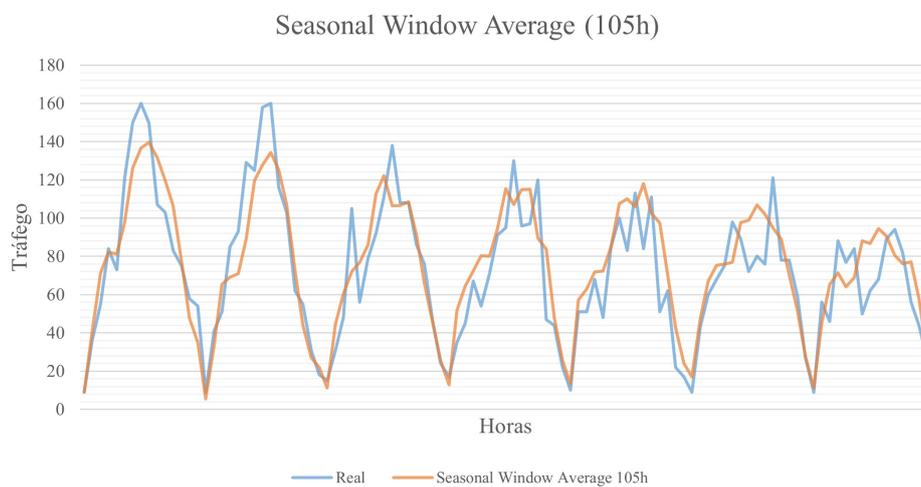


FIGURA 4.6. Previsão do modelo SeasWA (semanal) na semana 2 do conjunto de teste

SARIMA

Para utilizar os modelos SARIMA começou por se fazer uma análise aos dados, para entender melhor como trabalhar as diferentes sazonalidades, uma vez que o modelo precisa desse mesmo parâmetro para obter um bom desempenho. A Figura 4.7 mostra como varia a tendência do tráfego da Loja Alfa, mediante os diferentes períodos de *rolling mean*.

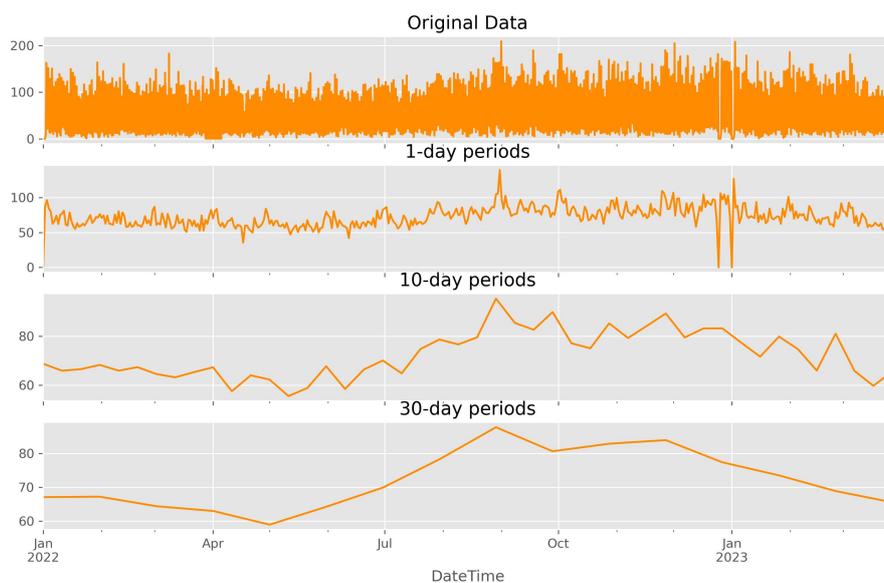


FIGURA 4.7. Rolling mean dos dados da Loja Alfa

Daqui dá para entender que os dados têm uma certa sazonalidade mensal e anual, subindo o tráfego no verão e ligeiramente na altura do Natal, ainda assim foi preciso aprofundar mais para entender como funcionava a sazonalidade diária e semanal, de forma a decidir qual o *input* a dar ao modelo SARIMA na componente sazonal. Depois de diferenciada a série temporal, foi analisado o seu gráfico de autocorrelações, ilustrado na Figura 4.8, que mostra que há uma correlação significativa nos múltiplos de 15, ou seja, na sazonalidade diária, e o mesmo acontece no valor de 105, ou seja, sazonalidade semanal, no entanto esta não parece ser diferente dos valores diários, o que dá a entender que a sazonalidade mais forte é mesmo a diária, daí seguirmos com valor 15 para a componente de sazonalidade dos modelos SARIMA.

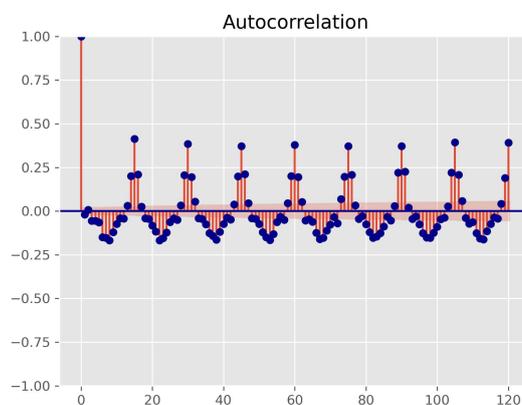


FIGURA 4.8. Autocorrelação dos dados da Loja Alfa

No que toca à *performance* dos modelos, foram testadas várias combinações de modelos SARIMA, dentro do que foi sugerido pela função `auto_arima`, algumas delas demorando uma quantidade significativa de tempo para treinar. A Tabela 4.2 mostra os vários testes realizados.

TABELA 4.2. Resultados das experiências SARIMA

Modelo	MAE	RelMAE
SARIMA(2,1,0)(2,1,0,15)	19,907	1,114
SARIMA(3,1,0)(1,1,0,15)	22,239	1,245
SARIMA(4,1,0)(2,1,0,15)	18,766	1,050
SARIMA(4,1,0)(3,1,0,15)	19,311	1,081
SARIMA(4,1,0)(6,1,0,15)	16,556	0,927
SARIMA(4,1,0)(7,1,0,15)	16,363	0,916
SARIMA(5,1,0)(2,1,0,15)	18,562	1,039
SARIMA(5,1,0)(6,1,0,15)	16,908	0,946

O modelo que obteve melhor *performance* foi o modelo (4,1,0)(7,1,0,15), e destes testes entendeu-se que um dos pontos mais importantes para uma melhor *performance* foi o parâmetro P, responsável pela componente auto regressiva da parte sazonal do modelo. Este, apesar de ter alto impacto na *performance*, também aumentou bastante o tempo gasto no treino do modelo. Este modelo obteve um MAE de 16,363 e um RelMAE de 0,916 e a Figura 4.9 ilustra a sua *performance* na semana 1 do conjunto de teste.

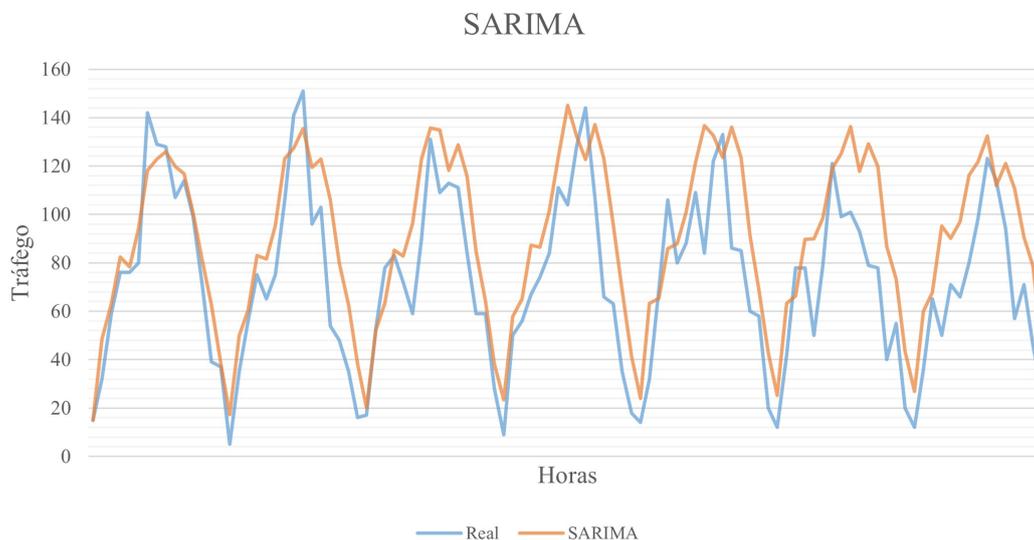


FIGURA 4.9. Previsão do modelo SARIMA na semana 1 do conjunto de teste

Random Forest

Para o modelo *Random Forest*, começaram por ser utilizadas apenas as variáveis *Year*, *Month*, *Day*, *Hour*, *WeekDay*, *Weekend*, *feels* e *rain*. A juntar a estas, utilizaram-se as variáveis que inseriam valores passados da série no modelo. Inicialmente utilizaram-se os valores das cinco últimas horas antes da hora que se pretendia prever, ou seja, de t-1 a t-5.

Uma segunda experiência envolveu não as cinco horas anteriores, mas os momentos sazonais anteriores, ou seja, utilizou-se $t-1$, $t-15$ e $t-105$. Com este novo caso a *performance* do modelo melhorou. Foi ainda testada a incorporação da variável relativa aos eventos externos mas, curiosamente, o modelo piorou, ainda que de forma imperceptível, os seus resultados. Depois de apurado o melhor modelo, experimentou-se a mesma configuração mas utilizando os dados sujeitos a transformação *MinMaxScaler*, mas a *performance* piorou. A Tabela 4.3 mostra os resultados obtidos para cada experimentação.

TABELA 4.3. Resultados das experiências *Random Forest*

Modelo	MAE	RelMAE	Comentários
RandomForest(1000)	12,903	0,722	conj. treino = 1000, shifts até 5
RandomForest(1000)	12,575	0,704	conj. treino = 1000, shifts 1h, 15h, 105h
RandomForest(1000)	12,914	0,723	conj. treino = 2000, shifts 1h, 15h, 105h
RandomForest(500)	12,602	0,705	conj. treino = 1000, shifts 1h, 15h, 105h
RandomForest(500)	12,616	0,706	conj. treino = 1000, shifts 1h, 15h, 105h, Jogo

Destes resultados se tiram várias indicações interessantes. O modelo começou por ser treinado com um conjunto de treino de tamanho 1000, no entanto, quando aumentado para 2000 o modelo piorou a sua *performance*. Em relação ao número de estimadores, apesar do modelo com melhor *performance* ter sido alcançado com 1000 estimadores, a mesma configuração de modelo com apenas 500 estimadores só aumentou o RelMAE em 0,001 que é praticamente imperceptível, ou seja, dado que com 1000 estimadores o modelo demora mais a correr, a melhoria que oferece acaba por não compensar. Ainda assim, o melhor modelo foi o *Random Forest* com 1000 estimadores, utilizando as variáveis antes mencionadas e as variáveis criadas $t-1$, $t-5$ e $t-105$ obtendo um MAE de 12,575 e um RelMAE de 0,704. A Figura 4.10 mostra as previsões para a semana 6 do conjunto de teste.

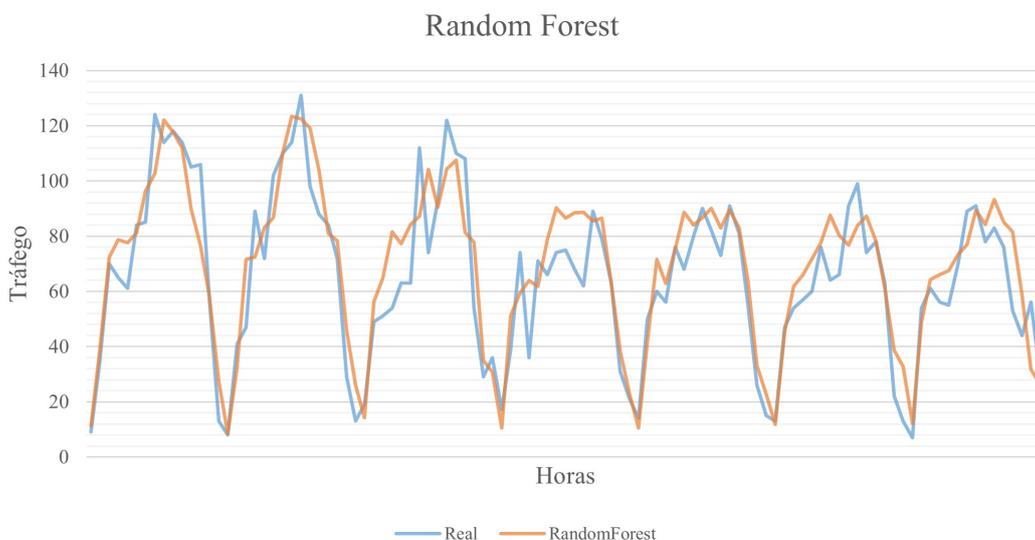


FIGURA 4.10. Previsão do modelo Random Forest na semana 6 do conjunto de teste

Exponential Smoothing

Com modelos *Exponential Smoothing* foram feitas experiências com emparelhamento dos parâmetros *damp* (True ou False) e o número de *timesteps* que constituem um ciclo sazonal (15 ou 105). A Tabela 4.4 ilustra os resultados obtidos para cada um destes testes.

TABELA 4.4. Resultados das experiências *Exponential Smoothing*

Modelo	MAE	RelMAE	Comentários
ExpSmoothing	16,090	0,901	trend=add, damp=True, seasonal=add, periods=15
ExpSmoothing	17,048	0,954	trend=add, damp=False, seasonal=add, periods=15
ExpSmoothing	18,662	1,045	trend=add, damp=True, seasonal=add, periods=105
ExpSmoothing	18,439	1,032	trend=add, damp=False, seasonal=add, periods=105

O melhor modelo de *Exponential Smoothing* obteve um MAE de 16,090 e um RelMAE de 0,901, sendo conseguido através do ciclo sazonal com 15 horas, ou seja, sazonalidade diária. A Figura 4.11 mostra as previsões do modelo para a semana 4 do conjunto de teste.

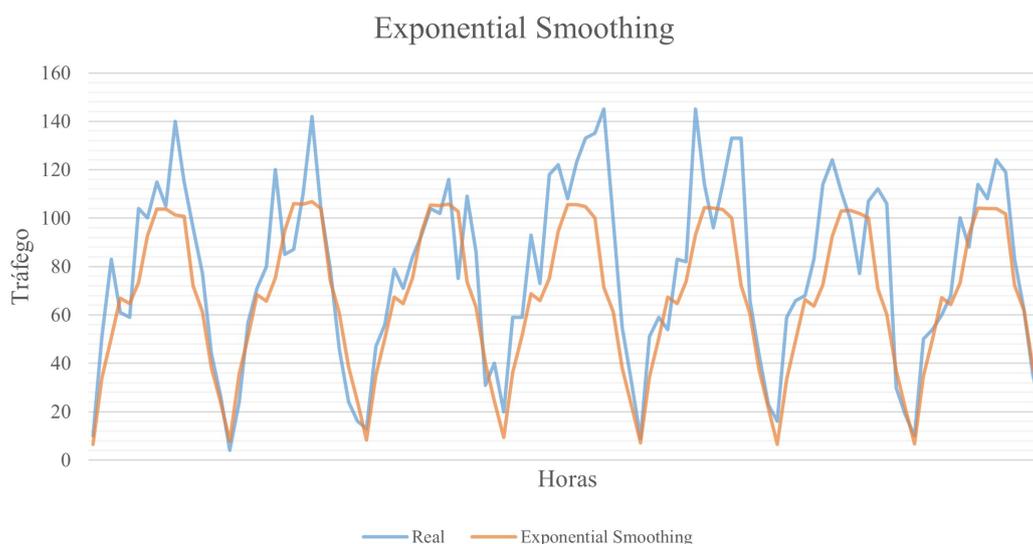


FIGURA 4.11. Previsão do modelo Exponential Smoothing na semana 4 do conjunto de teste

Prophet

Como descrito na metodologia, o modelo Prophet teve um processo de experimentação vasto. O conjunto de hiper parâmetros ‘*changepoint_prior_scale*’ e ‘*seasonality_prior_scale*’ foram emparelhados mediante todas as possibilidades dentro de um conjunto significativo de valores, no entanto, as experiências não se mostraram significativas (alterações entre 0,1 e 0,2 nos MAEs obtidos) e acabou por se optar seguir com os valores de 0,5 para ‘*changepoint_prior_scale*’, por ser o que apresentava ligeira vantagens para a *performance* e 10 para ‘*seasonality_prior_scale*’, que é o valor definido como *default* na documentação do Prophet. O efeito da sazonalidade ficou definido como aditivo, uma escolha que segue o conhecimento do mercado do retalho e que já foi utilizada nos modelos anteriores. Em

termos de sazonalidades, optou-se, segundo os resultados obtidos, por utilizar sazonalidade semanal e mensal. Apesar de se saber que existe também uma sazonalidade anual o conjunto de dados disponibilizado não contempla dois ciclos anuais completos, pelo que os padrões não seriam captados de forma correta. A Tabela 4.5 apresenta os resultados das diferentes experiências.

TABELA 4.5. Resultados das experiências Prophet

Modelo	MAE	RelMAE	Comentários
Prophet 1	13,847	0,775	seasonality_mode='additive'
Prophet 2	14,089	0,789	yearly_seasonality=True
Prophet 3	14,814	0,829	Prophet 1 + Dados logaritmizados
Prophet 4	13,846	0,775	Prophet 1 + <i>MinMax Scaler</i>
Prophet 5	13,490	0,755	Prophet 4 + 'feels'
Prophet 6	13,546	0,758	Prophet 1 + changepoint_prior_scale=0.5
Prophet 7	13,544	0,758	Prophet 6 + <i>MinMax Scaler</i>
Prophet 8	13,360	0,748	Prophet 7 + 'feels'
Prophet 9	13,364	0,748	Prophet 8 + 'rain'
Prophet 10	13,390	0,750	Prophet 8 + 'Feriado'
Prophet 11	13,330	0,746	Prophet 8 + 'WeekDay'
Prophet 12	13,308	0,745	Prophet 11 + 'Hour'

Destes testes retirámos que as variáveis “Feriado” e “Jogo” não ofereceram mais-valia aos modelos Prophet. A transformação com o *MinMaxScaler* fez uma grande diferença, permitindo que o modelo alcançasse uma melhoria significativa de *performance*. O modelo que melhor desempenho obteve juntou as sazonalidades semanal e mensal, ‘seasonality_prior_scale’ e ‘changepoint_prior_scale’ com valores 10 e 0,5 respetivamente, e ainda a utilização das variáveis *feels*, *WeekDay* e *Hour*, com a transformação *MinMaxScaler* na variável *Value (target)* e *feels*. Este modelo obteve um MAE de 13,308 que corresponde a um RelMAE de 0,745. A Figura 4.12 apresenta um dos testes feitos a este mesmo modelo, neste caso, na semana 7 do conjunto de teste.

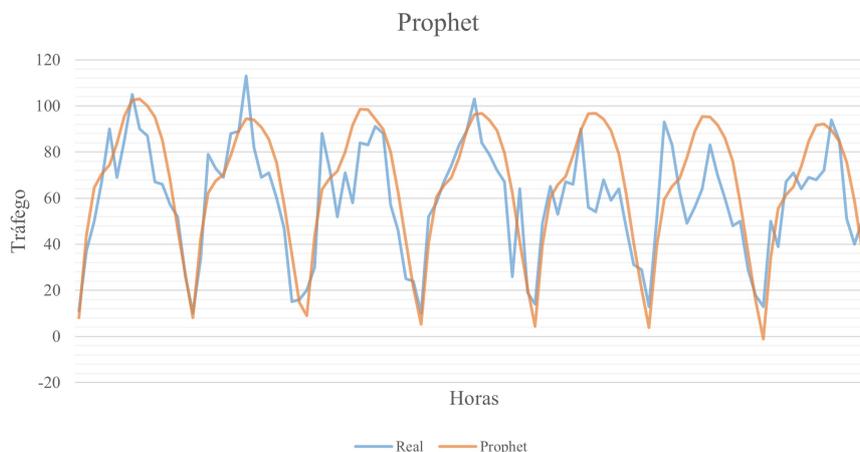


FIGURA 4.12. Previsão do modelo Prophet na semana 7 do conjunto de teste

XGBoost

O modelo XGBoost foi alimentado, à imagem dos modelos *Random Forest*, por todas as variáveis, com exceção da variável “temp” (temperatura) uma vez que a variável “feels” (temperatura sentida) dá praticamente a mesma informação. A estas juntaram-se ainda, de novo, as variáveis t-1, t-15 e t-105, que produziram melhores resultados nos modelos *Random Forest*. A Tabela 4.6 ilustra as diferentes experiências efetuadas.

TABELA 4.6. Resultados das experiências XGBoost

Modelo	MAE	RelMAE	Comentários
XGBoost	12,583	0,704	n_estimators=1000, com 1000 do conjunto de treino
XGBoost	12,514	0,700	n_estimators=1000, com 2000 do conjunto de treino
XGBoost	12,585	0,704	n_estimators=500, com 2000 do conjunto de treino
XGBoost	11,749	0,658	n_estimators=500, conjunto de treino inteiro
XGBoost	11,799	0,660	n_estimators=1000, conjunto de treino inteiro

Os testes iniciaram-se com 1000 estimadores, no entanto o modelo com 500 estimadores apresentou resultados idênticos, poupando assim tempo e capacidade computacional. Em relação ao conjunto de dados utilizado para treino, começou por se utilizar apenas uma porção do conjunto de treino, mas o melhor modelo acabou por ser obtido com o conjunto de treino a utilizar todos os dados disponíveis para tal. Foi interessante notar que, mesmo utilizando todos os dados disponíveis, o número de estimadores não fez diferença quando duplicado, tendo inclusive piorado a *performance*. Durante os testes, utilizou-se ainda o *MinMax scaler*, que em modelos anteriores tinha ajudado a melhorar a *performance*, mas de acordo com o anotado nos modelos *Random Forest*, este acabou por piorar a *performance* dos modelos de forma significativa. Todos estes testes foram feitos com uso da métrica MAE como valor objetivo a minimizar.

O melhor modelo acabou por ficar definido com 500 estimadores, utilizado todo o *dataset* disponível para conjunto de treino, alcançando um MAE de 11,749 e um RelMAE de 0,658. A Figura 4.13 ilustra as previsões do modelo para a semana 5 do conjunto de teste.



FIGURA 4.13. Previsão do modelo *XGBoost* na semana 5 do conjunto de teste

Long Short-Term Memory

No que toca aos modelos LSTM, como já apresentado na Metodologia, estes seguiram duas direções de estudo. Uma com previsão dos 105 valores diretamente e uma outra com a rede a prever apenas um valor, sendo depois utilizada a técnica de previsão *multistep-ahead* iterativa.

A Tabela 4.7 mostra como decorreram as várias tentativas de aprendizagem das redes LSTM de previsão direta, ou seja, cuja *output layer* era constituída por 105 unidades, bem como os resultados obtidos para cada uma delas.

TABELA 4.7. Resultados das experiências LSTM (direta)

Modelo	MAE	RelMAE	Comentários
LSTM 1	14,911	0,835	Value, duas camadas (20,70), epochs = 200, batch = 105
LSTM 2	14,947	0,837	Value, uma camadas (20), epochs = 300, batch = 105
LSTM 3	15,344	0,859	Value, duas camadas (50,70), dois dropouts de 0.2, batch = 105
LSTM 4	16,846	0,943	10 Features, uma camada (20), epochs = 40, batch = 105
LSTM 4	16,252	0,910	LSTM4 com MSE como loss function em vez de MAE
LSTM 5	17,784	0,995	Com 45 batch size
LSTM 4	15,439	0,864	Value de duas semanas, month, day, weekend, weekday, feels, rain
LSTM 6	14,160	0,793	LSTM4 com 315 timesteps de input
LSTM 7	15,778	0,883	Duas camadas (20,30), epochs=25, batch=105
LSTM 4	13,677	0,766	LSTM4, epochs = 50, batch = 105, timesteps = 315, scaler(-1,1)
LSTM 8	14,418	0,807	Três camadas (32, 32, 16) com ReLu nas duas últimas
LSTM 8	13,334	0,746	Três camadas (32, 32, 16) com ReLu nas três

Os testes começaram por ser conduzidos com redes simples, indo aumentado a sua complexidade com o número de casos implementados. Inicialmente começou por se usar o *dataset* formatado para entrar na camada de *input* com 105 *timesteps*, sendo depois testado também 210 *timesteps* (duas semanas) e 315 *timesteps* (três semanas), sendo que estes aumentos foram mostrando algumas melhorias na *performance* de previsão, mas ao mesmo tempo mostraram também aumentos no tempo de execução do modelo. O ponto de partida para o estudo foi perceber como variavam as formas de aprendizagem mediante o *batch size* e o número de *epochs*. A Figura 4.14 mostra as curvas da função custo para o treino e validação do modelo LSTM1, percebendo-se que apesar de haver uma aprendizagem rápida nos primeiros *epochs*, com o passar do tempo o modelo estagna em termos de aprendizagem, não mostrando capacidade de melhorar em termos de conjunto de treino ou validação.

Com isto pretendeu-se aumentar a complexidade dos modelos, uma vez que os testes relativamente ao *batch size* ideal para aprendizagem mostravam irremediavelmente o mesmo problema. Decidiu seguir-se utilizando testes com mais variáveis e aumentando o número de camadas, o que se mostrou proveitoso, uma vez que o modelo de previsão direta com melhor *performance* foi um modelo com três camadas de neurónios (32, 32 e 16 neurónios), *batch size* de 105 (uma semana de dados) e 20 *epochs*, melhorando o MAE para 13,334, que corresponde a um RelMAE de 0,746. Para além destas alterações, também o facto de alterar o *range* do *scaler* de (0,1) para (-1,1) fez melhorar a *performance* dos modelos no geral, sendo adotado daí em diante como *scaler* predefinido.

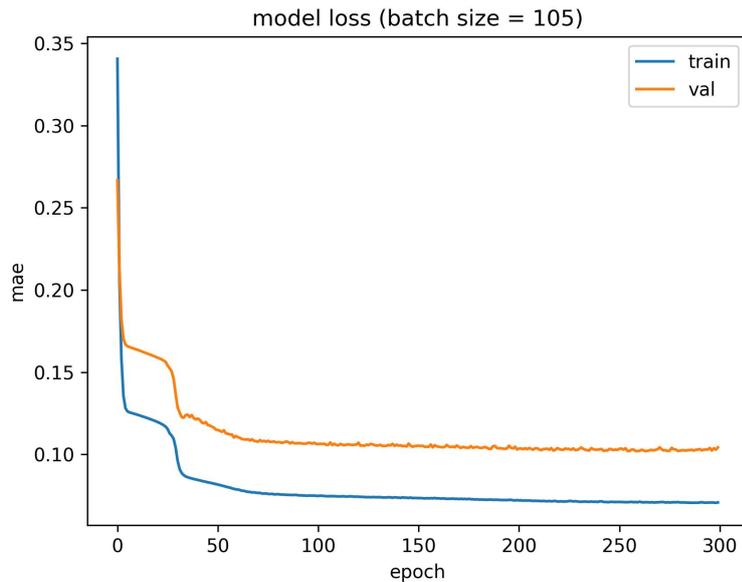


FIGURA 4.14. *Loss function de treino e validação de uma rede LSTM (direta)*

Em relação à modelação de uma LSTM para previsão de um valor e sua utilização numa estratégia de *multistep-ahead* iterativa, começou-se o estudo utilizando já as aprendizagens dos modelos testados para a previsão de todos os valores diretamente. Alterou-se a *output layer* para devolver apenas um valor e iniciaram-se os testes com LSTM de duas camadas e diferentes configurações de *batch size* e número de *epochs*. Ao longo dos testes, a *loss function* foi sendo alterada, entre MSE e MAE, conforme o que parecia ser uma melhor adaptação e estabilidade do modelo no processo de aprendizagem. A Tabela 4.8 mostra a evolução dos vários testes nesta fase.

TABELA 4.8. Resultados das experiências LSTM (direta)

Modelo	MAE	RelMAE	Comentários
LSTM iter 1	14,768	0,827	Duas camadas (32,64), 105 input, batch=105
LSTM iter 1	15,463	0,866	Duas camadas (32,64), 210 input, batch=45
LSTM iter 2	13,688	0,766	Uma camada (32), 100 epochs, batch = 45, 8 Features
LSTM iter 3	14,212	0,796	Duas camadas (16,32), 8 features
LSTM iter 4	13,113	0,734	Uma camada (128), 10 epochs, batch=45, 8 Features
LSTM iter 5	12,954	0,725	Duas camadas (128,16), 8 features
LSTM iter 7	13,329	0,746	Uma camada (256), 20 epochs, adam, batch=105
LSTM iter 7	13,052	0,731	MAE como <i>loss function</i>
LSTM iter 7	12,832	0,718	50 epochs
LSTM iter 9	12,381	0,693	Duas camadas (32,64), 105 input, learn rate 0,0001, MAE
LSTM iter 10	14,090	0,789	Três camadas (64,64,32), 105 input, learn rate 0,0001, MAE

Nesta vertente de estudo, o *learning rate* começou a ter uma maior preponderância, uma vez que foi utilizado várias vezes para estabilizar o processo de aprendizagem, “esticando” o tempo de aprendizagem de forma que o modelo não estagnasse tão facilmente. O modelo que apresentou melhores resultados, começou por ser um caso de claro *overfitting*, como comprova o gráfico da Figura 4.15. Apesar do conjunto de treino apresentar

uma clara continuação de aprendizagem e adaptação por parte do modelo, o conjunto de validação começou a perder capacidade, dando a entender que se estava a adaptar demasiado aos dados de treino não tendo depois agilidade para lidar com dados novos.

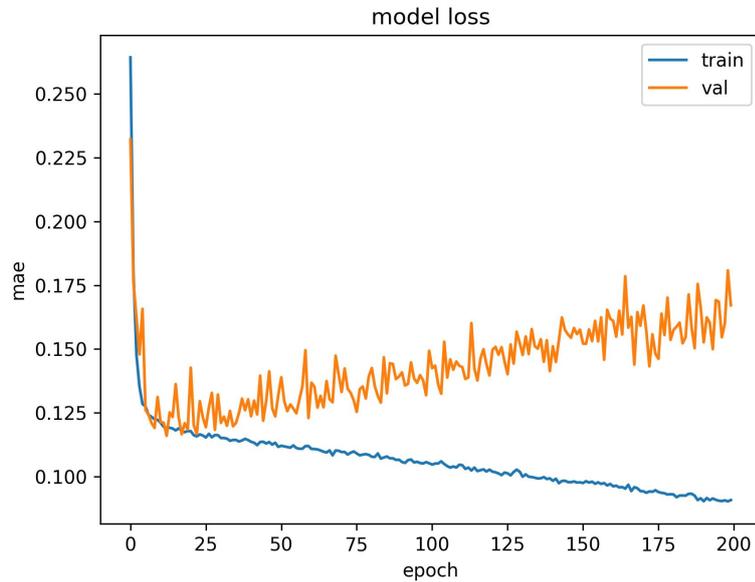


FIGURA 4.15. *Loss function de treino e validação de uma rede LSTM (iterativa)*

Neste caso optou-se então por diminuir o número de *epochs* e reduzir o *learning rate* do valor *default* 0,001 para 0,0001, alcançando então o melhor modelo desta abordagem, com um MAE de 12,381 e um RelMAE de 0,693 tornando-se o único modelo LSTM com uma performance abaixo dos 0,7 no RelMAE. A Figura 4.16 mostra a sua performance na semana 8 do conjunto de teste.

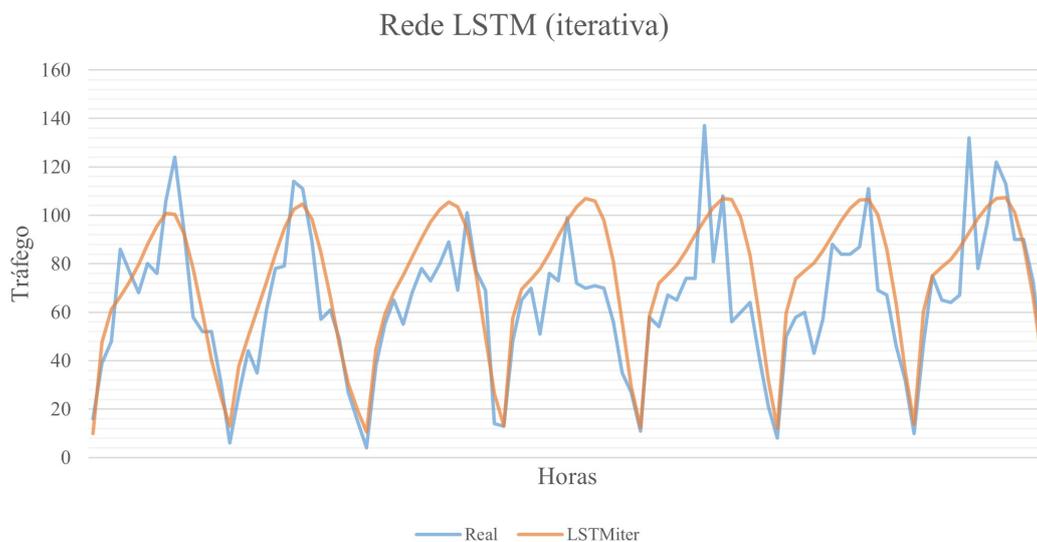


FIGURA 4.16. *Previsão do modelo LSTM (iterativa) na semana 8 do conjunto de teste*

Modelo *Ensemble*

Com todos os modelos testados, procedeu-se então à criação do modelo *ensemble*, um dos pontos de maior interesse no desenvolvimento deste estudo. Neste momento o problema foi abordado como se se tratasse de um problema de otimização, disponibilizando todos os dados de previsões para que estes trabalhassem com objetivo de reduzir o MAE. Foram testadas várias possibilidades e a Tabela 4.9 mostra os resultados obtidos para as que obtiveram resultados mais satisfatórios.

TABELA 4.9. Resultados das experiências *Ensemble*

<i>ExpSmoothing</i>	<i>XGBoost</i>	<i>LSTM iter</i>	<i>RandomForest</i>	MAE	<i>RelMAE</i>
0%	50%	50%	0%	11,567	0,647
0%	70%	30%	0%	11,515	0,645
0%	0%	50%	50%	11,851	0,663
5%	65%	30%	0%	11,493	0,643

Um dos destaques vai para o modelo que faz a média entre o modelo de previsões iterativas LSTM e o modelo *Random Forest*, em que cada um deles obteve individualmente MAEs de 12,381 e 12,575 respetivamente, mas que quando combinados alcançam um MAE de 11,851 que corresponde a um RelMAE de 0,663. Em relação ao melhor modelo possível, chegou-se a um *ensemble* de três modelos diferentes. A previsão composta por 5% de *Exponential Smoothing*, 65% de XGBoost e 30% LSTM iterativo obteve um MAE de 11,493 que corresponde por sua vez a um RelMAE de 0,643. A Figura 4.17 mostra a *performance* deste último modelo na semana 3 do conjunto de teste.

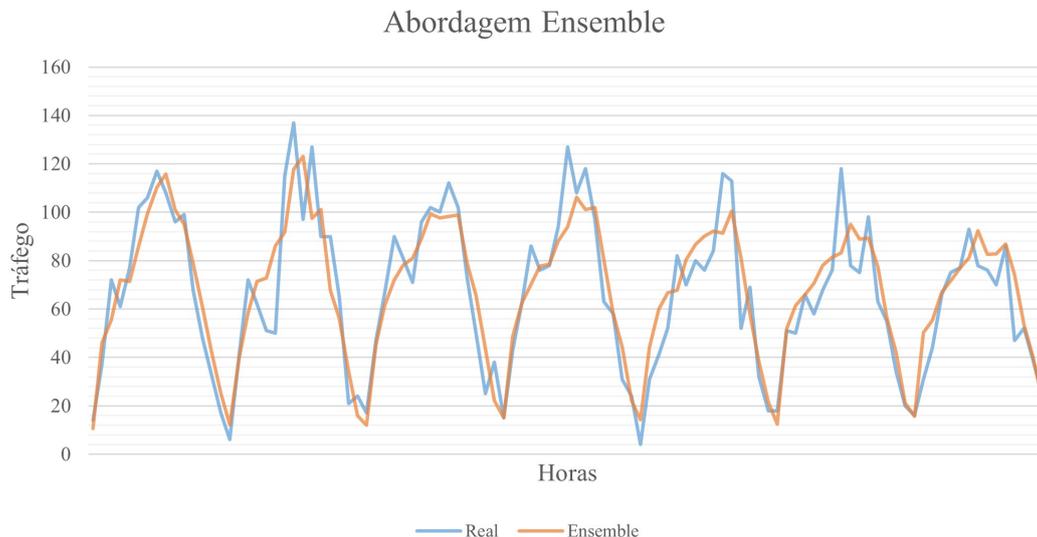


FIGURA 4.17. Previsão do modelo *Ensemble* na semana 3 do conjunto de teste

4.3. Discussão

Depois de apurados e testados todos os modelos esta é a lista final de modelos para previsão de tráfego horário (ver Tabela 4.10).

TABELA 4.10. Modelos e resultados finais das várias experiências

Modelo	MAE	Rel MAE
<i>Naive Forecast (baseline)</i>	17,864	1,000
<i>Seasonal Window Average (15h)</i>	14,210	0,795
<i>Seasonal Window Average (105h)</i>	14,690	0,822
<i>SARIMA</i>	16,363	0,916
<i>Random Forest</i>	12,575	0,704
<i>Exponential Smoothing</i>	16,090	0,901
Prophet	13,308	0,745
<i>XGBoost</i>	11,749	0,658
<i>LSTM (direta)</i>	13,334	0,746
<i>LSTM (iterativa)</i>	12,381	0,693
<i>Ensemble (ExpSmooth + XGB + LSTM)</i>	11,493	0,643

A primeira nota a salientar é o facto de todos os modelos terem conseguido diminuir erro face ao modelo *baseline*, ou seja, todos obtiveram um valor de RelMAE inferior a 1, considerada a *baseline*.

Dada a fórmula do RelMAE, descrita no capítulo da Metodologia, podemos interpretar o seu valor como sendo a percentagem de erro que não conseguimos retirar do modelo *baseline*, ou de forma inversa, a diferença entre o RelMAE é a percentagem de erro que conseguimos diminuir com o modelo em causa. Seguindo esta linha de pensamento, podemos afirmar que com o modelo *ensemble*, o melhor modelo do estudo, conseguimos diminuir o erro em 35,7%, que é um valor bastante significativo sabendo que a intenção é que estes valores sirvam de input para decisões de gestão relacionadas com a organização de turnos, *marketing*, gestão de equipas e produto.

O segundo ponto a realçar, numa perspetiva mais negativa é a falta de impacto que algumas variáveis exógenas tiveram, nomeadamente, a variável “Jogo” relativa a eventos desportivos e culturais perto da Loja Alfa. Apesar de haver um claro sentido lógico por trás da crença de que esta seria uma variável importante, a realidade é que não foi acrescentado o valor esperado aos modelos.

Para terminar, salientar o facto da abordagem através de modelos de regressão (via algoritmos de aprendizagem máquina) ter apresentado resultados tão favoráveis. Facilmente podem ser postos de lados os modelos que não suportam séries temporais, no entanto modelos como *Random Forest* ou *XGBoost* têm estado presentes em vários artigos de investigação e também neste estudo acabaram por acrescentar bastante valor, sendo que o *XGBoost* acabou mesmo por ser o modelo individual que melhor resultado obteve.

CAPÍTULO 5

Conclusões

A presente dissertação tinha como objetivo fornecer previsões de tráfego que servissem de ferramenta às equipas de gestão da rede de lojas MEO. Estas previsões foram pensadas de duas perspetivas. A primeira perspetiva diz respeito a previsão de tráfego para um horizonte temporal de vinte e quatro meses, com frequência mensal, para cada uma das lojas da rede e para a rede de lojas MEO como um todo. Desta forma pretende-se que um gestor da rede MEO consiga ter um *input* válido de como é previsto que o tráfego das suas lojas se venha a desenvolver no futuro, bem como uma ideia de como os padrões de tráfego da rede como um todo se enquadram na estratégia da empresa, em termos de dotação, oferta e espaços (Fildes et al., 2022). A segunda perspetiva aborda o tema da organização de equipas mais ao pormenor, produzindo previsões para um horizonte temporal de uma semana com frequência horária. Com esta segunda abordagem, pretende-se que a organização de equipa se torne, cada vez mais, um processo *data-driven*, combatendo as dificuldades da indefinição associada à volatilidade horária tão característica no setor do retalho (Chuang et al., 2016).

Do capítulo Revisão de Literatura, foi possível concluir que a investigação não está ainda totalmente direcionada para a previsão de tráfego, preferindo ir na direção de outros indicadores que caracterizem a procura, como as vendas (Lam et al., 1998), tendo este sido um ponto encarado também como um desafio no decorrer desta dissertação. A verdade é que existem redes de lojas, como no caso do retalho de Telecomunicações, em que as vendas não são o fator principal que caracteriza a procura, sendo necessário abordar essa questão de forma diferente, procurando através do tráfego preparar recursos para oferecer um atendimento de excelência a qualquer cliente, independentemente do motivo que o leve a visitar uma loja física.

Em relação ao tráfego mensal, a grande dificuldade do estudo passou por perceber qual a melhor forma de lidar com os intervalos anómalos decorrentes do abalo no tráfego, sentido durante a pandemia COVID-19. A abordagem de substituição dos intervalos através de interpolação apresentou melhores resultados, quando comparada com a abordagem com uso do modelo Prophet. No que toca às previsões produzidas, o modelo SARIMA obteve clara vantagem na capacidade de generalização para várias lojas, sendo inclusive utilizado o modelo SARIMA para previsão do tráfego na rede de lojas MEO como um todo. Naturalmente, as lojas com mais tráfego têm dados mais estáveis em termos de volatilidade, o que permitiu que o modelo produzisse melhores previsões quando comparadas com as previsões obtidas para lojas com menor tráfego mensal. O MAPE obtido para cada loja, apresentado na tabela 4.1, foi bastante satisfatório, sendo que para esta

abordagem, a rede de lojas não contava ainda com nenhuma ferramenta que produzisse este tipo de inputs. No que toca, em particular, à previsão de tráfego da rede de lojas para os próximos vinte e quatro meses, esta incorpora, já hoje, o relatório mensal da MEO, sendo atualizado todos os meses de forma a servir de *input* para a gestão de topo do grupo Altice. A figura 5.1 ilustra como tem sido a *performance* do modelo até à data, para as previsões a um mês e respetivas previsões futuras, estando à data com um MAPE de 2,74% para o ano de 2023.

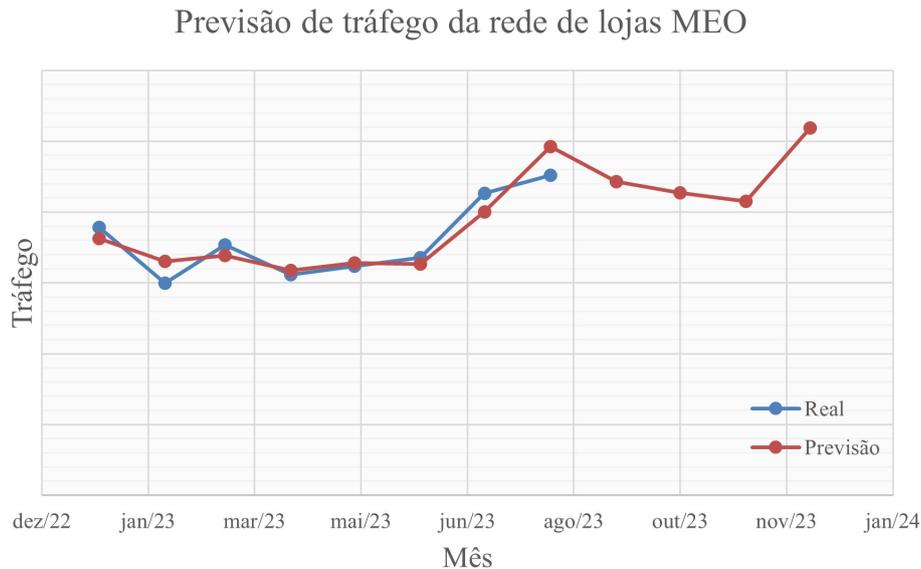


FIGURA 5.1. Previsão de tráfego mensal da rede de lojas MEO para 2023

No caso da previsão de tráfego com periodicidade horária, a abordagem foi diferente. Por haver já, na empresa, uma utilização frequente dos dados passados para gerir os períodos futuros, foi definido que o modelo *Naive Forecast* serviria de *baseline* sobre a qual todos os modelos seriam avaliados, com uso da métrica RelMAE, como a abordagem utilizada no artigo (Sousa et al., 2022). Este é um tema de relevância significativa uma vez que o tema do tempo de espera e qualidade de atendimento são focos principais na rede de lojas MEO que, por conhecimento do mercado, sabemos que são altamente influenciados pela capacidade da equipa de loja disponível aos clientes. Este caso, com base no artigo (Abrishami et al., 2017) e na recolha de feedback de elementos das equipas da rede MEO, foi estudado com uso de variáveis exógenas, nomeadamente feriados, condições climatéricas e eventos externos.

Começámos por definir o erro do método existente, neste caso *Naive Forecast*, que produz previsões com MAE de 17,864. Foram depois feitas experiências com os modelos *Seasonal Window Average*, SARIMA, *Random Forest*, *Exponential Smoothing*, Prophet, XGBoost e redes LSTM, sendo que cada um destes modelos foi analisado individualmente, tendo em conta as suas particularidades e híper parâmetros, para que pudessem alcançar uma melhor *performance*. De entre estas experiências destacam-se as abordagens nos

modelos *Random Forest* e XGBoost que, para além de usarem as variáveis bases, usaram também variáveis criadas para inserir nas árvores de decisão inputs que dessem informação da série temporal, nomeadamente t-1 (valor da hora anterior), t-15 (valor de há um dia) e t-105 (valor de há uma semana). Destaca-se também a diferença entre resultados obtidos para os dois tipos de estratégias *multistep-ahead* (An & Anh, 2015), aplicadas nas redes LSTM. A rede LSTM cuja abordagem seguiu a estratégia direta obteve um RelMAE de 0,746 (MAE de 13,334), enquanto a rede LSTM com abordagem com estratégia iterativa obteve um RelMAE de 0,693 (MAE de 12,381).

Para finalizar o estudo dos diferentes modelos, e com base no artigo (Sousa et al., 2022), foi testada uma abordagem ensemble que juntou vários modelos. De entre as várias experiências, destaca-se a que juntou *Exponential Smoothing*, XGBoost e a rede LSTM (iterativa), que obteve uma melhor *performance* que qualquer um dos modelos individualizadas, acabando por ser a abordagem com mais sucesso da presente dissertação, obtendo um RelMAE de 0,643 (MAE de 11,492).

Este é um resultado muito positivo e que cumpre o propósito desta dissertação. Obtendo um RelMAE de 0,643 este modelo melhora, em média, 35,7% do erro obtido com o modelo usado atualmente na rede de lojas MEO. Este é um valor bastante significativo e que permite que esta ferramenta seja vista como prometedora para o futuro da gestão das equipas de loja.

Fazendo agora uma retrospectiva de todo processo, em termos de dificuldades, durante a fase *Modeling*, o tempo computacional necessário para correr as diferentes experiências um dele acabou por constituir um grande desafio. Este foi um tema que teve de ser abordado, tal como na gestão de recursos, como um equilíbrio entre o recurso (tempo) e o proveito percebido (*performance*), para que não se deixassem experiências importantes para trás mas também não compromettesse a execução do projeto como suposto. Também as variáveis exógenas se mostraram um desafio, não para o presente estudo, mas para um projeto sustentável de futuro. Variáveis como as condições climáticas são muito difíceis de captar para períodos de frequência horária que tenham um horizonte temporal superior a 5 ou 6 dias. Sendo que este estudo foi levado a cabo para um horizonte temporal de previsão de 7 dias, e sendo que foram trabalhados períodos passados, este caso não se interpôs como um entrave, no entanto, não deixa de ser um obstáculo a analisar no futuro. Também os eventos externos carecem de alguma atenção. Ao não existirem bases de dados já estruturadas com as datas dos mesmos, a criação de bases de dados de raiz acabou por ser necessária, sendo também este um desafio para uma eventual aplicação futura dos modelos (solução eventualmente possível com técnicas de *webscraping*).

Em relação a estudos futuros, esta dissertação deixa duas linhas de investigação em aberto. A primeira diz respeito à profundidade do estudo de cada um dos modelos usado nas diferentes experiências. Esta dissertação, mais do que profundidade de investigação num dado modelo, procurou criar uma base diversa de modelos que pudesse também ser experimentada numa abordagem *ensemble*, deixando, portanto, em aberto, agora com

mais dados sobre a performance de cada um deles no contexto desejado, um estudo mais profundo de forma a melhorar a sua performance. A segunda linha de investigação e, provavelmente, a mais relevante das duas, prende-se com o facto desta ferramenta ter uma potencialidade muito grande quando aplicada a um horizonte temporal a prever de um mês. No entanto, este é um estudo que apresenta grandes desafios, uma vez que estaríamos a tratar um problema de *multistep-ahead* com 720 *timesteps* (ou 450 *timesteps*, se contarmos apenas com as 15 horas do dia em que a Loja Alfa está aberta), o que colocaria em causa o uso de algumas variáveis externas, como condições climatéricas.

Referências Bibliográficas

- Abrishami, S., & Kumar, P. (2019). Using Real-World Store Data for Foot Traffic Forecasting. *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*, 1885–1890. <https://doi.org/10.1109/BigData.2018.8622551>
- Abrishami, S., Kumar, P., & Nienaber, W. (2017). Smart stores: A scalable foot traffic collection and prediction system. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10357 LNAI, 107–121. https://doi.org/10.1007/978-3-319-62701-4_9
- Adhikari, R., & Agrawal, R. K. (2012). A Novel Weighted Ensemble Technique for Time Series Forecasting (pp. 38–49). https://doi.org/10.1007/978-3-642-30217-6_4
- An, N. H., & Anh, D. T. (2015). Comparison of Strategies for Multi-step-Ahead Prediction of Time Series Using Neural Network. *2015 International Conference on Advanced Computing and Applications (ACOMP)*, 142–149. <https://doi.org/10.1109/ACOMP.2015.24>
- Bates, J. M., & Granger, C. W. J. (1969). The Combination of Forecasts. *Journal of the Operational Research Society*, 20(4), 451–468. <https://doi.org/10.1057/jors.1969.103>
- Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The Role of Big Data and Predictive Analytics in Retailing. *Journal of Retailing*, 93(1), 79–95. <https://doi.org/10.1016/j.jretai.2016.12.004>
- Chandra, R., Goyal, S., & Gupta, R. (2021). Evaluation of Deep Learning Models for Multi-Step Ahead Time Series Prediction. *IEEE Access*, 9, 83105–83123. <https://doi.org/10.1109/ACCESS.2021.3085085>
- Chuang, H. H. C., Oliva, R., & Perdikaki, O. (2016). Traffic-Based Labor Planning in Retail Stores. *Production and Operations Management*, 25(1), 96–113. <https://doi.org/10.1111/poms.12403>
- Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4), 559–583. [https://doi.org/10.1016/0169-2070\(89\)90012-5](https://doi.org/10.1016/0169-2070(89)90012-5)
- Cortez, P., Matos, L. M., Pereira, P. J., Santos, N., & Duque, D. (2017). Forecasting Store Foot Traffic Using Facial Recognition, Time Series and Support Vector Machines. In *Advances in Intelligent Systems and Computing* (Vol. 527, pp. 267–276). https://doi.org/10.1007/978-3-319-47364-2_26
- Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, 38(4), 1283–1318. <https://doi.org/10.1016/j.ijforecast.2019.06.004>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441. <https://doi.org/10.1037/h0071325>
- Junior, C., Gusmão, P., Moreira, J., & Tome, A. M. M. (2021). Time series forecasting in retail sales using LSTM and prophet. In *Handbook of Research on Applied Data Science and Artificial Intelligence in Business and Industry* (pp. 241–262). IGI Global. <https://doi.org/10.4018/978-1-7998-6985-6.ch011>
- Kumar, K. V., & Anitha, R. (2022). A Novel Ensemble Model by Combining LSTM, BiLSTM, and Facebook Prophet Algorithm to Forecast Stock Prices. *2022 Third International Conference*

- on *Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, 1044–1047. <https://doi.org/10.1109/ICICICT54557.2022.9917634>
- Lam, S., Vandenbosch, M., & Pearce, M. (1998). Retail Sales Force Scheduling Based on Store Traffic Forecasting. *Journal of Retailing*, 74(1).
- Larson, R. C. (1987). PERSPECTIVES ON QUEUES: SOCIAL JUSTICE AND THE PSYCHOLOGY OF QUEUEING. *Operations Research*, 35(6), 895–905. <https://doi.org/10.1287/opre.35.6.895>
- Lee, H. S., Kesavan, S., & Deshpande, V. (2021). Managing the impact of fitting room traffic on retail sales: Using labor to reduce phantom stockouts. *Manufacturing and Service Operations Management*, 23(6), 1580–1596. <https://doi.org/10.1287/msom.2020.0884>
- Livieris, I. E., & Pintelas, P. (2022). A novel multi-step forecasting strategy for enhancing deep learning models' performance. *Neural Computing and Applications*, 34(22), 19453–19470. <https://doi.org/10.1007/s00521-022-07158-9>
- Lu, G., Du, R. Y., & Peng, X. (2022). The Impact of Schedule Consistency on Shift Worker Productivity: An Empirical Investigation. *Manufacturing and Service Operations Management*, 24(5), 2780–2796. <https://doi.org/10.1287/msom.2022.1132>
- Maçaira, P. M., & Cyrino Oliveira, F. L. (2016). Another look at SSA.Boot forecast accuracy. In *International Journal of Energy and Statistics* (Vol. 04, Issue 02). <https://doi.org/10.1142/s2335680416500083>
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54–74. <https://doi.org/10.1016/j.ijforecast.2019.04.014>
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4), 1346–1364. <https://doi.org/10.1016/j.ijforecast.2021.11.013>
- Panay, B., Baloian, N., Pino, J. A., Peñafiel, S., Frez, J., Fuenzalida, C., Sanson, H., & Zurita, G. (2021). Forecasting key retail performance indicators using interpretable regression. *Sensors*, 21(5), 1–18. <https://doi.org/10.3390/s21051874>
- Perdikaki, O., Kesavan, S., & Swaminathan, J. M. (2012). Effect of traffic on sales and conversion rates of retail stores. *Manufacturing and Service Operations Management*, 14(1), 145–162. <https://doi.org/10.1287/msom.1110.0356>
- Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A., ... Ziel, F. (2022). Forecasting: theory and practice. In *International Journal of Forecasting* (Vol. 38, Issue 3, pp. 705–871). Elsevier B.V. <https://doi.org/10.1016/j.ijforecast.2021.11.001>
- Seaman, B. (2018). Considerations of a retail forecasting practitioner. *International Journal of Forecasting*, 34(4), 822–829. <https://doi.org/10.1016/j.ijforecast.2018.03.001>
- Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing*, 5(4), 13–22.
- Sousa, M., Tomé, A. M., & Moreira, J. (2022). Long-term forecasting of hourly retail customer flow on intermittent time series with multiple seasonality. *Data Science and Management*, 5(3), 137–148. <https://doi.org/10.1016/j.dsm.2022.07.002>
- Van Den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., & De Boeck, L. (2013). Personnel scheduling: A literature review. *European Journal of Operational Research*, 226(3), 367–385. <https://doi.org/10.1016/j.ejor.2012.11.029>

- Wang, X., Liono, J., McIntosh, W., & Salim, F. D. (2017). Predicting the city foot traffic with pedestrian sensor data. *ACM International Conference Proceeding Series*, 1–10. <https://doi.org/10.1145/3144457.3152355>
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1–3), 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)