



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ (PUCPR)
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA (PPGIA)

Predição de Picos em Séries Temporais com Aplicação em Redes Elétricas Inteligentes

Rodrigo Krüger

Curitiba - Paraná

2023



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ (PUCPR)
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA (PPGIA)

Predição de Picos em Séries Temporais com Aplicação em Redes Elétricas Inteligentes

Rodrigo Krüger

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná, como parte dos requisitos para a obtenção do título de Mestre em Informática.

Área de concentração: Ciência de Dados

Orientador: Vinícius Mourão Alves de Souza

Curitiba - Paraná

2023

Abstract

The growing popularity of electronic devices and sensors in general has been responsible for the generation of a huge amount of sequential data called time series. These data are widely used by machine learning algorithms to perform different tasks in applications in the areas of health, biology, energy, among others. Among different possibilities, a task that is still little explored is the prediction of peaks. Peaks indicate significant events in a time series, such as a sudden increase in the volume of sales of a product at a certain time of the year or a sudden increase in the pressure of a river given the increase in the water level. Thus, the prediction of peaks with a considerable time in advance can be important information for decision-making, such as planning stock replacement or the development of a flood warning system. The peak prediction task involves predicting both the magnitude and the time at which a peak will occur in future observations of a time series. However, most studies consider the prediction of only one of these values or the prediction of values independently. This project investigates *multi-target* solutions that consider the correlation between the target attributes of a problem to improve predictions. In addition, the work evaluates different descriptive characteristics for the construction of predictive models in a smart grid problem. In this application, the goal is to predict, a week in advance, which day of the week and which will be the peak of the daily electricity consumption of a given customer according to its historical data, making it possible to use this information in overload warning systems, savings recommendation and dynamic pricing solutions.

Keywords: Peak prediction, time series, multi-target regression, extrinsic regression, energy consumption, smart grids.

Resumo

A crescente popularidade de dispositivos eletrônicos e sensores em geral, tem sido responsável pela geração de uma enorme quantidade de dados sequenciais denominados séries temporais. Estes dados são amplamente utilizados por algoritmos de aprendizagem de máquina, para a realização de diferentes tarefas em aplicações nas áreas da saúde, biologia, energia, entre outras. Dentre diferentes possibilidades, uma tarefa que ainda é pouco explorada é a predição de picos. Picos indicam eventos significativos em séries temporais, como o aumento repentino no volume de vendas de um produto em uma determinada época do ano ou a elevação súbita na pressão de um rio dado o aumento do nível da água. Assim, a predição de picos com um tempo considerável de antecedência pode ser uma importante informação para a tomada de decisões, como planejamento de reposição de estoque ou o desenvolvimento de um sistema de alerta de enchentes. A tarefa de predição de picos envolve prever tanto a magnitude como o instante em que um pico irá ocorrer nas observações futuras de uma série temporal. Entretanto, a maior parte dos trabalhos considera a predição de somente um destes valores ou a predição dos valores de maneira independente. Este projeto investiga soluções *multi-target* que consideram a correlação entre os atributos alvo de um problema para a melhoria das predições. Além disso, o trabalho avalia diferentes características descritivas para a construção dos modelos preditivos em um problema de redes elétricas inteligentes. Nesta aplicação, o objetivo é prever com uma semana de antecedência, qual dia da semana e qual será o pico de consumo diário de energia elétrica de um dado cliente de acordo com seus dados históricos, sendo possível utilizar esta informação em sistemas de alerta de sobrecarga, recomendação de economia e tarifação dinâmica.

Palavras-chave: Predição de picos, séries temporais, regressão *multi-target*, regressão extrínseca, consumo de energia, redes elétricas inteligentes.

Agradecimentos

O Mestrado é visto como uma oportunidade para desenvolver habilidades de pesquisa, análise crítica e resolução de problemas complexos, características altamente valorizadas no mercado de trabalho. Para alguns, é um caminho para se destacar em suas carreiras e alcançar novas oportunidades profissionais. Para mim, o Mestrado vai além: é a realização de um sonho. Independentemente da motivação específica, a busca por um Mestrado é um marco significativo na trajetória acadêmica e profissional, impulsionada pelo desejo de crescimento pessoal e contribuição para o progresso da sociedade.

A conquista de um projeto do porte de um Mestrado é resultado de uma enorme jornada colaborativa, na qual o apoio e o envolvimento de diversas pessoas são elementos fundamentais para o sucesso alcançado. Desde o momento da concepção da pesquisa até a sua conclusão, a contribuição de professores, orientadores, colegas, familiares e amigos se faz presente, desempenhando papéis essenciais ao longo de todo o percurso. As trocas de conhecimento, as orientações recebidas, os estímulos emocionais e o suporte incondicional constituem pilares que sustentam o desenvolvimento deste projeto acadêmico complexo e desafiador. É na conjunção desses esforços, na interação entre mentes dedicadas e no compartilhamento de ideias, que se edifica o alicerce sólido para o êxito do Mestrado, destacando que nenhuma conquista de grande envergadura é fruto do esforço solitário, mas sim do trabalho colaborativo e do espírito coletivo.

Aos meus queridos pais, Alexandre Krüger Júnior e Edite Fátima Krüger, agradeço pelo amor incondicional e apoio ao longo de toda a minha trajetória acadêmica. Foi por intermédio das orientações e influências de vocês que escolhi a área de tecnologia da informação para exercer minha profissão, que agora é complementada com este trabalho de Mestrado.

Agradeço à minha amada esposa, Ana Claudia Serrato Krüger, por seu amor incondicional, apoio incansável e compreensão durante toda a minha jornada de Mestrado. Sua presença ao meu lado foi um farol de esperança e motivação, tornando os momentos desafiadores mais leves para que as conquistas fossem mais significativas. Seu incentivo constante e o seu encorajamento foram fundamentais para manter minha determinação ao enfrentar as demandas acadêmicas. Sou privilegiado por tê-la como minha parceira de vida e tenho a felicidade de dividir esta conquista com você.

Ao estimado orientador deste trabalho de Mestrado, Prof. Vinícius Mourão Alves de Souza. Sua notável perspicácia e vasto conhecimento na área foram fundamentais para alicerçar meu desenvolvimento acadêmico e aprimorar o conteúdo desta dissertação. Além disso, sua inestimável paciência em escutar minhas indagações e inquietações, bem como suas orientações cuidadosas e motivadoras, foram cruciais para o êxito deste trabalho. Sua dedicação incansável em acompanhar cada etapa do processo, sempre demonstrando disponibilidade e interesse genuíno pelo meu crescimento intelectual, foi inspiradora e enriquecedora.

À Pontifícia Universidade Católica do Paraná (PUC-PR) pela excelência e liderança exemplar na área de educação, que a tornam uma referência em seu segmento. Sou grato por ter tido a oportunidade de fazer parte desta comunidade acadêmica e de ser beneficiado pelo alto padrão de ensino, corpo docente qualificado e recursos de aprendizagem que me permitiram aprofundar meu conhecimento em minha área de estudo. É com imensa gratidão que reconheço a PUC-PR como uma instituição inspiradora, que promove o crescimento intelectual e profissional de seus estudantes de maneira excepcional.

Sumário

Abstract	i
Resumo	ii
Agradecimentos	iii
Lista de Tabelas	viii
Lista de Figuras	xi
1 Introdução	1
1.1 Justificativa e motivação	3
1.2 Objetivos e hipótese	4
1.2.1 Principais resultados e contribuições	5
1.3 Organização	6
2 Fundamentação Teórica	7
2.1 Considerações iniciais	7
2.2 Redes elétricas inteligentes	7
2.3 Séries temporais	9
2.4 Predição de picos de consumo	11
2.5 Regressão extrínseca	15
2.5.1 Aplicações	16
2.5.2 Algoritmos	17
2.6 Regressão <i>multi-target</i>	19

2.6.1	Métodos globais e locais	20
2.6.2	ERC	21
2.6.3	SST	22
2.6.4	DSTARS	23
2.7	Considerações finais	24
3	Trabalhos Relacionados	25
3.1	Considerações iniciais	25
3.2	Predição de picos de consumo de energia	25
3.3	Predição de picos em diferentes aplicações	29
3.4	Discussão	32
3.5	Considerações finais	33
4	Construção e Análise de Dados	34
4.1	Considerações iniciais	34
4.2	Dados de consumo de energia elétrica	34
4.3	Pré-processamento e construção dos dados	36
4.4	Correlação entre as variáveis-alvo	40
4.5	Análise dos dados	42
4.6	Considerações finais	44
5	Predição de picos de consumo diário de energia	45
5.1	Considerações iniciais	45
5.2	Objetivos	45
5.3	Metodologia de desenvolvimento do trabalho	47
5.4	Métodos de avaliação	49
5.5	Considerações finais	51
6	Avaliação Experimental	53
6.1	Considerações iniciais	53
6.2	Descrição do protocolo experimental	53

6.2.1	Linhas de base	54
6.2.2	Método de validação	55
6.3	Análise geral dos resultados	56
6.3.1	Janela de semanas	56
6.3.2	Matriz de características	60
6.3.3	Algoritmos	63
6.3.4	Melhor configuração	64
6.4	Modelos individuais vs. modelo global	68
6.5	Análise individual dos resultados	70
6.5.1	Predição da magnitude do pico	72
6.5.2	Predição da posição do pico	76
6.6	Respostas às questões de pesquisa	80
6.7	Considerações finais	83
7	Conclusões	85
7.1	Considerações iniciais	85
7.2	Limitações	85
7.3	Trabalhos futuros	86
	Referências	86

Lista de Tabelas

5.1	Combinações de características utilizadas nos experimentos de predição.	49
6.1	Algoritmos utilizados nos experimentos e seus parâmetros.	55
6.2	Resultados de MAE para a tarefa de predição da magnitude do pico de consumo de energia elétrica.	75
6.3	Resultados de MSE para a tarefa de predição da magnitude do pico de consumo de energia elétrica.	75
6.4	Resultados de RMSE para a tarefa de predição da magnitude do pico de consumo de energia elétrica.	75
6.5	Resultados de MAPE para a tarefa de predição da magnitude do pico de consumo de energia elétrica.	76
6.6	Resultados de Acurácia para a tarefa de predição da posição do pico de consumo de energia elétrica.	79
6.7	Resultados de Precisão para a tarefa de predição da posição do pico de consumo de energia elétrica.	79
6.8	Resultados de Recall para a tarefa de predição da posição do pico de consumo de energia elétrica.	79
6.9	Resultados de F1-score para a tarefa de predição da posição do pico de consumo de energia elétrica.	80

Lista de Figuras

2.1	Exemplo de série temporal que representa o comportamento de um papel na bolsa de valores.	9
2.2	Predição da magnitude e posição do pico de consumo de energia com 7 dias de antecedência.	12
2.3	Exemplo de framework para predição de picos de energia.	13
2.4	Métodos de predição de picos de consumo.	14
2.5	Representação gráfica do treinamento do ERC	21
2.6	Representação gráfica do treinamento do SST	22
2.7	Representação gráfica do treinamento do DSTARS	23
4.1	Localização da cidade de Los Alamos, NM, EUA, onde os dados de consumo de energia elétrica foram coletados e fornecidos pela concessionária LADPU.	35
4.2	Medidor inteligente de energia (<i>smart meter</i>) Landis+Gyr E470.	35
4.3	Exemplo de dados de consumo de energia.	36
4.4	Primeiros registros do arquivo de dados de consumo obtidos pelo medidor identificado como 49991.	37
4.5	Transformação da série de consumo instantâneo em uma série com o consumo acumulado diário.	38
4.6	Janela deslizante que percorre toda a série temporal para formação da matriz de características.	39
4.7	Matriz de características construída a partir dos dados de consumo.	40
4.8	Correlação entre os atributos-alvo observada entre os 100 consumidores.	41
4.9	Matriz de características construída a partir dos picos.	42
4.10	Distribuição da posição do pico para a base de 100 consumidores.	43

4.11	Distribuição da magnitude do pico para a base de 100 consumidores.	43
4.12	Análise da magnitude do pico em função do horário.	43
4.13	Magnitude dos picos de acordo com os meses e dias da semana. . . .	44
6.1	Distribuição das métricas MAE e RMSE conforme a janela de semanas, para predição da magnitude.	57
6.2	Distribuição das métricas acurácia, precisão, recall e f1-score conforme a janela de semanas, para predição da posição.	57
6.3	Distribuição do erro médio absoluto por algoritmo, conforme a janela de semanas, para predição da magnitude.	58
6.4	Distribuição da acurácia por algoritmo, conforme a janela de semanas, para predição da posição.	59
6.5	Distribuição das métricas MAE e RMSE conforme a matriz de características para predição da magnitude e considerando as diferentes abordagens de algoritmos de predição.	61
6.6	Distribuição da acurácia e f1-score conforme a matriz de características para predição da posição e considerando as diferentes abordagens de algoritmos de predição.	62
6.7	Distribuição das métricas MAE e RMSE conforme o algoritmo para predição da magnitude, utilizando métodos <i>single-target</i>	64
6.8	Distribuição da acurácia e f1-score conforme o algoritmo para predição da posição, utilizando métodos <i>single-target</i>	65
6.9	Distribuição das métricas MAE e RMSE conforme o algoritmo para predição da magnitude, utilizando métodos <i>multi-target</i>	66
6.10	Distribuição da acurácia e f1-score conforme o algoritmo para predição da posição, utilizando métodos <i>multi-target</i>	67
6.11	Distribuição das métricas MAE e RMSE dos melhores algoritmos para predição da magnitude.	69
6.12	Diferença crítica dos melhores algoritmos para predição da magnitude.	70
6.13	Distribuição da acurácia e f1-score dos melhores algoritmos para predição da posição.	71
6.14	Diferença crítica dos melhores algoritmos avaliados para predição da posição.	72
6.15	Distribuição do MAE dos algoritmos comparados a sua versão de treinamento a partir de dados de todos os consumidores.	73

6.16	Distribuição da acurácia dos algoritmos comparados a sua versão de treinamento a partir de dados de todos os consumidores.	74
6.17	Gráficos de dispersão das predições da magnitude do pico do algoritmo Linear Regression para quatro consumidores.	77
6.18	Séries temporais das predições da magnitude do pico do algoritmo Linear Regression para quatro consumidores.	78
6.19	Matriz de confusão das predições da posição do pico do algoritmo XGBoost para quatro consumidores.	81
6.20	Série temporal das predições da posição do pico do algoritmo XGBoost para quatro consumidores.	82

Capítulo 1

Introdução

A maioria dos fenômenos observados na natureza podem ser continuamente medidos ao longo do tempo e representados por dados sequenciais denominados séries temporais. Por isso, dados de séries temporais estão presentes em problemas e aplicações de diferentes áreas de conhecimento, como medicina, astronomia, economia, entre outras áreas. Algumas destas aplicações são a identificação de arritmia cardíaca em eletrocardiogramas, detecção de intrusão em sistemas computacionais, a identificação de transações fraudulentas em cartões de crédito, classificação de eventos sísmicos, previsão de valores de ações na bolsa, entre muitas outras.

Devido ao baixo custo para obtenção de uma grande quantidade de dados temporais, principalmente em aplicações envolvendo sensores com uma alta taxa de amostragem, séries temporais têm sido amplamente utilizadas como entrada para algoritmos de aprendizagem de máquina e Mineração de Dados, para a realização de tarefas como classificação, agrupamento, descoberta de padrões recorrentes, detecção de anomalias e *forecasting* (Souza et al., 2014; Mueen et al., 2009; Aghabozorgi et al., 2015; Parmezan et al., 2019). Entretanto, grande parte dos algoritmos tradicionais não lidam bem com este tipo de dado, devido a características como a alta dimensionalidade e elevada quantidade de ruído (Ratanamahatana et al., 2005). Ou ainda, não fazem uso de características como alta correlação entre os exemplos para a obtenção de melhores resultados. Algoritmos tradicionais como Máquina de Vetor de Suporte (SVM - *Support Vector Machine*) (Cortes e Vapnik, 1995), Árvores de Decisão (Quinlan, 1993) e Redes Neurais Artificiais (ANN - *Artificial Neural Network*) (Jain et al., 1996), assumem que os exemplos são obtidos de maneira independente de uma mesma distribuição de dados (Dundar et al., 2007). Diferentemente de outros tipos de dados, as observações das séries temporais não são mutuamente independentes. Assim, um dos desafios ao utilizar estes dados em modelos preditivos é a incorporação de informações sobre a correlação entre os exemplos ou entre os diferentes atributos preditivos.

Neste projeto de mestrado, é explorada a tarefa de predição de picos em dados de séries temporais. Picos indicam eventos significativos, como o aumento repentino no

volume de vendas de um produto em uma determinada época do ano ou o aumento significativo do valor de uma ação em uma determinada hora do dia. A partir de uma série temporal com observações de um fenômeno até o instante de tempo t , a tarefa consiste em prever a partir deste instante qual será o valor máximo (magnitude do pico) a ser observado nos valores futuros da série até o horizonte h , ou seja, prever qual será o valor máximo da série entre os instantes t e $t + h$. Além de prever o valor do pico, também se tem o interesse em prever simultaneamente qual será o instante de tempo i em que este valor irá ocorrer (posição do pico).

A predição de pico pode ser compreendida como uma tarefa de regressão, que utiliza dados de séries temporais como atributos de entrada e tem como objetivo induzir um modelo capaz de realizar o mapeamento entre uma sequência de valores contínuos e duas variáveis contínuas de saída - a magnitude do pico e sua posição. Recentemente, esta tarefa de regressão no domínio de séries temporais foi denominada *regressão extrínseca* (Tan et al., 2021). Enquanto algoritmos tradicionais de regressão não consideram a relação dos atributos com o tempo, a regressão extrínseca leva em consideração que os atributos de entrada são observações relacionadas entre si e que possuem uma organização temporal. Além disso, conforme discutido anteriormente, a tarefa proposta neste trabalho tem o objetivo de prever duas variáveis de saída simultaneamente: a magnitude e a posição do pico. Em regressão, quando existem múltiplas saídas relacionadas a um mesmo conjunto descritivo de atributos, denomina-se o problema de *regressão multi-target* ou *multi-output* (Borchani et al., 2015). Desse modo, a tarefa de predição de pico em séries temporais é abordada neste projeto de maneira inédita como um problema de *regressão extrínseca multi-target*.

É importante destacar que embora relacionada, a tarefa de predição de picos se diferencia de tarefas como a detecção de picos (Palshikar et al., 2009) e *forecasting* (Parmezan et al., 2019). Na tarefa de detecção de picos, o objetivo é localizar a posição dos picos em uma série temporal de consulta, não sendo uma tarefa de predição de valores futuros. Desse modo, um algoritmo de detecção de picos retorna um vetor com a posição de um ou mais máximos locais encontrados em uma série temporal de entrada. Neste contexto, um máximo local ou pico, é uma observação em que seus vizinhos apresentam magnitudes inferiores. Por sua vez, na tarefa de *forecasting*, o valor a ser predito representa uma extrapolação da série temporal de entrada e possui alta relação com as últimas observações da série. Por outro lado, na tarefa de predição proposta neste trabalho, não é possível assumir que a saída possua relação direta com os últimos valores observados nos dados de entrada. Entretanto, é importante observar que a tarefa de *forecasting* pode ser uma abordagem para a predição de picos a partir da previsão de um conjunto de valores futuros e a seleção da observação de maior valor.

1.1 Justificativa e motivação

A tarefa de predição de picos pode ser empregada em aplicações de domínios variados. Por exemplo, na área de saúde, a predição de picos de demanda no atendimento médico à pacientes de um hospital em um determinado dia ou horário pode ser utilizada para adequações na escala de trabalho de médicos e enfermeiros durante a semana. Em gestão urbana, a previsão de um grande volume de precipitação de chuva em um determinado horário do dia pode ser utilizada por sistemas de alerta de enchentes com horas de antecedência. Na área de tecnologia, a predição de picos de demanda em um centro de processamento de dados pode auxiliar no gerenciamento dos recursos disponíveis e definição de políticas de acesso. Em geral, boa parte das aplicações que lidam com o gerenciamento de recursos, serviços ou produtos com base em demanda, pode se beneficiar da tarefa de predição de picos investigada neste projeto.

Em específico, os métodos explorados e propostos neste projeto possuem aplicação direta em um problema de predição de pico de consumo de energia elétrica, utilizando dados de centenas de residências com medidores inteligentes de energia (*smart meters*). Os dados e o problema abordado são provenientes de uma parceria entre pesquisadores da Pontifícia Universidade Católica do Paraná (PUCPR), *University of New Mexico* (UNM) e *The New Mexico Established Program to Stimulate Competitive Research* (NM EPSCoR)¹ no âmbito do projeto *NM SMART Grid Center*, um projeto interdisciplinar financiado pela agência de fomento norte-americana *National Science Foundation* (NSF)², para a produção e entrega de energia elétrica da próxima geração com base no desenvolvimento científico e tecnológico.

Nesta aplicação, o objetivo é prever com dias de antecedência (por exemplo, 7 dias), quando e qual será o pico de consumo de energia diário de um dado cliente de acordo com seus dados históricos medidos por um *smart meter*. A predição da magnitude e dia da semana dos picos para um conjunto de consumidores é essencial para diferentes aplicações em redes elétricas inteligentes (*smart grids*), como sistemas de alerta e tarifação dinâmica. Por exemplo, se houver a previsão de que uma quantidade elevada de consumidores terá um consumo acima de um determinado valor simultaneamente em um mesmo dia, poderá ser emitido um alerta de possível sobrecarga. Este sistema poderá ser utilizado, por exemplo, para a definição de valores de tarifas de maneira dinâmica de acordo com a demanda e visando a redução da sobrecarga. Além disso, poderá auxiliar no gerenciamento da distribuição considerando diferentes fontes energéticas ou em sistemas de recomendação de economia.

¹<https://www.nmepscor.org/>

²RII Track-1: The New Mexico SMART Grid Center: Sustainable, Modular, Adaptive, Resilient, and Transactive – https://www.nsf.gov/awardsearch/showAward?AWD_ID=1757207

1.2 Objetivos e hipótese

O principal objetivo deste projeto é investigar e avaliar métodos de regressão para a tarefa de predição de picos em dados de séries temporais que representam o consumo de energia. Nesta tarefa, os métodos deverão ser capazes de prever de maneira antecipada e com precisão, tanto a magnitude como o dia da semana que o pico irá ocorrer nas observações futuras da série temporal de consumo. Tal predição deverá ser realizada a partir de um modelo preditivo, capaz de realizar o mapeamento de uma entrada de dados composta por uma sequência de valores contínuos (série temporal) para uma saída com duas variáveis, sendo uma contínua (magnitude) e outra discreta (posição, dia da semana). Nesta direção, os objetivos específicos deste trabalho são:

- Propor uma solução que seja independente da aplicação responsável pela geração dos dados. Para isso, os modelos serão treinados a partir de dados brutos de séries temporais ou a partir da extração de características genéricas dos dados, sem considerar o uso de fontes externas como informações meteorológicas no caso de dados de consumo de energia;
- Realizar uma ampla avaliação experimental, considerando diferentes abordagens para a construção de modelos de regressão para a tarefa de predição de picos, como modelos *single-target* e *multi-target*, rasos e profundos, treinados com dados de um único consumidor ou com dados de vários consumidores;
- Explorar diferentes configurações da aplicação de consumo de energia (atributos, tamanho dos exemplos, intervalos de antecedência) com o objetivo de identificar os melhores cenários para a tarefa de predição;
- Disponibilizar publicamente para a comunidade científica, os algoritmos implementados e os conjuntos de dados construídos durante o desenvolvimento deste projeto, visando estimular a reprodução dos resultados e a proposta de novas abordagens para o problema.

Tendo em vista o objetivo geral e específicos deste projeto, pode-se declarar a seguinte hipótese de pesquisa:

A adaptação de métodos de regressão extrínseca para que sejam capazes de gerar múltiplas saídas (multi-target), pode levar a resultados de predição de picos em séries temporais mais precisos que os métodos de regressão convencionais single-target, ao considerar a relação temporal dos dados de entrada e a correlação entre as variáveis de saída.

1.2.1 Principais resultados e contribuições

Uma das principais contribuições deste trabalho foi verificar a capacidade de duas abordagens (regressão extrínseca e regressão *multi-target*) para a tarefa de predição de picos em dados de consumo de energia. Nessa direção, este trabalho fornece uma discussão com base em uma avaliação experimental com métodos de aprendizagem de máquina rasos e profundos utilizando abordagens *single-target* e *multi-target*. Uma discussão destes resultados também é apresentada no artigo Kruger et al. (2023).

O trabalho contribuiu para afirmar que métodos de regressão *multi-target* não geram melhores resultados caso a correlação entre as variáveis de saída seja baixa, como é o caso da posição e magnitude dos picos de consumo diário de energia. Além disso, tais métodos são mais adequados em problemas com mais de duas variáveis de saída, em que existe maior probabilidade de correlação entre diferentes combinações de variáveis de saída e que podem ser exploradas pelos métodos para a melhoria das predições.

Sobre o tamanho dos exemplos (i.e., quantidade de atributos) utilizados para o treinamento dos modelos preditivos, o trabalho explorou janelas de dados de consumo de 1 a 4 semanas utilizando diferentes métodos de regressão rasos, profundos e *multi-target*. Com os experimentos foi concluído que os resultados, tanto de modelos rasos como profundos, pioram na medida que o comprimento da série de entrada aumenta, considerando o horizonte de predição de 1 semana à frente.

O trabalho também buscou explorar qual combinação de características proporciona os melhores resultados preditivos. Dentre as características avaliadas, estão a soma do consumo diário, o pico de consumo diário, o dia da semana do pico e diferentes combinações entre elas. Neste sentido o trabalho contribuiu para demonstrar que, em um contexto onde há pouca correlação entre as variáveis de saída, utilizar características diferentes para treinar modelos *single-target* para cada variável de saída, proporciona os melhores resultados na tarefa de predição de picos.

Embora tenha proporcionado melhores resultados, utilizar características diferentes para predição de cada variável de saída descaracteriza o principal propósito da técnica de regressão *multi-target*. Deste modo, foi identificada uma limitação dos métodos de regressão *multi-target*, que reduzem o desempenho de métodos rasos e profundos *single-target*, quando não há correlação entre as variáveis de saída.

Os resultados do trabalho mostraram também que modelos de predição construídos com dados de consumo de todos os consumidores - ou modelos globais - favoreceram a predição da posição, principalmente quando os modelos utilizam métodos *multi-target*. Porém, o mesmo comportamento não ocorre para predição da magnitude, reforçando a limitação destes métodos em gerar desempenho ótimo para todas as variáveis da aplicação.

Além disto, o trabalho buscou avaliar os resultados da predição da posição do pico, quando encarado como tarefas de classificação e regressão. Neste sentido,

o trabalho contribuiu para demonstrar que na aplicação de predição de picos de energia, os melhores resultados foram obtidos quando se considera a predição como uma tarefa de classificação.

1.3 Organização

O restante deste documento está organizado da seguinte maneira:

- O Capítulo 2 descreve conceitos relacionados com séries temporais, regressão extrínseca e regressão com múltiplas variáveis de saída;
- No Capítulo 4, é discutido o conceito de Redes Elétricas Inteligentes, relacionando com a tarefa de predição de picos explorada neste projeto e apresentação dos conjuntos de dados que serão utilizados nas etapas de avaliação das soluções investigadas;
- No Capítulo 3 são apresentados trabalhos relacionados à tarefa de predição de picos, discutindo as técnicas utilizadas e desempenho obtido para tal tarefa;
- No Capítulo 5 é detalhada a proposta deste trabalho, juntamente com a metodologia de desenvolvimento empregada;
- No Capítulo 6, são apresentados os resultados obtidos na avaliação experimental considerando as diferentes abordagens investigadas neste trabalho;
- Por último, no Capítulo 7, são discutidas as conclusões obtidas a partir da investigação realizada neste projeto e avaliação experimental conduzida, limitações e direções para trabalhos futuros.

Capítulo 2

Fundamentação Teórica

2.1 Considerações iniciais

Neste capítulo serão apresentados os principais fundamentos e conceitos relacionados aos temas de estudo deste projeto, como redes elétricas inteligentes, séries temporais, predição de picos, regressão extrínseca e regressão *multi-target*. No contexto deste projeto de pesquisa, os dados de consumo são estruturados como séries temporais, que posteriormente são submetidas a tarefas de mineração de dados, para construir modelos de regressão capazes de prever a magnitude do pico e sua posição. Neste projeto explora-se o conceito de regressão extrínseca como uma maneira de generalizar estes modelos de regressão, além de relacioná-los ao conceito de regressão *multi-target* como meio de aumentar o desempenho dos modelos, a partir da relação entre as variáveis alvo (magnitude do pico e sua posição). A seguir são apresentados detalhadamente cada conceito, juntamente com técnicas utilizadas para construir modelos de regressão *multi-target*.

2.2 Redes elétricas inteligentes

De acordo com Xue-Song et al. (2010), redes de distribuição de energia elétrica do tipo *smart grids* possuem um elevado nível de automação, sendo capazes de fornecer dados em tempo real sobre o fluxo de distribuição de energia no sentido das unidades consumidoras, assim como no sentido da central de distribuição. Por toda a extensão da rede são empregados inúmeros dispositivos de monitoramento que fornecem dados em tempo real sobre a tensão, corrente elétrica, potência, temperatura, vibração, entre outras informações úteis, não somente para monitorar a rede, como para realizar ações preventivas ou corretivas, ou até mesmo proteger o sistema de distribuição. Os mesmos autores também definem algumas características essenciais das redes *smart grid*, que estão relacionadas abaixo:

1. Auto-reparável: fornece mecanismos de gerenciamento capazes de prever, detectar e se recuperar de falhas.
2. Interativa: possuem sistemas sofisticados de integração entre as unidades consumidoras e as distribuidoras, de maneira que seja sempre possível manter o fluxo de energia em todas as unidades consumidoras da rede.
3. Forte: robusta o suficiente para resistir a desastres naturais ou provocados pelo homem e a ataques cibernéticos.
4. Otimização: fornecer mecanismos de monitoramento dos ativos da rede que torne possível reduzir custos de manutenção, custos de investimento e desperdício de energia elétrica.
5. Compatibilidade: é capaz de ser utilizada em diferentes meios de geração, distribuição e armazenamento de energia.
6. Integração: incluir otimização de processos, integração de informações e padronizações para refinamento do gerenciamento da rede

Com essas características, é possível perceber grandes diferenças no processo de distribuição de energia a partir de uma rede convencional e uma *smart grid*, principalmente em relação ao monitoramento em tempo real das unidades consumidoras. Este monitoramento possibilita ajustes finos e rápidos no sistema de distribuição, para otimizar a distribuição de energia em todos os pontos da rede, reduzindo custos de distribuição, mas também evitando quedas ou falhas na distribuição. A robustez que uma *smart grid* possui é obtida a partir de uma vasta intercomunicação entre os nós da rede, constantemente monitorados e ajustados, para conquistar a eficiência e robustez que se espera deste sistema.

A capacidade de se adaptar a situações adversas é o que mais diferencia uma *smart grid* de um sistema convencional de distribuição. Porém, a implantação de uma rede como essa também é mais complexa e possui um custo de construção mais elevado. A grande maioria dos países desenvolvidos têm projetos de implantação de *smart grids* em execução por toda a extensão de seus territórios, como Estados Unidos, países da Europa, China e Japão. Mas, devido ao elevado custo e também a falta de um padrão maduro para os componentes necessários, que facilitaria a construção destes componentes por mais empresas pelo mundo, ainda não é comum encontrar projetos de redes *smart grid* em execução nos países em desenvolvimento como o Brasil.

Em virtude da diversidade de informações que são coletadas em uma rede *smart grid*, abre-se espaço para analisar, ou até mesmo prever, comportamentos do sistema de distribuição de energia, aumentando ainda mais a eficiência do sistema. Dentre as tarefas possíveis de serem realizadas com os dados coletados de uma *smart grid*, pode-se citar a predição de falhas na distribuição devido a problemas nos componentes da *smart grid*. Seria possível prever se um transformador requer substituição

baseado no seu perfil de trabalho, ou ainda prever eventos de manutenção em componentes da *smart grid*. Estas duas tarefas são muito importantes para as companhias de distribuição, pois ambas têm impacto direto na continuidade da distribuição de energia. Em muitos países, incluindo o Brasil, quedas na distribuição de energia podem acarretar em elevadas multas para as companhias de distribuição, o que motiva a busca por mecanismos cada vez mais eficientes que previnem eventos de ruptura na distribuição de energia, algo possível de ser feito em uma rede *smart grid*.

Além dessas tarefas, pode-se citar também a predição de picos de consumo de energia elétrica de cada unidade consumidora, o principal tema de estudo deste trabalho. Nessa tarefa, modelos de aprendizagem de máquina são construídos utilizando os dados de consumo coletados pelos *smart meters* de uma *smart grid*, com o objetivo de prever a magnitude e o dia da semana que o pico de consumo ocorrerá na próxima semana. Maiores detalhes sobre essa tarefa são tratados na Seção 2.4.

2.3 Séries temporais

Séries temporais podem ser definidas como um conjunto de observações quantitativas, ordenadas cronologicamente no tempo. Formalmente, uma série temporal T , de comprimento n , é uma sequência ordenada de valores reais t_i , medidos no tempo i e em intervalos igualmente espaçados, podendo ser representada por $T = (t_1, t_2, \dots, t_n)$. Um exemplo de série temporal é a observação diária do preço de fechamento de uma determinada ação ao longo de 30 dias, conforme ilustrado na Figura 2.1.

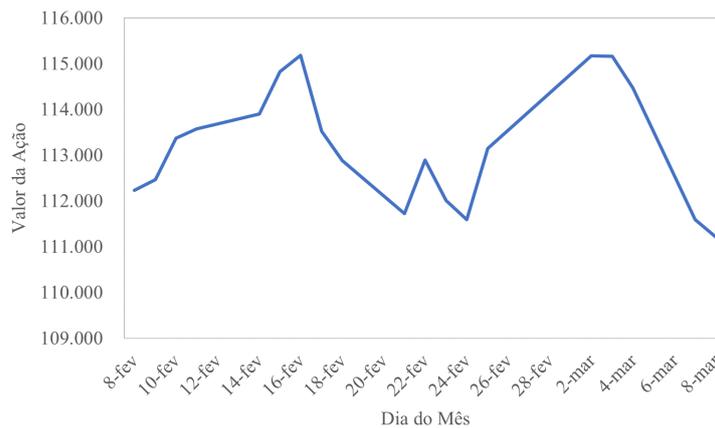


Figura 2.1: Exemplo de série temporal que representa o comportamento de um papel na bolsa de valores.

Dentre as diferentes tarefas de mineração de dados possíveis com séries temporais, destacam-se as seguintes (Ratanamahatana et al., 2005):

- **Indexação:** Também conhecida como consulta por similaridade (*query by content*), a tarefa tem como objetivo buscar em um banco de dados que armazena um conjunto de séries temporais, qual a série mais similar (ou as mais similares) à uma série de consulta Q . Para isso, utiliza-se uma medida de similaridade ou medida de distância $dist$ (por exemplo, distância euclidiana), para medir a similaridade entre a série de consulta Q com as séries P , armazenadas no banco de dados a partir de comparações $dist(Q, P)$. Esta busca pode considerar que Q e P possuem o mesmo comprimento ou que P é uma subsequência de Q .
- **Classificação:** determinar que tipo de objeto ou ações uma série temporal desconhecida pode representar, dado um conjunto pré-definido de classes composto de valores categóricos. Por exemplo, qual animal o som representado por uma série temporal pertence, ou que tipo de doença cardíaca um exame de eletrocardiograma (ECG) exhibe.
- **Regressão:** similar à tarefa de classificação, mas ao invés de considerar um conjunto de classes composto por valores categóricos, considera valores numéricos. Por exemplo, dada uma série temporal que representa a quantidade de acessos ao *website* de um determinado produto ao longo do dia, pode-se prever quantas vendas serão realizadas.
- **Agrupamento:** a partir de um conjunto de várias séries temporais, encontrar grupos naturais entre as séries a partir de suas similaridades. O resultado obtido pelo agrupamento fornece um número incerto de grupos, na grande maioria das vezes, assim como ocorre com o agrupamento de outros tipos de dados.
- **Forecasting:** a partir da análise do comportamento dos valores passados de uma série temporal, efetuar a previsão dos valores futuros da mesma série. Por exemplo, com base no comportamento de um papel de bolsa de valores, determinar os valores futuros deste mesmo papel.
- **Sumarização:** para séries temporais muito grandes, a tarefa consiste na criação de uma nova série temporal a partir de um subconjunto destes valores, possibilitando a sua representação gráfica de maneira mais intuitiva.
- **Segmentação:** consiste na divisão de uma série temporal em uma sequência de segmentos. Cada segmento é composto por uma subsequência de observações da série que possui características ou propriedades em comum. Uma aplicação da tarefa de segmentação é na identificação de atividades físicas em dados de sensores de movimento, sendo possível identificar segmentos que representam atividades como caminhar, correr, repousar, pedalar, entre outras.
- **Detecção de anomalias:** por meio da análise do comportamento da série temporal, identificar seções que apresentam valores inesperados ou anormais. Por

exemplo, dado uma série temporal que represente a temperatura de uma caldeira, identificar pontos da série com valores acima do esperado ou valores negativos, que possivelmente representam problemas no sensor de temperatura.

A tarefa abordada neste projeto é a predição de picos em séries temporais (regressão) com aplicação na predição de picos de consumo de energia elétrica. Devido às características do problema, esta tarefa será explorada como um problema de regressão extrínseca *multi-target*. Desse modo, os temas de regressão extrínseca e regressão *multi-target* são discutidas nas Seções 2.5 e 2.6, respectivamente.

2.4 Predição de picos de consumo

A adoção de *smart grids* permite a troca de informações em tempo real entre os consumidores e as estações de distribuição. Além disto, por meio de dispositivos como *smart meters*, é possível obter dados de consumo de cada cliente com um alto nível de granularidade, com medições a cada 15 minutos. Tais análises servem para construir sistemas de gerenciamento de demanda, responsáveis por ajustar os parâmetros da distribuição de energia de uma *smart grid*, aumentando a eficiência total do sistema, reduzindo falhas e custos operacionais.

A análise do padrão de consumo também é extremamente útil para construção de modelos capazes de prever os momentos de picos de consumo. Um pico de consumo em uma série temporal Q pode ser definido como um valor $x \in Q$, onde $x = \max(Q)$. Na tarefa de predição de picos também é relevante determinar sua posição p na série Q , ou seja:

$$p = \operatorname{argmax}_{0 \leq x \leq m} (Q) \quad (2.1)$$

Destaca-se que a predição é realizada antes do recebimento das observações da série temporal Q , tendo como base os dados do passado de Q . Por conveniência, podemos atribuir estes dados do passado a uma série temporal P . Além disso, considera-se que Q possui um comprimento m , referente ao tempo de antecedência que a predição é realizada.

A Figura 2.2 ilustra a maneira como a tarefa de predição de picos é abordada neste projeto. Nessa imagem, cada observação da série temporal representa a soma do consumo de energia elétrica ao longo de um dia. Desse modo, a série temporal P representa os dados de consumo previamente ao momento da predição por um período de X dias. A série temporal Q , representa os dados de consumo que ainda serão gerados no intervalo de sete dias ($m = 7$) e que o modelo de predição busca prever a magnitude e a posição do maior valor.

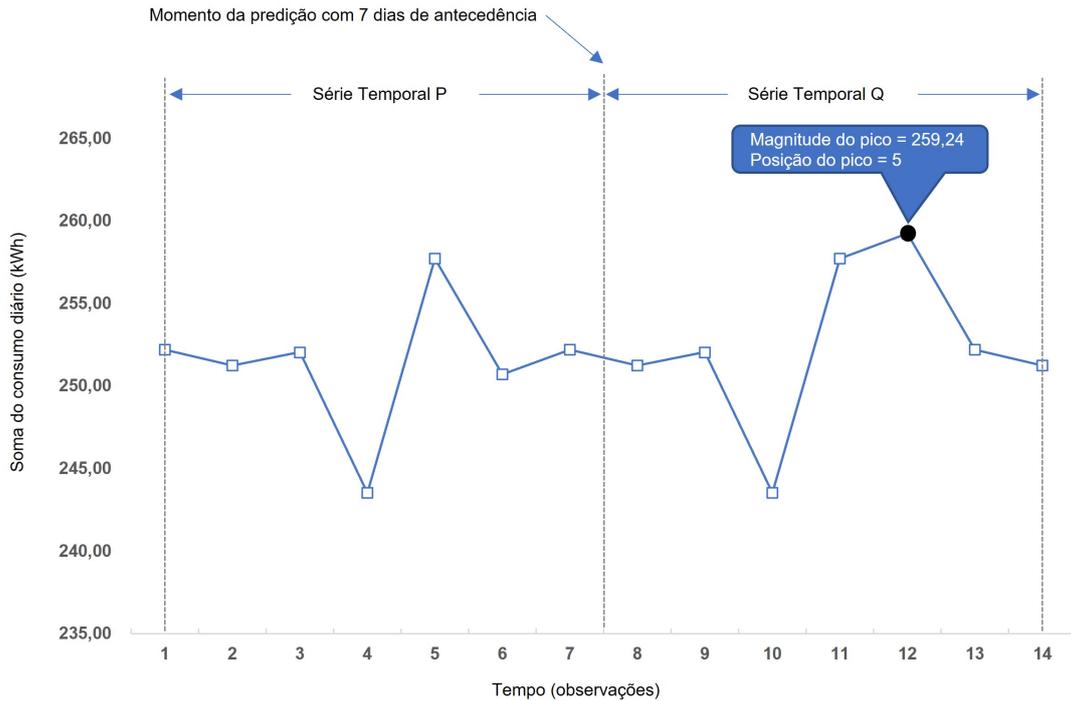


Figura 2.2: Predição da magnitude e posição do pico de consumo de energia com 7 dias de antecedência.

Para uma distribuidora de energia, prever o pico de consumo é uma tarefa importante, pois permite traçar com antecedência ações para manter o sistema funcionando, utilizando o mínimo possível de fontes não renováveis de energia, com o mínimo possível de sobrecarga em diferentes dispositivos da rede. Saber com antecedência um momento de pico de consumo é útil para determinar o momento que o sistema deverá ser abastecido por baterias, ao invés das usinas elétricas de pico.

Cabe ressaltar que a ativação destas usinas elétricas de pico para o fornecimento de energia não é um método econômico ou ambientalmente amigável (Dai et al., 2021), pois geralmente o custo para produzir energia por essas usinas é maior por utilizarem combustíveis fósseis, que demandam vários processos para transformação do combustível em energia elétrica. Por esta razão, é comum a aplicação de sobretaxas nos momentos de pico ou emissão de alertas aos clientes para a economia nestes períodos.

A Figura 2.3, adaptada de Dai et al. (2021), ilustra o processo realizado para predição de picos de consumo de energia, utilizando variáveis relevantes do domínio da aplicação. As variáveis coletadas pelo próprio sistema da *smart grid* são chamadas de endógenas. Dentre elas, pode-se citar as séries temporais que representam o consumo no intervalo de medição desejado para a construção do modelo. Geralmente escolhe-se o intervalo de uma hora entre as medições para a construção destas séries temporais. Já as variáveis coletadas por sistemas externos à *smart grid* são

chamadas de exógenas. Dentre elas pode-se citar temperatura do ar, umidade, pressão barométrica, estação do ano, dia da semana, feriado, entre outras. Na entrada do sistema são escolhidas as variáveis relevantes ao problema, bem como o período adequado para construção do modelo. A saída do modelo de predição consiste nos valores de pico de consumo e posição, sendo que cada valor pode ser determinado por modelos individuais ou por modelos *multi-target*.

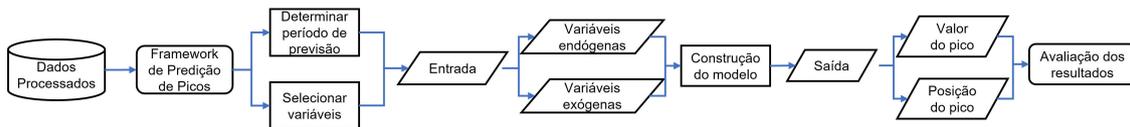


Figura 2.3: Exemplo de framework para predição de picos de energia.

Os modelos de predição de pico de consumo podem ser categorizado pelo horizonte desejado de predição. Os modelos de curto prazo são construídos para preverem a magnitude do pico no horizonte de algumas horas à até uma semana (Dai et al., 2021). Modelos de médio prazo possuem um horizonte de uma semana à até doze meses (Dai et al., 2021). Já modelos de longo prazo são construídos para preverem as magnitudes em horizontes acima de um ano (Dai et al., 2021). Cabe ressaltar que este trabalho de pesquisa está focado nos modelos de curto prazo, para atender a uma predição no horizonte de 7 dias.

Técnicas de pré-processamento podem ser aplicadas aos dados para limpeza, preenchimento de valores ausentes, remoção de *outliers* ou padronização dos dados dentro de uma mesma escala de valores. Este tratamento de dados é essencial pois estes problemas interferem diretamente no desempenho da predição. Uma vez que os dados estejam selecionados e pré-processados, na granularidade esperada, o restante da tarefa é realizado por um método de predição.

Com relação aos métodos de predição, a literatura possui variados métodos aplicáveis à tarefa de predição de picos (Dai et al., 2021), ilustrados pela Figura 2.4. Os primeiros métodos surgiram por volta de 1957, utilizavam a opinião de especialistas e experimentos empíricos para realização das predições. Na década de 80 surgiram os modelos estatísticos, baseados em regressão e autorregressão. Com a chegada das ANN - *Artificial Neural Network* (Yegnanarayana, 2009) por volta do início da década de 90, modelos avançados de predição baseados em aprendizagem de máquina e ANN começaram a surgir. Tais modelos constantemente superam os modelos estatísticos, por serem mais adaptáveis a eventos aleatórios das séries temporais e lidarem melhor com incertezas (Dai et al., 2021). Nos modelos de predição de curto prazo, o uso de aprendizagem de máquina é bastante recomendado, pois neste contexto o padrão de consumo é bastante variável e incerto (Dai et al., 2021). Além disto, este padrão tende a mudar drasticamente quando se pretende estudar padrões dentre diversos consumidores distintos, que é o caso desta pesquisa.

Com a evolução dos modelos de aprendizagem de máquina, outros métodos de predição de consumo foram surgindo. O SVM - *Support Vector Machine* (Drucker

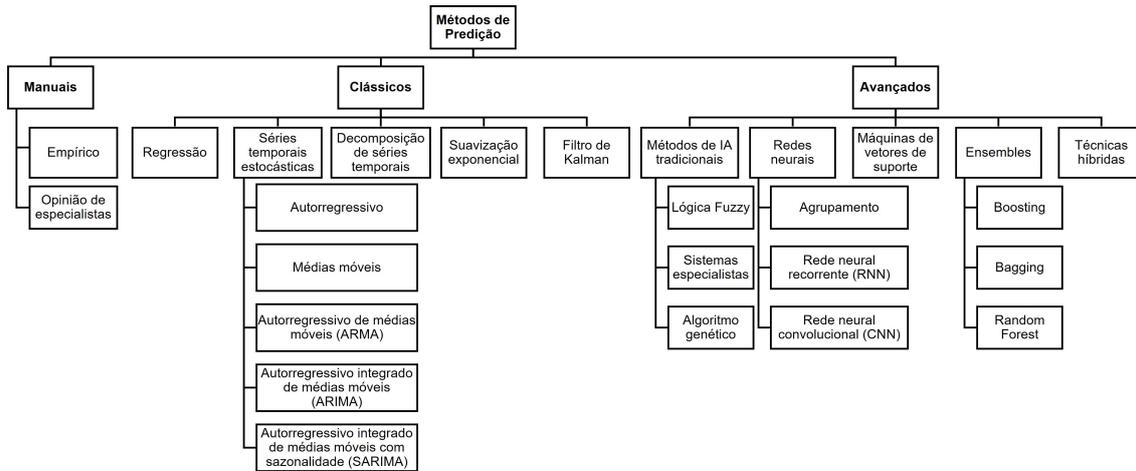


Figura 2.4: Métodos de predição de picos de consumo.

et al., 1996) é outra proposta de método de predição, por meio de sua variação SVR, que aplica o conceito do algoritmo para regressão. Métodos baseados em ensembles também foram explorados, com experimentos realizados pelos métodos baseados de *bagging* ou *boosting*. Além desses métodos, soluções baseadas em Deep Learning, por meio de seus algoritmos RNN - *Recurrent neural network* (Medsker e Jain, 2001), CNN - *Convolutional Neural Network* (O’Shea e Nash, 2015) ou LSTM - *Long short-term memory* (Gers et al., 2002), também possuem estudos relacionados (Dai et al., 2021), principalmente nos cenários onde é necessária uma análise de altos volumes de dados.

O trabalho de Dai et al. (2021) também cita alguns estudos que comparam a aplicação de métodos híbridos com os demais já citados, combinando a aplicação de modelos estatísticos com modelos baseados em aprendizagem de máquina. Tais modelos surgiram em virtude de diferentes problemas que existem no mundo real relacionados à predição de pico, onde etapas de pré-processamento ou etapas intermediárias podem ser necessárias antes da predição de fato. Além disso, os métodos híbridos foram mais eficientes em cenários onde há flutuação elevada das magnitudes de pico. Alguns modelos construídos para determinação do pico e posição de consumo por meio de classificação aplicaram lógica *fuzzy* para determinação do pico com maior acurácia.

É importante ressaltar que a escolha do método de predição tem forte relação com o contexto do problema, no que tange a disponibilidade dos dados na granularidade necessária, além do horizonte desejado para predição. Isso pode causar uma alta incerteza sobre o comportamento da série temporal, que reflete diretamente no desempenho do modelo preditivo em construção. Por exemplo, um dado consumidor pode ter uma baixa variabilidade dos dados em um certo dia, por ser um dia de feriado. Mas pode ter uma alta variabilidade em uma data comemorativa, não relacionada a um evento amplamente conhecido, como o dia do aniversário do con-

sumidor. A escolha do método de predição de pico necessita levar em consideração estes fatores, pois impactarão diretamente no seu desempenho.

2.5 Regressão extrínseca

Em aprendizagem de máquina, regressão consiste em prever um valor numérico contínuo a partir de um conjunto de atributos descritivos de um problema (Russell e Norvig, 2002). Em problemas envolvendo séries temporais, os atributos descritivos são compostos pelas observações do passado da série temporal. Quando o valor a ser predito pelo regressor representa uma continuação da série de entrada, ou seja, o seu valor futuro, tem-se a tarefa de *forecasting*, conforme discutido anteriormente. Entretanto, em muitas ocasiões se tem o interesse em prever um valor que não representa necessariamente o valor futuro da série, podendo ser até mesmo um valor externo à série temporal de entrada. Um exemplo concreto é a predição da frequência cardíaca de uma pessoa com base em dados de acelerômetro (Phan et al., 2008). Note que, neste exemplo, a predição (frequência cardíaca) possui um significado diferente dos dados de entrada (dados de acelerômetro), embora possam ser encontradas correlações. Assim, o valor a ser predito não representa uma continuação da série temporal de entrada. Para este cenário, se dá o nome de *regressão extrínseca* ou TSER - *Time Series Extrinsic Regression* (Tan et al., 2021).

De acordo com Tan et al. (2021), a regressão extrínseca se enquadra entre as tarefas de classificação e *forecasting*. Se o valor a ser predito é um valor futuro da série, representa um problema de *forecasting*. Por outro lado, se o valor a ser predito não tem relação com os valores futuros, mas é um valor discreto dentro de um conjunto finito de possibilidades, representa um problema de classificação de séries temporais. Enquanto na regressão extrínseca, o valor a ser predito é um valor contínuo que não se enquadra em um conjunto finito de possibilidades (classe) e também não representa um valor futuro da série.

Dado um conjunto de treinamento $\mathcal{D} = \{(T_1, y_1), (T_2, y_2), \dots, (T_m, y_m)\}$ composto por m pares de séries temporais rotuladas (T_i, y_i) , em que T_i é uma série temporal de comprimento n e $y_i \in \mathbb{R}$ é o atributo-alvo da série T_i , um modelo de regressão extrínseca para séries temporais é uma função capaz de prever um valor contínuo \hat{y} para uma série desconhecida Q com base nos dados observados em \mathcal{D} .

Algoritmos de aprendizagem de máquina tradicionais para a tarefa de regressão, como Floresta Aleatória (Breiman, 2001) e Árvore de Regressão (Loh, 2011), podem ser diretamente aplicados na tarefa de regressão extrínseca com dados de séries temporais. Entretanto, estes algoritmos possuem a suposição que os atributos são independentes entre si e não possuem relação com o tempo, o que não é verdade quando os dados de entrada são séries temporais. Assim, o desempenho preditivo destas abordagens pode ser prejudicado neste cenário. Além disso, estes algoritmos

podem ter dificuldade em lidar com a alta dimensionalidade dos dados, uma característica recorrente em séries temporais. Uma alternativa para estas questões é realizar a extração de novos atributos que descrevem as características dos dados de séries temporais. Neste caso, uma limitação desta abordagem é que a solução muitas vezes se torna dependente da aplicação, já que cada problema possui um conjunto específico e diferente de atributos.

Também encontram-se algoritmos baseados em Deep Learning (Dechter, 1986), como o LSTM, que consideram uma relação com o tempo para os atributos no processo de predição. Entretanto, estes algoritmos podem apresentar desempenho insatisfatório quando é necessário treinar modelos usando séries temporais muito extensas, além de apresentar alto tempo para treinamento. Já em um modelo construído a partir da TSER, os valores preditos podem ser determinados com base na análise de toda a série temporal e não com base apenas em uma janela da série, tampouco em um intervalo curto de tempo ao final da série.

2.5.1 Aplicações

Para suportar a pesquisa da TSER, Tan et al. (2021) criaram alguns conjuntos de dados de *benchmark*¹, úteis para analisar o desempenho de diferentes modelos de aprendizagem de máquina nessa tarefa. As séries temporais disponíveis nestes datasets são de diversas áreas do conhecimento, como monitoramento de saúde, monitoramento de energia, monitoramento de ambiente, análise de sentimento e previsões. A seguir é apresentada uma visão geral sobre algumas destas aplicações.

Em monitoramento do ambiente, pode-se monitorar qualquer evento relacionado ao meio ambiente, como nível de poluição, pluviosidade, rendimento de colheitas e níveis de inundações. Cabe ressaltar que fenômenos como estes podem exercer influência entre si. Por exemplo, a pluviosidade pode afetar diretamente o nível de inundação e um mecanismo de predição que utilize especificamente os dados da série temporal se torna muito relevante. Por meio da TSER, é possível criar um modelo de regressão capaz de determinar o nível de inundação com base na análise das séries temporais que representam a pluviosidade de uma região, sem entrar no mérito de quais características influenciam um fenômeno ou outro. Outra aplicação interessante é baratear a predição de um fenômeno a partir do uso de dados que sejam mais fáceis ou baratos de serem coletados. Por exemplo, pode-se prever o nível de chuva de uma região a partir de dados de temperatura do ar. Neste caso, sensores que capturam a temperatura do ar são mais baratos e mais fáceis de instalar quando comparados a sensores de nível de chuva.

Em monitoramento de saúde, a tarefa consiste em monitorar sinais vitais de um certo indivíduo, como pressão sanguínea, batimentos cardíacos ou nível de saturação de oxigênio. Geralmente os dados necessários para estas análises são oriundos

¹Disponível em <http://tseregression.org>

de dispositivos vestidos pelo indivíduo, como o esfigmomanômetro para leitura da pressão sanguínea, o eletrocardiógrafo para leitura dos batimentos cardíacos e o oxímetro para leitura de níveis de saturação de oxigênio. A partir destes dispositivos, se obtém as séries temporais correspondentes, de onde pode-se aplicar a TSER para determinar eventos futuros na saúde do indivíduo, como estimar a taxa de batimento cardíaco, a taxa de respiração e o nível de saturação de oxigênio no sangue de um paciente.

Na análise de sentimentos, temos as interpretações e classificações expressas em algum texto. O sentimento sobre um texto é definido como um *score*, onde intervalos deste valor classificam um texto como positivo, neutro ou negativo, sendo útil para determinar se o texto expressa algo bom ou ruim sobre um certo assunto. A análise de sentimentos em redes sociais é bastante utilizada por empresas que desejam detectar, de maneira eficiente, qual o sentimento expresso pela sua marca nas redes sociais, ou ainda, encontrar consumidores que estejam descontentes com o produto ou serviço prestado. A TSER torna-se útil neste cenário por permitir uma análise completa de toda a série temporal, podendo fornecer *scores* por toda a extensão da série temporal e não somente em passados recentes.

2.5.2 Algoritmos

Como mencionado anteriormente, a TSER tem como premissa que os valores da série temporal estejam mapeados como características, permitindo o uso de algoritmos tradicionais de aprendizagem de máquina e não somente algoritmos que levam o tempo em consideração no aprendizado, como é o caso do RNN e do LSTM.

Dependendo do mecanismo de leitura dos dados na origem, a partir de sensores ou *streams* de dados, uma etapa de pré-processamento é necessária para criar um vetor de características de tamanho $D \times L$, onde D representa o número de dimensões e L o comprimento da série temporal. Por exemplo, uma série temporal de 3 dimensões e 100 valores deverá ser representada por um vetor de 300 características, como mencionado por Tan et al. (2021).

É importante salientar que é uma prática comum considerar o uso de toda a série temporal na construção do vetor de características. A extração de características pode ser feita por meio do cálculo manual dos atributos ou a partir do uso de ferramentas capazes de extrair informações mais genéricas. Na literatura encontramos alguns algoritmos que efetuam extração de características em séries temporais. O algoritmo FFORMA - *Feature-based Forecast Model Averaging* (Montero-Manso et al., 2020) cria um metamodelo de predição utilizando características extraídas automaticamente das séries temporais. Dentre as características exploradas por este modelo temos a entropia espectral, tendência, sazonalidade, linearidade e autocorrelação. Além do FFORMA, temos o pacote *tsfeatures*², o HCTSA - *Highly*

²Disponível em <https://github.com/robjhyndman/tsfeatures>

comparative time-series analysis (Fulcher e Jones, 2017) e o Catch22 (Lubba et al., 2019).

Após a composição do vetor de características, seja com os dados de séries temporais ou com características extraídas da série, pode-se aplicar algoritmos de aprendizagem de máquina tradicionais como o SVM, por meio de sua variação SVR - *Support Vector Regression* que aplica os conceitos do SVM para fins de regressão. Da mesma maneira, outros algoritmos clássicos como o ANN, Floresta Aleatória (Breiman, 2001), KNN - *k nearest neighbour* (Sammut e Webb, 2011) e o XGBoost - *Extreme Gradient Boosting* (Chen et al., 2015), em suas variações construídas para regressão, são algoritmos possíveis de serem utilizados para a TSER.

Alguns estudos propõem o uso de subsequências representativas da série original (chamadas de *shapelets*) para a construção de um modelo preditivo mais eficiente (Ye e Keogh, 2009). Embora a eficiência deste modelo seja relevante, o custo para a extração e avaliação destes padrões pode ser proibitivo em séries de alta dimensionalidade. Por esta razão, algumas variações foram propostas para reduzir a complexidade destes algoritmos, como o *Shapelet Transform* (Lines et al., 2012) e o *Shapelet Ensemble* (Bagnall et al., 2015).

O trabalho de Dempster et al. (2020) propôs um algoritmo de classificação de séries temporais, capaz de trabalhar com uma fração de tempo se comparado aos demais apresentados anteriormente. O Rocket - *Random convolutional kernel transform* transforma uma série temporal, utilizando um alto número de kernels convolucionais aleatórios, treinando um classificador baseado em *ridge regression*, responsável pela tarefa de predição. Cada kernel possui características próprias como pesos, viés e *padding*. Quando aplicado em uma série temporal, gera um vetor de características resultante da tarefa de convolução. Em cada um destes vetores é calculada uma proporção de números positivos, produzindo um segundo vetor de 2 números que significa a proporção de características presentes no kernel. Em uma série temporal qualquer, se aplicado o Rocket com 10.000 kernels, será gerado um vetor de características resultante de 20.000 entradas, que servirá de base para o treinamento e validação do modelo baseado em *ridge regression*. Comparado a outros modelos de classificação de séries temporais como o HIVE-COTE (Lines et al., 2016) e o InceptionTime (Ismail Fawaz et al., 2020), em uma avaliação experimental com 85 conjuntos de dados de *benchmark*, o Rocket obteve o melhor desempenho nas comparações. Inicialmente o Rocket foi desenvolvido para tarefas de classificação, mas o trabalho de Dempster et al. (2020) adaptou o algoritmo para que trabalhe também com tarefas de regressão.

Considerando abordagens de Deep Learning, o trabalho de Tan et al. (2021) atestou a eficiência do ResNet (*Residual Networks*), FCN (*Fully Convolutional Neural Networks*) e do Inception nas tarefas de TSER, realizando benchmarks em 19 conjuntos de dados. O experimento foi realizado comparando estes algoritmos com algoritmos clássicos como o Floresta Aleatória, o XGBoost e o SVM. Os métodos de

Deep Learning superaram os métodos clássicos, embora não tenha sido encontrada diferença estatisticamente significativa para o Rocket.

2.6 Regressão *multi-target*

Na área de aprendizagem de máquina, em específico os problemas de aprendizagem supervisionada, os modelos de predição podem fornecer valores categóricos ou numéricos. No primeiro caso o problema é denominado classificação, enquanto no segundo caso, regressão. Os modelos preditivos construídos para prever apenas um único valor numérico, discreto ou contínuo, são chamados também de modelos *single-target* (ST), ou de único alvo.

Entretanto, é comum encontrar aplicações que envolvam múltiplos alvos para um certo conjunto de características. Por exemplo, a predição de peso e altura de uma pessoa, baseado em atributos descritivos de sua imagem. Ou ainda, dado uma série temporal que represente o consumo de energia elétrica de uma unidade consumidora, utilizada como fonte de dados para construção de um modelo preditivo através da TSER, tendo tal modelo a capacidade de prever o valor de pico e sua posição. Este último problema tem relação com o problema tratado por este projeto de pesquisa e será relatado em capítulos posteriores. Por possuírem a capacidade de definir múltiplos valores de saída, contínuos ou discretos, estes modelos são chamados de modelos *multi-target* (MT) (Aho et al., 2012), ou de múltiplos alvos.

Formalmente, um problema de predição de múltiplas variáveis é definido como uma função h , que relaciona um conjunto X de m características a um conjunto Y de d variáveis

$$h : X_{1..m} \rightarrow Y_{1..d}$$

As variáveis de saída dos modelos MT podem ser induzidas a partir de diversos modelos ST que operam de maneira independente. Porém, o estado da arte dos algoritmos de MT exploram relações que tais variáveis possuem, utilizando estas relações para otimizar o resultado final dos modelos de predição. A correlação entre as variáveis de saída pode ser determinada através de algum modelo estatístico ou preditivo. Na literatura existem modelos que utilizam simplesmente correlação linear (coeficiente *Pearson*), mas existem modelos preditivos sofisticados que utilizam a variação do erro médio quadrático (MSE) para atestar que alguma variável possui forte relação com outras variáveis. Frequentemente se observa que os modelos MT que exploram estas relações superam os modelos MT que trabalham com regressores ST independentes.

Na próxima seção serão discutidos alguns algoritmos de predição *multi-target*, focando em problemas de regressão, por estar relacionado ao contexto deste trabalho.

A literatura fornece diversos algoritmos relacionados a regressão *multi-target*. Entretanto, serão apresentados algoritmos que representam o estado da arte do tema, ocultando possíveis variações de tais algoritmos, por não terem tanta relevância com o tema deste trabalho.

2.6.1 Métodos globais e locais

Os algoritmos de regressão MT se diferenciam pelo modo que os problemas são tratados. Dentre as categorizações possíveis, existem os métodos globais e locais (Kocev e Ceci, 2015). Os métodos globais se baseiam na transformação de um algoritmo de aprendizagem de máquina conhecido, para que passe a lidar com múltiplas saídas e relações que eventualmente existam entre elas. Dentre as transformações possíveis temos a função de otimização do SVM, o critério de separação de nós do Árvores de Regressão, entre outros. Já os métodos locais buscam adaptar o problema ao invés de adaptar um algoritmo conhecido. Para isto, modificações nos dados de entrada são realizadas, para explorar relações existentes entre as variáveis.

Dentre os métodos globais, Xu et al. (2013) propuseram uma nova forma de calcular a função de perda do algoritmo SVM, para tratar problemas *multi-target* e possíveis relações que as variáveis de saída tenham. O trabalho de Kocev et al. (2007) estuda uma abordagem utilizando uma variação do algoritmo Floresta Aleatória denominada MORF (*Multi-objective Random Forest*), que fornece o tratamento de múltiplas variáveis de saída, aplicando uma estratégia de classificação hierárquica e agrupamento do conjunto de dados. A técnica MORF é considerada um dos melhores métodos de regressão MT. Porém, métodos locais mais recentes foram capazes de superar seu desempenho preditivo (Mastelini et al., 2020).

Ainda nos modelos globais, outros métodos baseados em ensembles também são relevantes, como o FIRE (Fitted Rules Ensemble) de Aho et al. (2012), que também utiliza o Floresta Aleatória e tem o diferencial de endereçar maneiras de fornecer um modelo que seja facilmente interpretável, para tratar problemas onde a explicação da predição é importante. Além destes, o trabalho de Hadavandi et al. (2015) propõe um modelo MT por meio de um ensemble de ANN baseado em Boosting, juntamente com a extração de características PCA.

Conforme dito anteriormente, métodos locais de regressão MT tratam o problema por meio de modelos ST conhecidos. Esta abordagem frequentemente supera os resultados obtidos pelos métodos globais. Dentre inúmeros modelos documentados na literatura, será discutido com maiores detalhes 3 métodos relevantes a esta pesquisa: o ERC (*Ensemble of Regression Chains*), o SST (*Stacked Single-Target*) e o DSTARS (*Deep Structure for Tracking Asynchronous Regressor Stacking*).

2.6.2 ERC

O método ERC (*Ensemble of regression chains*), proposto em Spyromitros-Xioufis et al. (2016), é um método MT local que funciona a partir do estabelecimento de uma sequência aleatória das variáveis de saída e suas características, por onde o modelo de predição utiliza os resultados induzidos da primeira sequência de características como instâncias de treinamento da próxima cadeia e assim sucessivamente para todas as sequências estabelecidas no problema. A primeira variável-alvo de saída é induzida a partir das características relacionadas a ela e esta servirá de característica para o segundo modelo criado para a segunda cadeia, que representa as instâncias da segunda variável-alvo e assim por diante até percorrer todas as cadeias estabelecidas para o conjunto de dados. Caso o número de alvos seja maior do que três, o algoritmo permite que no máximo 10 cadeias sejam criadas, como proposto por Spyromitros-Xioufis et al. (2016), a fim de otimizar o desempenho do algoritmo.

A Figura 2.5 ilustra como o processo de treinamento é tratado pelo algoritmo ERC. O conjunto de características é representado por X e as variáveis que correspondem a classe ou a uma grandeza contínua são representadas por y_1 até y_d . As variáveis de saída da predição do modelo estão representadas por y_i^p , onde $i \in \{1, 2, \dots, d\}$. É possível notar que cada variável y_i^p é utilizada como característica da cadeia seguinte e em cada cadeia é criado um modelo de regressão novo. O processo se repete até que todas as cadeias sejam calculadas, formando-se assim o conjunto final com todas variáveis de saída determinadas para o problema.

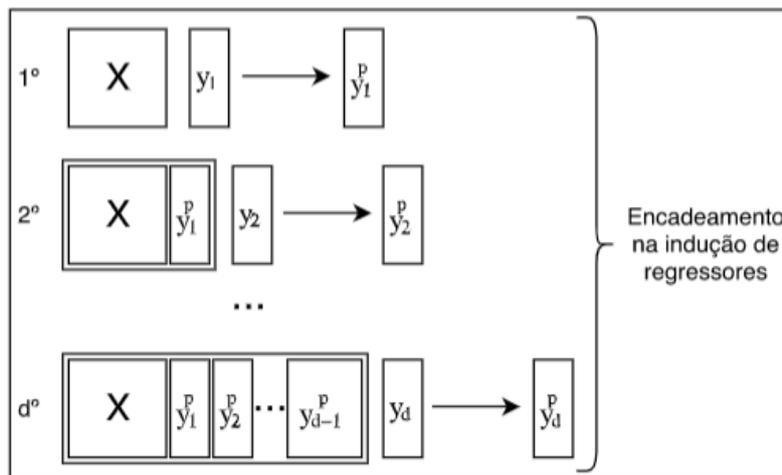


Figura 2.5: Representação gráfica do treinamento do ERC. Fonte: Mastelini et al. (2020).

A complexidade do algoritmo ERC é $O(10db)$, onde d é o número de variáveis-alvos, b a complexidade do regressor base utilizado, seguindo as recomendações de Spyromitros-Xioufis et al. (2016) de limitar em até 10 o número de cadeias, caso o problema tenha mais de 3 alvos. Para os problemas com até três alvos, a comple-

xidade do ERC é da ordem de $O(d^2(d-1)!b)$, considerando que para cada cadeia gerada, d regressores são construídos.

2.6.3 SST

Também proposto em Spyromitros-Xioufis et al. (2016), o método SST (*Stacked Single-Target*) utiliza duas camadas de regressores ST, onde na primeira camada é realizada uma indução das variáveis de saída individualmente e na segunda camada é feita uma indução adicional com um modelo preditivo para cada variável-alvo, utilizando todas as variáveis induzidas na primeira camada em conjunto com as características do conjunto de dados.

A Figura 2.6 representa como o processo de treinamento é realizado pelo algoritmo SST. Dado um conjunto de características X e um conjunto Y de d variáveis-alvos, modelos preditivos ST são induzidos para cada variável-alvo na primeira camada, gerando um segundo conjunto X^{est} por meio da concatenação do conjunto X com as variáveis-alvos y_i^p , induzidas na primeira camada do modelo preditivo. Em uma segunda camada são induzidos novos modelos preditivos ST para as variáveis-alvos do problema, utilizando o conjunto X^{est} gerado na primeira camada para induzir as variáveis-alvos y_i^F , que formam o conjunto final das variáveis de saída do problema, onde $i \in \{1, 2, \dots, d\}$.

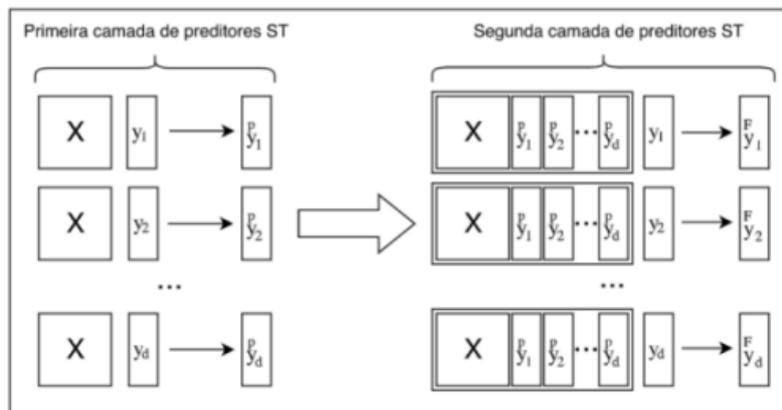


Figura 2.6: Representação gráfica do treinamento do SST. Fonte: Mastelini et al. (2020).

Sob a ótica de complexidade, o algoritmo SST é dependente da complexidade b dos modelos preditivos ST utilizados na indução e d variáveis-alvos relacionadas ao problema. Como a indução é realizada em duas camadas, é possível afirmar que a complexidade do SST é da ordem de $O(2db)$. É importante salientar que o tamanho do conjunto de dados tem influência no custo computacional do SST, mas este tamanho está abstraído na complexidade b do modelo preditivo ST utilizado.

2.6.4 DSTARS

Proposto por Mastelini et al. (2020), o método DSTARS é baseado na geração de camadas profundas de modelos de predição, em contraste com outros modelos de regressão MT que trabalham de forma rasa, como o ERC e o SST. Esta arquitetura do DSTARS é motivada pelo fato que modelos mais profundos de indução tendem a gerar resultados melhores e são menos suscetíveis a *overfitting* quando o conjunto de dados de treinamento se torna muito grande. Outra característica importante do DSTARS é o fato que o algoritmo leva em consideração que nem sempre as variáveis de saída possuem correlação entre si. Portanto, para os casos onde as variáveis de saída não possuem correlação entre si, o algoritmo trata o problema como uma regressão ST, simplesmente.

A Figura 2.7 demonstra o fluxo esquemático do algoritmo DSTARS. O conjunto de dados é inicialmente particionado em conjuntos de treinamento e validação, utilizando alguma estratégia de particionamento conhecida pela literatura, como por exemplo, a validação cruzada. Feito isso, uma primeira etapa de indução com modelos ST é feita para cada variável de saída. Na figura 2.7, o resultado desta indução está representado pela variável \hat{Y}_i^1 , onde $i \in \{1, 2, \dots, f\}$ e f representa o número de variáveis-alvos do problema relatado.

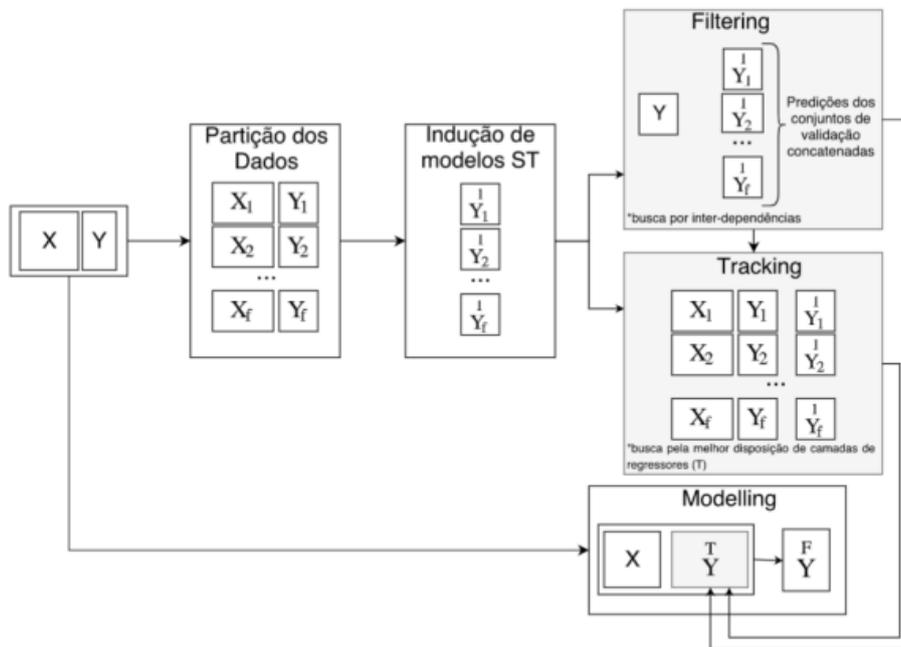


Figura 2.7: Representação gráfica do treinamento do DSTARS. Fonte: Mastelini et al. (2020).

Após a primeira indução ST sobre o conjunto de dados particionado, o algoritmo DSTARS efetua 3 procedimentos: *Filtering*, *Tracking* e *Modelling*. Na etapa *Filte-*

ring, o algoritmo procura pelas variáveis de saída que possuem correlação e somente elas estarão sujeitas a empilhamentos de modelos preditivos. Já as variáveis de saída não correlacionadas serão simplesmente tratadas por modelos preditivos ST rasos.

Para determinar a correlação, o algoritmo pode utilizar métodos de correlação linear simples, como o Pearson, ou utilizar métodos mais refinados. Um exemplo é o *Out-of-bag error* (OOBE) utilizado pelo Floresta Aleatória. O OOBE indica o erro calculado da predição de amostras fora dos *baggings* definidos pelo algoritmo. É esperada uma redução deste erro ao inserir uma variável-alvo que possua correlação com outras. Uma vantagem deste método é a possibilidade de operar com correlações não-lineares, diferentemente do método de Pearson, que opera exclusivamente com correlações lineares.

Na etapa *Tracking*, o algoritmo busca a melhor disposição de modelos de predição ST, realizando empilhamentos de modelos preditivos nas características, sempre que o método de validação determinar que tal empilhamento resultou numa redução do erro da indução.

A última etapa *Modelling* realiza a indução final das variáveis-alvos, utilizando a estrutura resultante da etapa *Tracking*. Tal estrutura utiliza um conjunto de características concatenado às variáveis induzidas que obtiveram no menor erro, representadas pela variável $\overset{T}{Y}$. A variável $\overset{F}{Y}$ representa a saída do modelo de indução, para cada variável-alvo do problema relatado.

2.7 Considerações finais

No campo de mineração de séries temporais, dentre diversas tarefas apresentadas, a predição de picos é uma tarefa que pode ser realizada utilizando métodos determinísticos, como o uso de heurísticas para determinar um valor futuro com base no passado. Utilizar métodos baseados em aprendizagem de máquina na predição de picos têm a vantagem de se adaptarem melhor a diferentes comportamentos das séries temporais, proporcionando melhor desempenho e escalabilidade em diferentes contextos. A regressão extrínseca é um método que facilita o desenvolvimento de modelos de aprendizagem de máquina, pelo fato de utilizar somente dados brutos das séries temporais, dispensando a coleta de dados externos às séries. Combinado com a regressão *multi-target*, configura um método robusto de predição de diversas variáveis de uma série temporal, como a magnitude a posição de um pico em uma série temporal.

Capítulo 3

Trabalhos Relacionados

3.1 Considerações iniciais

Neste capítulo são apresentados trabalhos da literatura relacionados à tarefa de predição de picos em séries temporais. Estes trabalhos utilizam métodos de aprendizagem de máquina ou abordagens estatísticas para prever a magnitude do pico e sua posição nas observações futuras. São discutidos trabalhos que realizam a tarefa de predição de picos em aplicações envolvendo o consumo de energia elétrica, como este projeto, além de outros problemas, relacionados à predição de enchentes ou secas, predição de fluxo de água ou predição de consumo de baterias. Em ambos os casos, em geral os trabalhos abordam o problema como uma tarefa de regressão ou classificação.

3.2 Predição de picos de consumo de energia

Em Mughees et al. (2021), o problema de predição de picos é tratado a partir da análise do padrão de comportamento de consumo de energia elétrica, considerando uma distinção do padrão de consumo entre dias úteis e feriados. De acordo com os autores, a diferença no comportamento do consumo dos clientes nestes dias dificulta o desenvolvimento de um modelo único de predição, pois além do comportamento padrão de cada consumidor, características regionais das unidades consumidoras podem alterar esses padrões. Os autores propõem um modelo de predição utilizando redes Bi-LSTM (*Bi-directional long short term memory*) (Melamud et al., 2016), escolhida em virtude das características de seu funcionamento, como a possibilidade do armazenamento de informações sobre o passado (camada backward, atrelada as características) e também sobre o futuro (camada forward, relacionada aos picos previstos). Na avaliação experimental, o modelo foi comparado com a rede LSTM (do tipo shallow e deep), ANN e SVM. Na etapa de treinamento dos modelos,

foram utilizados dados históricos de consumo obtidos por medidores mecânicos e fornecidos pela companhia de distribuição de energia elétrica FESCO (*Faisalabad Electric Supply Company*), entre maio de 2015 a setembro de 2017. Dessa base, foram extraídas 9 características para o treinamento e teste do modelo: semana do ano, dia do mês, dia da semana, dia especial (feriado ou não), temperatura máxima, ponto de orvalho máximo, precipitação acumulada, pico do dia anterior e pico da semana anterior. Foram consideradas 818 amostras entre o período de maio de 2015 a julho de 2017 para o treinamento e validação dos modelos, restando outras 6 semanas remanescentes entre agosto e setembro de 2017, utilizadas na etapa de teste. Para avaliação do desempenho dos regressores, foi utilizado o erro absoluto médio percentual (MAPE). Comparando o MAPE de todos os modelos, o modelo Bi-LSTM superou o SVM em 73%, o ANN em 68%, o shallow LSTM em 78% e o deep LSTM em 82%, considerando dias especiais (ou feriados). Já nos dias úteis, o modelo Bi-LSTM superou o ANN em 0,8%, o shallow LSTM em 40%, o deep LSTM em 6% e o SVM em 2%. Em termos absolutos, o modelo Bi-LSTM obteve um MAPE de 3,67% considerando dias úteis do mês de agosto de 2017, 5,92% para o mês de setembro e 4,93% para o mês de outubro. Já para os dias especiais, foi obtido um MAPE de 2,84% o mês de agosto, 4,90% para o mês de setembro e 2,73% para o mês de outubro.

Visando otimizar o custo de distribuição de energia, Santos et al. (2020) propõem a tarefa de predição da magnitude do pico para auxiliar na definição de preços com 24 horas de antecedência. Para determinar as magnitudes, foi proposto um regressor baseado em redes ANN, treinado a partir de dados históricos, gerados pelos sistemas de distribuição de energia das cidades de Lisboa, Porto e Évora. Foram extraídas 10 características baseadas em autocorrelação que representam o consumo de energia ocorrido na última semana a partir do ponto da predição. Juntamente com essas características, foi calculado um padrão de consumo que representa o momento da predição. O padrão foi calculado considerando a identificação do tipo do dia (dia útil, final de semana ou feriado) e uma média da carga de consumo para cada dia comparado ao seu correspondente dia no ano anterior. Além disso, foi acrescentado um ruído na entrada do regressor para adicionar variabilidade aos padrões de consumo, uma vez que estes dados variam muito pouco a cada intervalo de predição. Na predição com 24 horas de antecedência, os melhores resultados foram obtidos a partir do uso do vetor de características completo que inclui as características de autocorrelação, os padrões de consumo e o ruído. Neste caso, foi obtido um MAPE de 3,008% para a cidade de Lisboa, 2,986% para Porto e 4,221% para Évora.

Em Soman et al. (2020), é introduzido um modelo para sobretaxar o consumo de energia elétrica de clientes considerados de alta demanda, como centros comerciais, universidades, centros de convenções e fábricas. De acordo com os autores, tal modelo busca motivar os clientes a economizarem energia elétrica e evitarem situações de pico de consumo na rede de distribuição. Para reduzir o faturamento de energia elétrica com sobretaxa, são propostos sistemas de distribuição suporta-

dos por baterias. Estes sistemas fornecem energia elétrica à rede nos momentos de pico e são carregadas nos momentos de baixa demanda. Para determinar o exato momento para utilização das baterias, foram consideradas duas abordagens. A primeira abordagem determina o uso das baterias em momentos predeterminados. A segunda abordagem se baseia na predição da magnitude e posição dos picos por meio de sistemas inteligentes, que realizam a previsão destes valores com base no perfil de consumo de cada unidade consumidora. Para determinar o horário exato dos picos com uma antecedência de 24 horas, foi proposto um modelo baseado na LSTM. O modelo utiliza os dados de consumo dos últimos 2 dias, em conjunto com características como estação do ano, ocorrência de feriado, temperatura externa e umidade. O modelo foi treinado à partir de dados históricos coletados de uma micro-grid de 156 construções. Para fins de avaliação, foram testados dois modelos LSTM, um com 2 camadas de 100 e 80 neurônios e outro com 4 camadas de 100, 90, 80 e 70 neurônios. Também foram comparados modelos baseados em regressão linear e ANN. Os modelos LSTM obtiveram um MAPE de 4.1% e 3.7%, nas arquiteturas de 2 e 4 camadas respectivamente para a predição da magnitude dos picos. Já os modelos de regressão linear e ANN obtiveram um MAPE de 5.9% e 3.8%, concluindo que o LSTM de 4 camadas possui o melhor desempenho. A predição da posição dos picos foi abordada como um problema de classificação. Para a tarefa, os modelos LSTM obtiveram uma acurácia de 16% e 47%, novamente nas arquiteturas de 2 e 4 camadas, enquanto os modelos de regressão linear e ANN obtiveram uma acurácia de 42% e 26%. Foi observado que a acurácia aumenta significativamente se os modelos forem treinados para prever os k picos do dia, chegando a 95% de acurácia no LSTM de 4 camadas, para a predição dos 4 picos de um dia.

Em Dowling et al. (2018), é discutido o problema de predição de picos coincidentes a partir de dados da ERCOT (*Electric Reliability Council of Texas*). Picos coincidentes representam o momento em que todo o sistema de distribuição de energia está sob alta demanda devido ao pico de consumo de muitos consumidores ao mesmo tempo. Neste caso, aplica-se uma sobretaxa de até 100 vezes o valor da tarifa regular para estes consumidores durante o período. A necessidade de uma solução robusta para a tarefa se deve a limitações da solução até então utilizada pela ERCOT, baseada em um modelo de regressão quantílica para a emissão de alertas de pico de demanda. Em uma avaliação com dados de consumo de 2017, esse modelo forneceu predições cerca de 4.000 MW superiores ao valor real. Para reduzir esse erro, foram adotadas medidas como o envio de alertas em apenas 10 dias do mês, com 5 a 10 minutos de antecedência. Ainda assim, uma alta taxa de alarmes falsos foi observada. Dessa maneira, os autores propõem um novo modelo de predição de picos de energia, capaz de mostrar a probabilidade de um pico ocorrer nas próximas 24 horas. O modelo foi construído com base numa FDA (*função de distribuição acumulada*), que modela a probabilidade de ocorrência de um pico dentro das próximas 24 horas, a partir do ponto de predição. Essa FDA foi construída empiricamente para as observações reais, e unificada com dados históricos de consumo da ERCOT entre 2010 a 2016, combinado com dados de clima como temperatura, umidade, ve-

locidade do vento, direção do vento, visibilidade, pressão atmosférica, precipitação acumulada para cada uma das 19 maiores cidades atendidas pela ERCOT. Este intervalo de 2010 a 2016 foi utilizado para treinamento de uma rede ANN, enquanto os dados de 2017 foram utilizados para o teste do modelo. A avaliação do desempenho do regressor foi feita por um indicador de limite de perda, que consiste na somatória dos desvios da FDA empírica com a FDA prevista pelo regressor, dividido pela FDA empírica. Por último, foi proposto o recall binário, com definição similar à precisão binária, com a diferença que o somatório destes desvios é dividido pelo FDA previsto. Analisando a precisão binária, o modelo proposto supera o modelo de regressão quantílica em todos os intervalos da função de distribuição acumulada. Já o recall binário evidencia que o modelo fornece valores para a função de distribuição acumulada acima de 0,98 em pelo menos 50% do tempo.

No trabalho de Gajowniczek et al. (2017), é proposta uma abordagem para determinar se um certo valor de consumo de energia elétrica pode ser considerado um pico ou não. O método consiste em pré-processar as séries temporais, agrupando os valores em um percentil, onde somente os valores posicionados no percentil 99 ou superior deve ser considerado um pico. Nesse trabalho, são considerados dados históricos obtidos do sistema de distribuição de energia da Polônia, entre 1 de janeiro de 2008 a 31 de dezembro de 2015, gerando um montante de 70.128 observações, separadas em 1 hora entre si. Foram extraídas características dos dados como a hora da medição, dia do mês, dia da semana, indicador do mês, indicador de feriado, carga média observada nos intervalos anteriores e carga no mesmo período da semana anterior. Também foram consideradas medidas de temperatura e umidade durante o período, já que estas variáveis externas influenciam diretamente na magnitude dos picos, principalmente em momentos de alto consumo observados durante o verão. Para predição dos picos, foi proposto um classificador baseado em ANN, utilizando os dados de 2008 a 2013 para treinamento, dados de 2014 para validação e 2015 para testes. Foram utilizadas as medidas de acurácia e especificidade para avaliação do desempenho do classificador. O classificador proposto foi comparado com um modelo baseado na função quantil, em que a magnitude dos picos estão localizadas no 99° quantil. Após uma etapa de otimização dos valores dos parâmetros da ANN, o classificador obteve 90,5% de acurácia e 96,2% de especificidade, modelado por uma rede neural de nove neurônios.

Em Chemetova et al. (2016), é realizada a tarefa de predição de picos de consumo nas imediações das cidades de Évora e Porto em Portugal. Para isso, são utilizados dados coletados pelos sistemas de medição de consumo nos transformadores utilizados pelo sistema de distribuição de energia. Para predição da magnitude dos picos, foi treinado um regressor baseado ANN, com uma camada oculta, utilizando o ano de 2013 para treinamento e o de 2014 para validação. A partir dessa base, foram extraídas 13 características, onde cada uma representa um valor de consumo posicionado num intervalo de 1 semana. As características foram selecionados a partir de uma análise de autocorrelação, determinando uma sequência de características que apontam tendência de evolução. Outras variáveis, como fatores climáticos, fo-

ram consideradas no estudo, mas foram descartadas por não possuírem correlação com os valores de consumo. Utilizando a medida MAPE, o trabalho concluiu que o regressor obteve 0,78% na predição do pico um dia à frente, e 1,01% uma semana à frente.

3.3 Predição de picos em diferentes aplicações

Como a quantidade de trabalhos que realizam a predição de picos em dados de consumo de energia é limitado, foram explorados trabalhos que realizam a tarefa em diferentes aplicações, como a predição da atividade magnética do sol, predição de irradiância solar sobre uma certa região e predição de enchentes. Estes trabalhos são discutidos a seguir.

Em Prasad et al. (2022), discute-se sobre o ciclo magnético do sol e seus impactos sobre as operações dos astronautas no espaço, satélites em órbita, sistemas de radar de comunicação terra-satélite e falhas de comunicação em sistemas de navegação de GPS. A tarefa de prever a atividade magnética do sol é fundamental para tomar decisões que evitem problemas nos sistemas afetados por este ciclo. Diversos indicadores podem ser utilizados para prever o ciclo solar, sendo o número de manchas solares observadas por um longo período de tempo um dos mais utilizados. Na literatura existem diversos métodos capazes de prever a amplitude do próximo ciclo, desde métodos estatísticos à modelos de aprendizagem de máquina treinados com dados de séries temporais que representam o ciclo magnético solar em períodos anteriores. Nesse trabalho, os autores propõem um modelo baseado em Deep Learning e redes LSTM, capaz de prever o mais forte ciclo magnético do sol, também conhecido como ciclo 25. A base de dados utilizada foi fornecida pelo Observatório Real da Bélgica, em Bruxelas, que possui 3.240 observações de manchas solares registradas entre janeiro de 1750 e dezembro de 2019. Para assegurar o desempenho do modelo, a rede foi configurada e treinada utilizando observações de manchas solares existentes entre 1898 e 2008, considerando 1320 meses para treinamento e 132 meses para teste. Desta maneira, a rede foi capaz de prever a magnitude do pico e posição dos ciclos 20 à 24, sendo que no ciclo 24 o modelo obteve um RMSE (Raiz do Erro Quadrático Médio) de 4,22 e um erro relativo de 1,99%. Complementando o trabalho, a rede foi treinada com observações de 1320 meses dos ciclos ocorridos entre 1910 e 2019, e 132 meses neste mesmo período para fins de teste. Nesta configuração, o modelo prediz que o pico do ciclo será de $171,9 \pm 3,4$ e ocorrerá em agosto de 2023 ± 2 meses.

Em Vanderstar et al. (2018), são discutidas soluções para a estimativa do índice GHI (*Global Horizontal Irradiance*). O índice representa a quantidade de irradiância direta e difusa que atinge a superfície da Terra e de onde se extrai energia elétrica a partir da energia solar por meio de placas fotovoltaicas. Estimar a quantidade de irradiância solar sobre uma certa região é um fator crítico para o planejamento de um sistema de captação de energia fotovoltaica, pois possibilita um dimensionamento mais preciso da quantidade e do tamanho das células fotovoltaicas. Além

disso, permite planejar com melhor eficiência um sistema de backup, utilizado nos momentos onde não é possível abastecer o sistema por meio da energia solar, como nos dias de céu encoberto ou durante a noite. Nesses períodos, a sustentação do sistema pode ocorrer através de baterias ou outras fontes de energia alternativas, como energia eólica ou demais sistemas baseados em métodos não renováveis. Prever a irradiação sobre uma região também é bastante importante para estipular o melhor momento de carga e descarga das baterias de suporte, aumentando a vida útil deste componente, diminuindo as chances das baterias estarem descarregadas nos momentos onde o sistema deveria ser abastecido por meio delas. O trabalho de Vanderstar et al. (2018) propõe um modelo baseado em ANN para prever a irradiação solar horizontal com 2 horas de antecedência. O modelo foi treinado utilizando dados obtidos por estações de monitoramento de irradiação solar, espalhados pela região noroeste da cidade de Alberta, no Canadá. Dados históricos destas estações foram fornecidos pelo Banco de Dados de Irradiação Nacional Solar do Canadá (NSRDB - *National Solar Radiation Database*), entre os anos de 1998 e 2015, onde cada medida foi realizada a cada meia hora. O modelo foi treinado utilizando dados de 1998 à 2014, com uma medida a cada hora, e o ano de 2015 foi utilizado integralmente para testes. A fim de otimizar o desempenho do modelo, foi aplicado algoritmo genético para encontrar a melhor combinação de estações de monitoramento a serem utilizadas no treinamento. Como resultado, o algoritmo selecionou 5 estações dentre 20 disponíveis. Nesta configuração, a ANN obteve um MAPE de 7,06% e um RMSE normalizado (NRMSE) de 10,8%, comparando a magnitude do pico encontrado pelo modelo ANN com o valor observado na base durante o ano de 2015.

O trabalho de Feng et al. (2020) apresenta o uso de modelos de aprendizagem de máquina para a previsão de enchentes e secas. Enchentes foram a primeira causa de mortes nos Estados Unidos no ano de 2014, nos eventos relacionados ao clima. Enquanto secas são responsáveis por perdas financeiras da ordem de 6 bilhões de dólares anualmente. A previsão da ocorrência desses fenômenos é fundamental para a tomada de decisões preventivas com impacto direto na sociedade. Os autores propõem um modelo baseado em redes LSTM, que tornaria possível uma aplicação em escalas continentais e não somente em determinadas regiões ou épocas do ano. O trabalho teve como base os dados fornecidos pela CAMELS (*Catchment Attributes and Meteorology for Large-sample Studies*, que estrutura séries temporais das médias das bacias hidrometeorológicas, juntamente com atributos de captação como topografia, características climáticas, cobertura por terra, solo e características geográficas, além de observações do fluxo de correntes de 671 captações realizadas em toda a extensão do território dos Estados Unidos. A maioria das medidas estão concentradas entre os anos de 1980 e 2014. O modelo construído foi baseado no estudo de Fang et al. (2018), que propôs uma rede LSTM para prever a umidade do solo sob determinadas condições climáticas e regionais. A rede foi treinada a partir de sequências de 365 dias com 6 características meteorológicas e 17 características estáticas, como profundidade do solo, máximo conteúdo de água, porosidade da superfície, entre outras. A variável de saída da rede fornece o valor da descarga de

água prevista 3 dias à frente. Para fins de comparação, o desempenho deste modelo foi comparado com uma rede ANN e ao modelo hidrológico SAC-SMA (*Sacramento Soil Moisture Accounting Model*) fornecido pela CAMELS. Para comparação dos modelos, foi aplicado o NSE (*Nash & Sutcliffe Model Efficiency*), frequentemente utilizado para comparar o desempenho de predição de modelos hidrológicos. O valor de NSE costuma variar entre 0 e 1, onde quanto mais próximo de 0 significa um modelo com capacidade preditiva próxima da média da série temporal em termos da soma do erro quadrado e quanto mais próximo de 1 orienta para um modelo capaz de prever valores próximos ao observado. O NSE médio observado para o modelo baseado em LSTM foi de 0,86, mais alto que 0,77 obtido pelo modelo baseado em ANN, e 0,64 obtido pelo modelo SAC-SMA. Esta avaliação sofreu algum tipo de variação conforme a região em estudo. Por exemplo, nas regiões montanhosas ou dominadas por neve, o modelo baseado em LSTM foi constantemente superior aos modelos de comparação. Entretanto, em regiões com baixa precipitação, o modelo foi praticamente equivalente aos demais modelo de comparação.

Le et al. (2021) também aborda o problema de predição de enchentes, mas utiliza a bacia do rio Vermelho no Vietnã como estudo de caso. O rio está localizado próximo a cidade de Hanói e sofre influência de outras 3 afluentes que podem potencialmente causar enchentes e transtornos a um grande número de pessoas. As características hidrológicas desta região orientam para uma forte influência das estações do ano sobre o escoamento deste sistema de rios. Além disto, a complexidade do sistema aponta para um cenário onde os picos de escoamento podem não ocorrer ao mesmo tempo ou na mesma região, sendo desta forma uma tarefa complexa combinar características relevantes para determinar o pico de escoamento em cada região, que poderia potencialmente resultar numa situação de enchentes próximas da cidade de Hanói. Para prever situações de enchentes, os autores propõem um modelo baseado em redes LSTM capaz de retornar o escoamento previsto na bacia do rio Vermelho, utilizando dados de séries temporais coletados de 7 estações hidrológicas posicionadas no sistema, entre os anos de 1995 e 2014. Diferentemente de outros trabalhos, os autores não propõem uma combinação dos dados de séries temporais e características externas como nível de água, características da topografia ou umidade do solo, devido a dificuldades em coletar estas informações em diversas regiões que compõem sistema. Os dados de 1995 a 2013 foram utilizados para treinamento da rede LSTM, e o ano de 2014 para testes do modelo. Para efeitos de comparação do desempenho do modelo, foram propostos avaliações juntamente com um modelo baseado em ANN e outro baseado em uma rede CNN de três camadas, onde a primeira camada executa uma convolução, a segunda executa um pooling e a terceira corresponde a porção *fully-connected*. Dos resultados obtidos, o modelo baseado em redes LSTM superou os demais modelos em comparação no cenário de predição do pico de escoamento de 1 dia à frente, fornecendo valores de NSE da ordem de 0,988, enquanto os modelos baseados em ANN e CNN fornecem valores de NSE da ordem de 0,982 e 0,978, respectivamente. No cenário de picos de escoamento de 2 dias à frente, novamente as redes LSTM demonstraram desempenho superior aos demais

modelo, fornecendo um valor de NSE da ordem de 0,930, enquanto ANN e CNN obtiveram valores de NSE da ordem de 0,908 e 0,897 respectivamente.

3.4 Discussão

Em geral, os trabalhos da literatura que realizam a tarefa de predição de picos abordam o problema com soluções de regressão convencionais, em que a saída retornada representa somente a magnitude do pico. Em relação à predição da localização dos picos, é usual o emprego de métodos estatísticos, como FDA ou faixas específicas de um percentil. Ao utilizar métodos de aprendizagem de máquina, os trabalhos abordam o problema de predição da localização dos picos como uma tarefa de classificação. Poucos trabalhos da literatura consideram a predição tanto da magnitude como da posição do pico e, quando consideram, tratam as tarefas como independentes. Desse modo, métodos de regressão *multi-target* capazes de considerar a relação entre ambas as variáveis, ainda não foram explorados para esta tarefa.

Nota-se que boa parte dos trabalhos de predição de picos realiza a extração de características dos dados sequenciais para a indução dos modelos preditivos ou utilizam características obtidas de fontes variadas de dados, como temperatura e umidade. Além de representar um custo adicional para o cálculo dos atributos em uma etapa de pré-processamento, é necessário encontrar um conjunto adequado de características de acordo com o problema. Assim, atributos relevantes para uma determinada aplicação podem ser pouco efetivos em outra aplicação. Para que seja desenvolvida uma solução independente da aplicação, pode-se utilizar os dados brutos de séries temporais como entrada (atributos) para a construção do modelo preditivo, conforme proposto neste projeto. Ainda, para que seja considerada a relação dos atributos com o tempo, pode-se considerar o uso de métodos de regressão extrínseca para a construção do modelo.

Na escolha dos algoritmos de aprendizagem de máquina, observa-se que grande parte dos trabalhos utiliza redes ANN para a construção dos modelos propostos ou para a construção dos modelos de comparação. Também é possível notar a popularidade de modelos de regressão baseados em Deep Learning, como as redes LSTM e suas variações. Nestes casos, é importante salientar a necessidade de um conjunto de dados maior, fazendo com que os trabalhos considerem dados de diferentes fontes.

Uma limitação compartilhada pela maioria dos trabalhos discutidos anteriormente, é a ausência da disponibilização dos conjuntos de dados avaliados, o que dificulta uma comparação direta de novas soluções com os resultados da literatura. Neste projeto, todos os algoritmos implementados e conjuntos de dados serão disponibilizados publicamente, permitindo a comparação futura de resultados obtidos por métodos de predição de picos.

3.5 Considerações finais

A literatura levantada apresenta diferentes abordagens utilizadas para prever picos de quaisquer natureza. Alguns métodos se baseiam em métodos estatísticos, alguns em heurísticas e outros em aprendizagem de máquina. Porém, poucos trabalhos exploram a predição de picos como um problema de aprendizagem de máquina que utiliza somente os dados brutos de séries temporais, requerendo a coleta de outras variáveis externas a aplicação. Além disso, a maioria dos trabalhos aborda a predição dos valores como tarefas independentes. Este trabalho possui uma abordagem diferenciada, que é propôr um método que utiliza somente os dados brutos de séries temporais para prever magnitude e posição dos picos em uma série temporal, por meio da mineração de dados brutos das séries, utilizando técnicas que visam explorar a correlação entre os atributos de entrada e a correlação entre os atributos-alvo.

Capítulo 4

Construção e Análise de Dados

4.1 Considerações iniciais

Neste projeto, a tarefa de predição de picos possui relação direta com dados de consumo de energia elétrica coletados por medidores inteligentes (*smart meters*). Esses dispositivos coletam informações de consumo de cada unidade de maneira frequente, permitindo formar conjuntos de dados do padrão de consumo de cada unidade em alta resolução, facilitando a construção de modelos preditivos. Os *smart meters* são peça fundamental na construção de *smart grids*, que são o estado da arte de sistemas de distribuição de energia elétrica, visando o aumento da eficiência, redução de falhas e custos de distribuição.

O presente capítulo apresenta uma descrição e análise dos conjuntos de dados construídos e avaliados durante o desenvolvimento deste projeto, descrevendo as principais características presentes nas informações que auxiliam a entender o padrão de comportamento das séries temporais de consumo.

4.2 Dados de consumo de energia elétrica

A tarefa de predição de picos, anteriormente apresentada na Seção 2.4, será conduzida em dados de consumo de energia elétrica fornecidos pela LADPU - *Los Alamos Department of Public Utilities*¹. A LADPU é uma concessionária pública e sem fins lucrativos que fornece eletricidade, gás natural e água para a cidade de Los Alamos, Estados Unidos. Los Alamos é uma cidade localizada no estado do Novo México ao sul dos Estados Unidos, conforme ilustrado na Figura 4.1, com uma população de aproximadamente 13 mil habitantes e cerca de 5.300 residências.

¹<https://www.losalamosnm.us/government/departments/utilities>

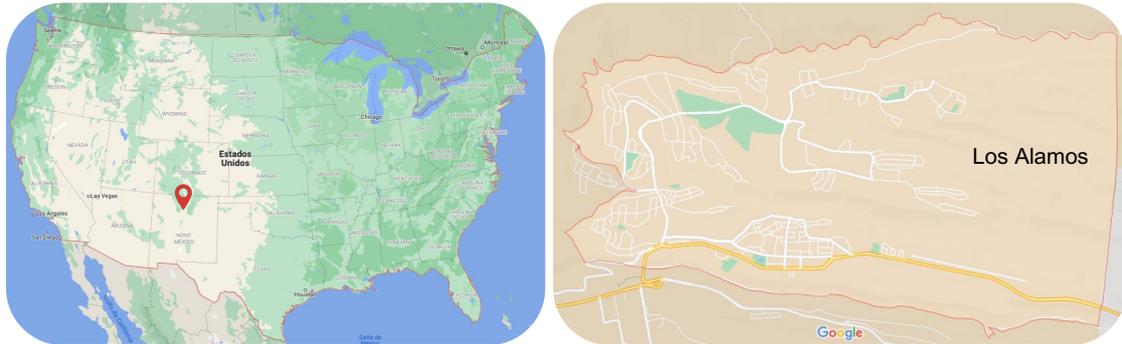


Figura 4.1: Localização da cidade de Los Alamos, NM, EUA, onde os dados de consumo de energia elétrica foram coletados e fornecidos pela concessionária LADPU.

O acesso aos dados de consumo de energia elétrica da LADPU foi obtido a partir de uma parceria entre pesquisadores da Pontifícia Universidade Católica do Paraná (PUCPR), *University of New Mexico* (UNM) e *The New Mexico Established Program to Stimulate Competitive Research* (NM EPSCoR)² no âmbito do projeto *NM smart grid Center*, um projeto interdisciplinar envolvendo diferentes instituições de pesquisa e órgãos públicos como a LADPU.

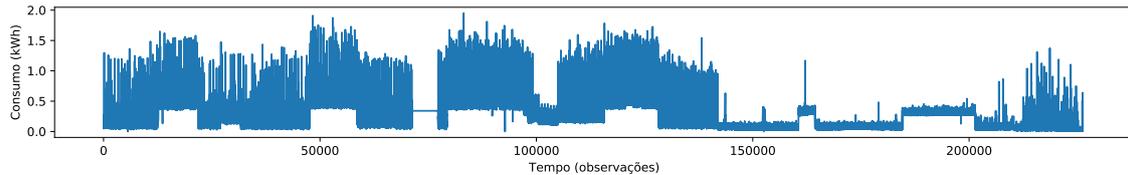
O conjunto de dados contempla informações de consumo de 1.757 residências com medidores inteligentes de energia Landis+Gyr, similares ao modelo ilustrado na Figura 4.2. Este dispositivo é responsável por realizar uma medição de consumo a cada 15 minutos, gerando 96 observações diárias. Para a maior parte dos usuários, os dados foram coletados durante um período de aproximadamente 6 anos, entre 30 de julho de 2013 a 31 de dezembro de 2019, gerando uma série temporal com mais de 200 mil observações para cada consumidor. Por questões de privacidade, os dados foram anonimizados para não vincular o padrão de consumo a identificação do consumidor. Atualmente, estes dados estão publicamente disponíveis para a comunidade científica no portal Dryad (Souza et al., 2020).



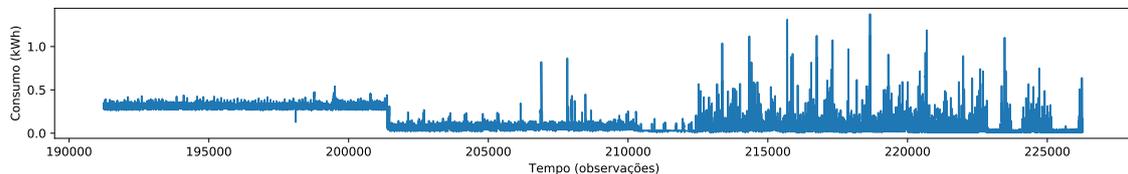
Figura 4.2: Medidor inteligente de energia (*smart meter*) Landis+Gyr E470.

²<https://www.nmepscor.org/>

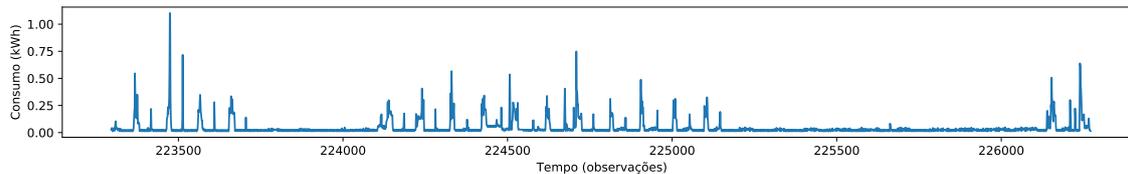
Na Figura 4.3, é apresentado os dados de um consumidor escolhido ao acaso, considerando: (a) todo o período de coleta de dados, (b) o consumo observado em 2019, (c) o consumo medido em dezembro de 2019 e (d) o consumo observado no último dia de coleta dos dados, em 31 de dezembro de 2019. É importante destacar que os dados originais possuem problemas como valores ausentes e medições duplicadas, que precisam ser tratados em uma etapa de pré-processamento, conforme será discutido na Seção 4.3.



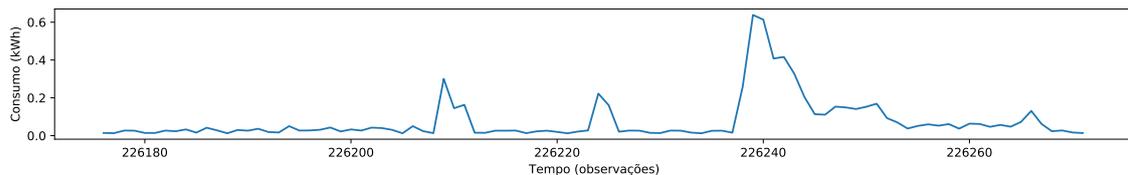
(a) Período entre 2013 e 2019



(b) Ano de 2019



(c) Dezembro de 2019



(d) 31 de dezembro de 2019

Figura 4.3: Dados de consumo de energia elétrica de um consumidor doméstico fornecidos pela LADPU considerando diferentes intervalos de tempo.

4.3 Pré-processamento e construção dos dados

Nesta seção, será apresentado o formato original dos dados e as etapas de pré-processamento realizadas para a construção dos conjuntos de dados utilizados para o treinamento e avaliação dos modelos preditivos explorados neste trabalho. Os dados de consumo de energia fornecidos pela LADPU são capturados por *smart meters* em

intervalos de 15 minutos e armazenados em um arquivo para cada consumidor. A Figura 4.4 ilustra o formato que os dados são originalmente armazenados. Embora o arquivo contenha diversas colunas de informação, somente as colunas *METER_FID*, *INTERVAL_TIME* e *INTERVAL_READ* são utilizadas neste trabalho. As colunas correspondem ao identificador do *smart meter* que realizou a coleta, data e hora da coleta e o valor de consumo ocorrido desde a última medição, respectivamente.

METER_FID	START_READ	END_READ	INTERVAL_TIME	INTERVAL_READ
49991	5174	16,49	10/07/2013 00:15	0,0924
49991	5174	16,49	10/07/2013 00:30	0,0924
49991	5174	16,49	10/07/2013 00:45	0,0888
49991	5174	16,49	10/07/2013 01:00	0,096
49991	5174	16,49	10/07/2013 01:15	0,084

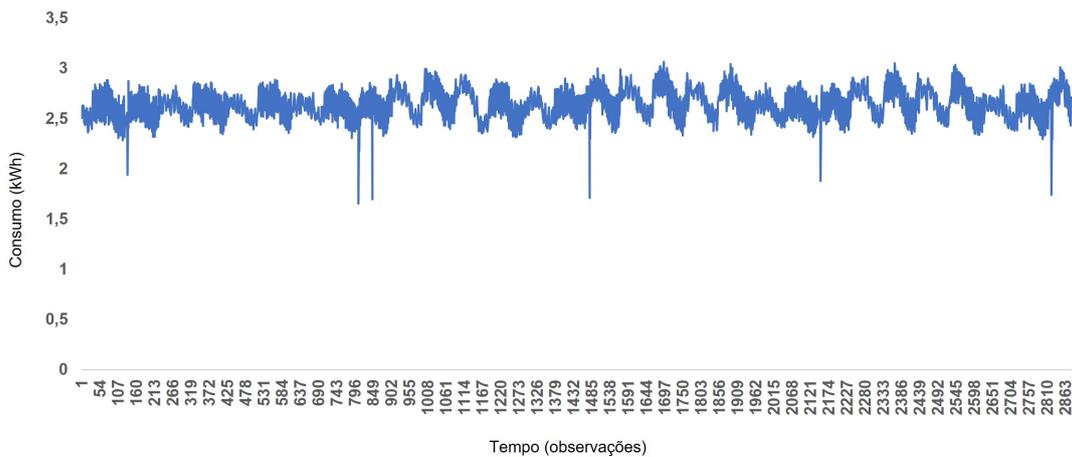
Figura 4.4: Primeiros registros do arquivo de dados de consumo obtidos pelo medidor identificado como 49991.

Devido a problemas durante a transmissão dos dados pela rede ou problemas relacionados ao seu armazenamento, o conjunto de dados apresenta medições incorretas como a ocorrência de valores ausentes em determinados períodos e o registro de mais de uma medição para um mesmo horário. Estes problemas foram tratados em uma etapa de pré-processamento. No caso de múltiplas medições, a maioria das ocorrências eram compostas por duas medições para um mesmo horário e, em alguns casos, a segunda medição era composta pelo valor zero ou com exatamente o mesmo valor da primeira medição. Para o tratamento deste problema, foi mantida somente a primeira medição registrada e as demais foram descartadas. Para o tratamento de valores ausentes, foi necessário identificar os horários sem o registro de medições e então preencher estas posições com o valor médio calculado entre a observação anterior e a observação seguinte à posição com um valor ausente.

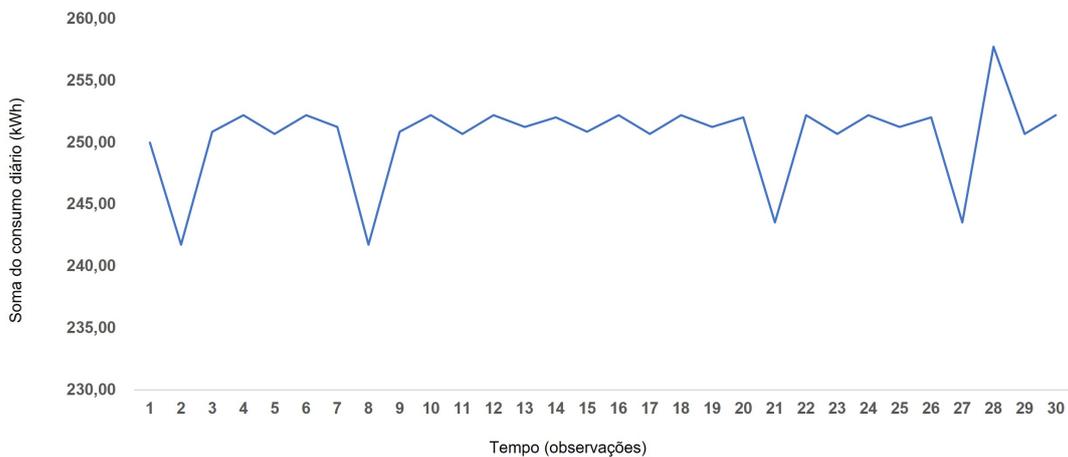
Embora o conjunto de dados original possua registros no período entre 2013 a 2019, optou-se por descartar os dados mais antigos pela elevada presença de valores ausentes em longas sequências (por exemplo, durante uma semana ou um mês) observada nos primeiros anos, além de priorizar dados mais recentes na tarefa de predição de picos. É importante destacar que o método de preenchimento de valores ausentes a partir da média das observações é efetiva em situações isoladas, em que poucos valores são faltantes. No caso de uma longa sequência de valores, esse procedimento pode levar a inclusão de um número elevado de valores incorretos, sendo necessário o uso de métodos mais adequados que fogem do escopo deste projeto. Desse modo, foram selecionados dados dos últimos três anos, entre 2017 e 2019. Especificamente, são considerados os dados de 2017 e 2018 para o treinamento dos modelos e dados de 2019 para a etapa de testes.

Preliminarmente, foi utilizada a série de consumo a cada 15 minutos para a construção dos modelos preditivos seguindo a abordagem de regressão extrínseca

para a tarefa de predição de picos. Entretanto, se confirmou que o uso dos dados neste formato não é suficiente para a construção de modelos precisos. Porém, a transformação das séries temporais com uma observação a cada 15 minutos em séries com o consumo diário acumulado, levou a geração de modelos mais precisos para a tarefa. Desse modo, após a limpeza dos dados, preenchimento dos valores faltantes e seleção dos dados mais recentes, a série foi sumarizada diariamente, somando todos os valores de consumo ocorridos em cada dia da semana, resultando em uma série de 1.095 observações (1 observação por dia durante três anos), conforme ilustrado na Figura 4.5.



(a) Série de consumo instantâneo a cada 15 minutos, durante 30 dias



(b) Série de consumo total diário, durante 30 dias

Figura 4.5: Transformação da série de consumo instantâneo em uma série com o consumo acumulado diário.

Para o treinamento dos modelos preditivos, é necessário transformar as séries temporais em matrizes de características de dimensões $M \times L$, em que M representa a quantidade de exemplos e L representa a quantidade de características descritivas

dos exemplos, sendo que os dois últimos valores de $L (y_1, y_2)$ são os atributos-alvo da tarefa de predição (magnitude do pico e o dia da semana do pico).

Para construir essa matriz, são utilizadas duas janelas de dados subsequentes w_i e f_i , que percorrem toda a série temporal de consumo de maneira sequencial conforme ilustrado na Figura 4.6. Todos os dados em w_i são utilizados como características descritivas dos exemplos e os atributos-alvo são obtidos a partir dos dados de consumo observados na janela f_i , que representam os valores futuros de w_i .

Partindo do pressuposto que boa parte dos consumidores possui padrões de comportamento que se repetem nos mesmos dias da semana, é provável que um consumidor realize atividades de maneira periódica (por exemplo, usar a máquina de lavar roupas por duas horas todo domingo próximo às 10 horas), definiu-se o tamanho de w_i em 7 dias, sendo possível capturar este padrão em um único exemplo. Como o objetivo da tarefa de predição é identificar um pico de consumo com 7 dias de antecedência, o tamanho da janela f_i é de 7 observações.

Cada deslocamento das janelas w_i e f_i pela série temporal de consumo é responsável pela geração de um novo exemplo na matriz de características. Para deslocar as janelas, considera-se um passo de 1 observação à frente, que representa o total de 1 dia de consumo. Assim, cada exemplo gerado representa uma predição de pico 7 dias à frente. Ao final do processo, é gerada uma matriz de características para cada um dos medidores com um total de 1.088 exemplos (1 observação por dia durante 3 anos), cada um com 7 atributos descritivos e 2 atributos-alvo.

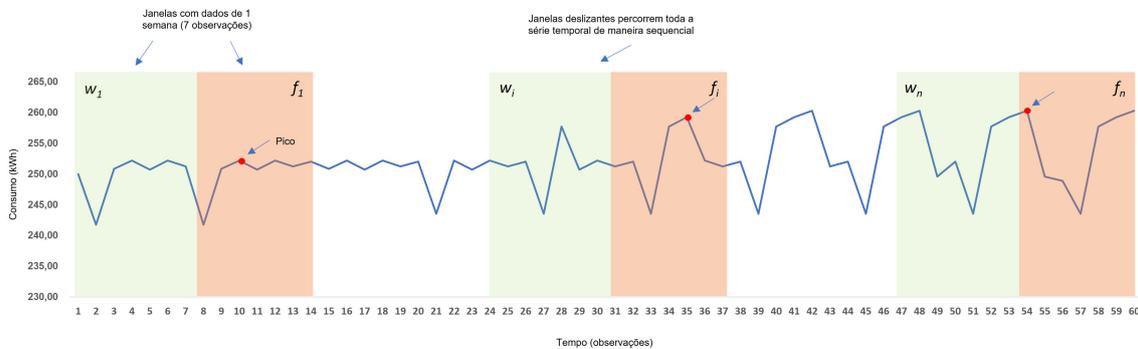


Figura 4.6: Janela deslizante que percorre toda a série temporal para formação da matriz de características.

A Figura 4.7 ilustra as primeiras observações da matriz de características gerada após as etapas de pré-processamento. Do lado esquerdo da matriz estão os atributos descritivos e do lado direito os atributos-alvo. Neste exemplo, é possível observar que as predições realizadas a partir da meia-noite do dia 01 de janeiro de 2017, levam a um pico de magnitude 20,762 que ocorrerá na quinta-feira³ da semana seguinte.

³Os dias da semana são representados numericamente de 0 a 6, onde 0 corresponde a segunda-feira e 6 o domingo

Ponto de Predição	0	1	2	3	4	5	6	Magnitude	Posição (dia da semana)
01/01/2017	20,438	20,203	20,050	32,557	24,728	20,369	20,524	20,762	3
02/01/2017	20,203	20,050	32,557	24,728	20,369	20,524	20,504	20,762	3
03/01/2017	20,050	32,557	24,728	20,369	20,524	20,504	20,258	20,762	3
04/01/2017	32,557	24,728	20,369	20,524	20,504	20,258	20,533	20,762	3
05/01/2017	24,728	20,369	20,524	20,504	20,258	20,533	20,527	20,762	3
06/01/2017	20,369	20,524	20,504	20,258	20,533	20,527	20,762	20,699	4
07/01/2017	20,524	20,504	20,258	20,533	20,527	20,762	20,699	20,586	5
08/01/2017	20,504	20,258	20,533	20,527	20,762	20,699	20,586	20,324	6
09/01/2017	20,258	20,533	20,527	20,762	20,699	20,586	20,324	20,257	0
10/01/2017	20,533	20,527	20,762	20,699	20,586	20,324	20,257	20,085	2
11/01/2017	20,527	20,762	20,699	20,586	20,324	20,257	19,826	20,085	2
12/01/2017	20,762	20,699	20,586	20,324	20,257	19,826	20,085	20,261	2
13/01/2017	20,699	20,586	20,324	20,257	19,826	20,085	19,903	20,261	2
14/01/2017	20,586	20,324	20,257	19,826	20,085	19,903	19,926	20,261	2

Figura 4.7: Matriz de características construída a partir dos dados de consumo.

Cabe ressaltar que esta base de dados da LADPU conta com 1.757 arquivos de dados de consumo, onde cada um representa o consumo de um cliente. Devido a restrições de tempo associados ao treinamento dos modelos para cada cliente, foi selecionado deste total um subconjunto de 100 consumidores, que representam os consumidores de maior consumo total de energia durante os três anos da análise. Desse modo, são geradas 100 matrizes de características conforme anteriormente ilustrado na Figura 4.7, sendo possível o treinamento individual de um modelo preditivo para cada um dos 100 consumidores selecionados. Nesta etapa de seleção, foram considerados os 100 consumidores que apresentam o maior consumo médio durante o período. Esta decisão se deu após a constatação, em experimentos preliminares, da dificuldade em prever picos em dados provenientes de consumidores com uma taxa reduzida de consumo. Nestes casos, pode-se observar que os picos não representam valores significativamente superiores ao consumo médio do cliente.

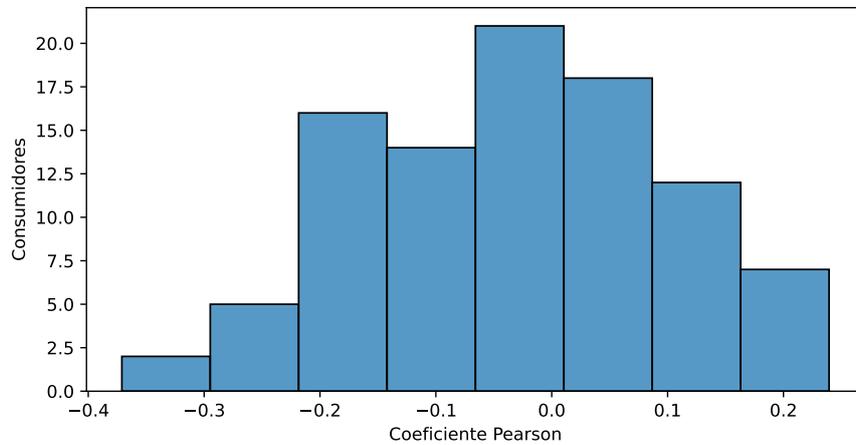
4.4 Correlação entre as variáveis-alvo

Esse trabalho propõe avaliar métodos de regressão *multi-target* como um mecanismo para aprimorar os resultados da predição, dada a premissa que existe correlação entre o valor de um pico e a sua posição. Tal premissa se dá pela intuição que consumidores repetem seus padrões de consumo ao longo das semanas, gerando valores de consumo similares em determinados dias da semana.

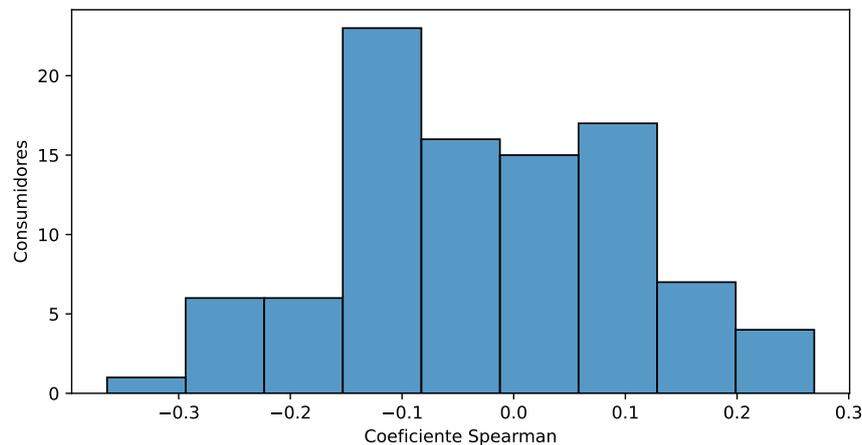
É importante ressaltar que os algoritmos de regressão *multi-target* exploram a correlação entre as variáveis-alvo, definindo novos conjuntos de dados que incluem as variáveis-alvo como características, ou ainda, estabelecendo um critério que condiciona a inclusão de uma variável-alvo no conjunto de características com a correlação existente dela com as demais variáveis.

Desse modo, foi calculada a correlação existente entre as variáveis-alvo magnitude e posição para o conjunto de 100 consumidores selecionados, utilizando os métodos de *pearson* e *spearman* para avaliar uma opção de método linear e não-linear. Por convenção, o trabalho definiu que haverá correlação entre as variáveis somente se ambos os coeficientes tiverem o valor superior a 0.6 em termos absolutos.

Para o conjunto de 100 consumidores, em nenhum deles os coeficientes de *pearson* e *spearman* atestaram correlação, conforme apresentado na Figura 4.8. Sendo assim, a análise foi ampliada para toda a base de 1.757 consumidores e, da mesma maneira, não foi encontrada correlação em nenhum desses conjuntos de dados. Desse modo, foi necessário definir um mecanismo que crie correlação entre as variáveis-alvo, de modo que os algoritmos de regressão *multi-target* forneçam melhores resultados.



(a) Distribuição do coeficiente Pearson para 100 consumidores.



(b) Distribuição do coeficiente Spearman para 100 consumidores.

Figura 4.8: Correlação entre os atributos-alvo observada entre os 100 consumidores.

O mecanismo explorado por este trabalho foi criar uma segunda matriz de características, utilizando os valores-alvo encontrados na etapa de pré-processamento

como características. Cada linha dessa nova matriz representa uma sequência de valores de pico e a variável-alvo é definida pelo valor de pico ocorrido imediatamente após o último valor da sequência. Desse modo, a sequência representa uma série de picos ocorridos em 7 dias e o valor da variável-alvo representa o valor de pico ocorrido 1 dia à frente desta série. As sequências são espaçadas uma posição adiante na série de picos, que representa 1 dia. O procedimento é executado para cada série de valores de cada variável-alvo. Desse modo, são geradas 2 matrizes de características, que representam as séries de picos de magnitude e posição, bem como as variáveis-alvo.

O trabalho avaliou os resultados obtidos quando se utiliza as matrizes de cada variável-alvo isoladamente, bem como os resultados utilizando uma matriz concatenada de todas as variáveis, além de outros mecanismos que explorem a correlação entre as variáveis. A Figura 4.9 ilustra uma matriz de características construída a partir da concatenação das matrizes de cada variável-alvo, que foi onde os algoritmos de regressão *multi-target* obtiveram seus melhores resultados. Maiores detalhes destes experimentos são detalhados no capítulo 6.

Ponto de Predição	0							1							Magnitude	Posição (dia da semana)
	0	1	2	3	4	5	6	0	1	2	3	4	5	6		
01/01/2017	20,7624	20,762	20,762	20,762	20,762	20,699	20,586	3	3	3	3	3	4	5	20,324	6
02/01/2017	20,7624	20,762	20,762	20,762	20,699	20,586	20,324	3	3	3	3	4	5	6	20,257	0
03/01/2017	20,7624	20,762	20,762	20,699	20,586	20,324	20,257	3	3	3	4	5	6	0	20,085	2
04/01/2017	20,7624	20,762	20,699	20,586	20,324	20,257	20,085	3	3	4	5	6	0	2	20,085	2
05/01/2017	20,7624	20,699	20,586	20,324	20,257	20,085	20,085	3	4	5	6	0	2	2	20,261	2
06/01/2017	20,6994	20,586	20,324	20,257	20,085	20,085	20,261	4	5	6	0	2	2	2	20,261	2
07/01/2017	20,586	20,324	20,257	20,085	20,085	20,261	20,261	5	6	0	2	2	2	2	20,261	2
08/01/2017	20,3244	20,257	20,085	20,085	20,261	20,261	20,261	6	0	2	2	2	2	2	20,261	2
09/01/2017	20,2572	20,085	20,085	20,261	20,261	20,261	20,261	0	2	2	2	2	2	2	20,312	6
10/01/2017	20,085	20,085	20,261	20,261	20,261	20,261	20,312	2	2	2	2	2	2	6	20,312	6
11/01/2017	20,085	20,261	20,261	20,261	20,261	20,312	20,312	2	2	2	2	2	6	6	20,312	6
12/01/2017	20,2614	20,261	20,261	20,261	20,312	20,312	20,312	2	2	2	6	6	6	6	20,499	2
13/01/2017	20,2614	20,261	20,261	20,312	20,312	20,312	20,499	2	2	2	6	6	6	2	20,499	2
14/01/2017	20,2614	20,261	20,312	20,312	20,312	20,499	20,499	2	2	6	6	6	2	2	20,499	2

Figura 4.9: Matriz de características construída a partir dos picos.

4.5 Análise dos dados

A partir do conjunto de dados de 100 consumidores, foi realizada uma análise levando em consideração os valores observados nos atributos-alvo dos exemplos (magnitude e horário dos picos). A Figura 4.10 ilustra um histograma com a distribuição da frequência da posição dos picos. É possível perceber que na maioria das vezes, os dias da semana que ocorre um pico é sábado e domingo. Com relação a magnitude do pico, percebe-se que este valor se concentra em torno de 40 kW, conforme ilustrado na Figura 4.11. Uma análise de ambas as variáveis em conjunto é fornecida na Figura 4.12.

Ao analisar o comportamento das magnitudes conforme os meses do ano, percebe-se que o padrão desta distribuição pouco se altera. Neste caso, nota-se um comportamento sensivelmente diferente nos meses de janeiro, fevereiro e março, em relação

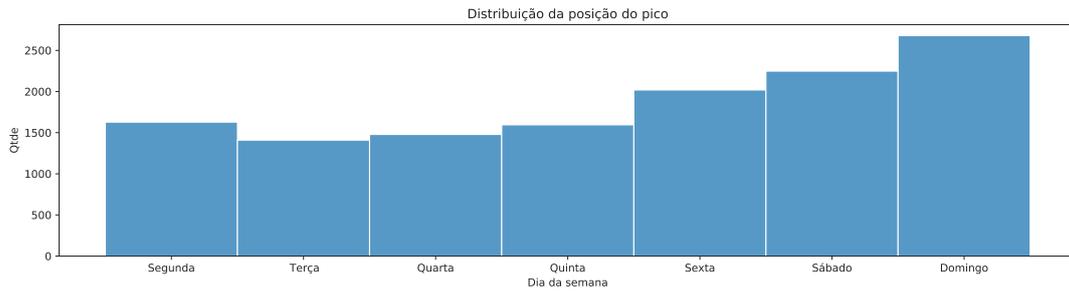


Figura 4.10: Distribuição da posição do pico para a base de 100 consumidores.

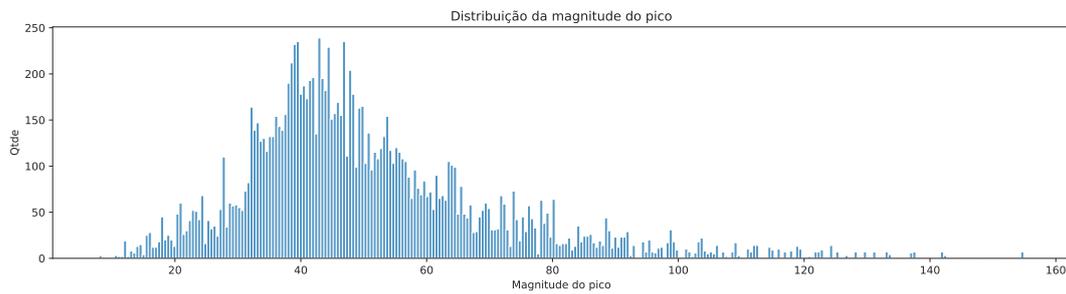


Figura 4.11: Distribuição da magnitude do pico para a base de 100 consumidores.

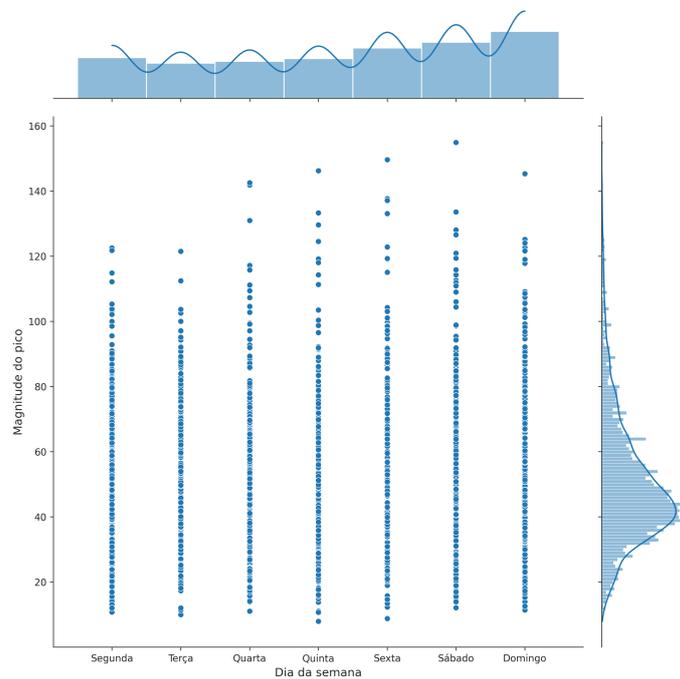


Figura 4.12: Análise da magnitude do pico em função do horário.

aos demais meses, conforme ilustrado na Figura 4.13, à esquerda. O mesmo ocorre

ao considerar os diferentes dias da semana, conforme ilustrado na Figura 4.13, à direita.

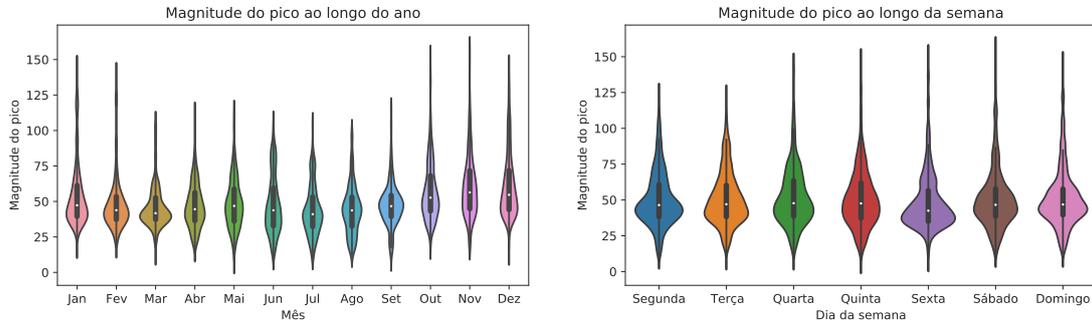


Figura 4.13: Magnitude dos picos de acordo com os meses e dias da semana.

4.6 Considerações finais

Neste capítulo, foi realizada uma revisão da literatura em relação ao estudo de caso abordado neste projeto, que são as redes elétricas inteligentes. Além disso, a tarefa de predição de picos foi formalmente definida.

A coleta de dados de consumo de energia elétrica por meio de *smart meters* gera a possibilidade de minerar estes dados para a realização de diferentes tarefas, como a predição de picos de consumo, conforme explorado neste projeto. Nesta tarefa, pode-se considerar métodos baseados em aprendizagem de máquina que levam em consideração padrões observados no consumo passado dos clientes. Para o uso destes métodos, é essencial a construção de conjuntos de dados um formato que seja adequado para esses algoritmos, conforme foi apresentado neste capítulo.

Em específico, este projeto busca explorar a correlação entre a magnitude e a posição do pico para a tarefa de predição, por meio de um método baseado na regressão extrínseca *multi-target*, que promete entregar resultados de predição melhores caso exista correlação entre as variáveis-alvo do problema em análise.

Também foi apresentado um mecanismo para melhorar o desempenho de predição, que consiste geração de uma segunda matriz de características construída a partir dos valores das variáveis-alvo determinadas a partir dos dados de consumos brutos da série de consumo. No Capítulo 6, os resultados obtidos por meio deste mecanismo são discutidos.

Capítulo 5

Predição de picos de consumo diário de energia

5.1 Considerações iniciais

Neste capítulo, são discutidos os principais aspectos da abordagem de pesquisa da tarefa de predição de picos de consumo de energia, utilizando regressão extrínseca e métodos de *multi-target*. Uma vez apresentado os conjuntos de dados na Seção 4.3, abaixo são discutidas as técnicas de regressão *multi-target* empregadas nesta pesquisa, bem como as métricas de avaliação utilizadas nas comparações de desempenho. A escolha das métricas está baseada no que geralmente se utiliza nos trabalhos relacionados. Da mesma maneira, são apresentados os algoritmos utilizados como regressor base dos modelos *multi-target*. Neste capítulo, também são apresentadas as questões que esta pesquisa responde, em conjunto com uma metodologia de trabalho que estabelece a maneira como as questões foram endereçadas durante todo o processo de pesquisa.

5.2 Objetivos

Este projeto tem como principal objetivo avaliar métodos de regressão para a tarefa de predição de picos em dados de séries temporais. Os métodos devem ser capazes de prever de maneira antecipada e com precisão, tanto a magnitude como o instante de tempo que o pico ocorrerá nas observações futuras da série temporal.

Para que a solução proposta seja independente da aplicação responsável pela geração dos dados, foram explorados algoritmos de regressão extrínseca que consideram os dados brutos das séries temporais como entrada do modelo ou a partir da extração de atributos das séries, mas sem a necessidade de considerar atributos provenientes de fontes externas aos dados do problema. Exemplos de algoritmos de

classificação que foram recentemente adaptados para o cenário de regressão extrínseca e apresentam resultados competitivos na literatura são os algoritmos Rocket, *Inception Network* e *Fully Convolutional Neural Networks*. Além destes algoritmos, também foram utilizadas abordagens que consideram a relação de dependência dos exemplos com o tempo, como o LSTM e o biLSTM.

Por se tratar de uma tarefa de predição com dois atributos-alvo (magnitude e posição do pico), também foram explorados algoritmos de regressão *multi-target*. Tais algoritmos exploram a correlação das variáveis de saída para a obtenção de resultados mais precisos. Em específico, foram explorados algoritmos *multi-target* estado-da-arte como o ERC - *Ensemble of Regression Chains*, o SST - *Stacked Single-Target* e o DSTARS - *Deep Structure for Tracking Asynchronous Regressor Stacking* em conjunto com os algoritmos de regressão extrínseca mencionados anteriormente.

Desse modo, a tarefa de predição de pico em séries temporais foi abordada neste projeto como um problema de *regressão extrínseca multi-target*. Para a avaliação dos métodos, foi considerada uma aplicação em redes elétricas inteligentes (*smart grids*). Nesta aplicação, o objetivo é realizar a predição do pico de consumo de energia elétrica (magnitude e dia da semana do pico) com uma semana de antecedência, considerando dados de 100 consumidores.

O desenvolvimento deste projeto visa responder um conjunto de questões de pesquisa relacionadas a tarefa de predição de picos e a aplicação estudada envolvendo dados de consumo de energia, tais como:

1. A combinação de algoritmos que consideram a correlação entre os atributos de entrada (regressão extrínseca) e a correlação entre os atributos-alvo (regressão *multi-target*), pode levar a resultados que superam abordagens de regressão mais simples que não consideram tais correlações?
2. Dentre os diferentes algoritmos de regressão e regressão *multi-target*, qual a combinação de algoritmos de ambas as abordagens leva aos melhores resultados preditivos para a tarefa de predição de picos?
3. Qual o impacto de se utilizar diferentes combinações de características extraídas das séries temporais? Por exemplo, utilizar a soma do consumo diário, a magnitude do pico, a posição do pico, ou uma combinação destes fatores?
4. Qual o impacto do tamanho dos exemplos utilizados no treinamento do modelo preditivo (por exemplo, dados de consumo de 1 dia, 1 semana, 2 semanas ou 1 mês) nos resultados de predição?
5. Para o problema de predição de picos, a melhor abordagem é a construção de um modelo individual para cada consumidor ou um modelo global com dados de diferentes consumidores?

A seguir, são discutidas as diferentes etapas consideradas no desenvolvimento deste projeto para cumprir com os objetivos apresentados e responder as questões de pesquisa levantadas.

5.3 Metodologia de desenvolvimento do trabalho

As primeiras etapas para o desenvolvimento deste projeto envolveram o levantamento e atualização bibliográfica sobre diferentes temas de estudo, com o objetivo de identificar o estado-da-arte em relação à tarefa de predição de picos e em relação a aplicação de redes elétricas inteligentes.

Inicialmente, foram explorados trabalhos relacionados que abordam a tarefa de predição de picos, que é o principal tema de estudo deste projeto. Neste levantamento, foram considerados trabalhos que utilizam métodos estatísticos ou de aprendizagem de máquina para realizar a tarefa de predição de picos em diferentes domínios, identificando aplicações que envolvem a predição de enchentes e secas, índice de irradiação solar e consumo de energia elétrica. Conforme discutido no Capítulo 3, as soluções propostas por estes trabalhos envolvem a extração de características dos dados e nem sempre processam dados de séries temporais, conforme é de interesse deste projeto. Uma limitação desta abordagem é a impossibilidade do uso de uma mesma solução em diferentes aplicações. Além disso, percebeu-se que a maioria dos trabalhos utilizam métodos convencionais de regressão em que somente um único valor é predito.

Dadas as limitações observadas pelos trabalhos relacionados, foi realizado um levantamento da literatura para identificar métodos adequados para a tarefa de predição proposta neste projeto e de acordo com as particularidades dos dados obtidos pela aplicação de redes elétricas inteligentes. Nessa direção, vem sendo realizado um estudo sobre as principais abordagens e algoritmos de regressão extrínseca e regressão *multi-target*, conforme discutido no Capítulo 2. Embora tenham sido identificados os algoritmos estado-da-arte de cada tema, é importante destacar a necessidade de atualização bibliográfica durante todo o processo de desenvolvimento deste projeto.

Após o levantamento bibliográfico dos principais temas, foram planejadas atividades de construção dos conjuntos de dados, bem como uma análise destes dados. Conforme discutido no Capítulo 4, foram realizadas atividade de pré-processamento de dados, como a seleção de 100 entre 1.757 consumidores com base no total de consumo de energia ao longo do período avaliado, preenchimento de valores ausentes, remoção de valores duplicados e janelamento para a construção das matrizes de características que serão processados pelos algoritmos. Com base nestes conjuntos de dados, foi realizada uma etapa de análise levando em consideração os atributos-alvo dos exemplos (magnitude e posição dos picos).

Após a etapa de construção e análise dos conjuntos de dados, foram realizados experimentos de predição de picos. Nesta etapa, foi avaliado o desempenho de

regressores convencionais como Linear Regression, SVR, XGBoost e o ANN. Nestes experimentos, os regressores foram avaliados considerando as tarefas de predição da magnitude e do pico como tarefas independentes. Além disso, estes regressores foram comparados com métodos de base (*baselines*), que levam em consideração somente informações do passado das séries, sem o uso de um modelo preditivo. Os resultados obtidos por estes regressores e os métodos de base foram utilizados como metas de desempenho que se espera superar com o desenvolvimento do trabalho. Maiores detalhes sobre os resultados são discutidos no Capítulo 6.

Os próximos passos incluíram a implementação de algoritmos de regressão extrínseca estado-da-arte com base na adaptação de algoritmos de classificação como o *Rocket*, *Inception Network* e o *Fully Convolutional Networks*, além de algoritmos que relacionam as grandezas com o tempo, como o LSTM e o biLSTM. Em seguida, foram implementados algoritmos de regressão *multi-target* como ERC, SST e DS-TARS. Após a implementação destes algoritmos, foram feitas adaptações necessárias para que os algoritmos *multi-target* sejam capazes de operar em conjunto com os algoritmos de regressão extrínseca.

Para determinar qual o conjunto de características fornece o melhor desempenho preditivo na aplicação em redes elétricas inteligentes, foram avaliadas seis diferentes conjuntos, conforme descrito a seguir:

- Dados brutos: leituras de consumo ocorridas a cada 15 minutos durante uma semana, totalizando 672 atributos por exemplo. Este conjunto contém as leituras brutas, não aplicando nenhum tipo de processamento posterior.
- Consumo total diário: soma do consumo total ocorrido dentro de um dia, durante uma semana. Cada exemplo contém 7 atributos, em que cada atributo representa o somatório de consumo diário.
- Magnitude do pico semanal: a partir do consumo total diário, é determinado qual o valor máximo de consumo observado 7 dias para trás. Para se ter o padrão de consumo dentro de uma semana, são gerados exemplos com 7 atributos, onde cada um representa o valor máximo de consumo ocorrido 7 dias para trás.
- Posição do pico semanal: da mesma maneira que o conjunto *magnitude do pico semanal*, porém, os exemplos armazenam a posição do pico ao invés da magnitude. A posição do pico corresponde a qual dia da semana ele foi observado, no intervalo de segunda a domingo.
- Catch22: a partir dos exemplos com o consumo bruto a cada 15 minutos durante 1 semana, calculam-se atributos utilizando método Catch22 (Lubba et al., 2019). O método efetua uma redução de dimensionalidade em cada amostra, resumindo 672 medições para 22 características como informação mútua, entropia, intervalo entre eventos extremos acima ou abaixo da média, periodicidade, etc.

- Rocket: a partir das amostras de consumo bruto, aplica-se a transformação Rocket (Dempster et al., 2020) conforme detalhado na Subseção 2.5.2. Utilizando uma configuração com 10 kernels, são gerados exemplos com 20 características.

Esses conjuntos de atributos foram considerados tanto individualmente como combinados, em alguns casos. Essas combinações foram codificadas, sequencialmente de A até G, para facilitar a compreensão. As combinações são apresentadas na Tabela 5.3. Por exemplo, a configuração *E* considera tanto a magnitude do pico semanal como a posição do pico semanal como atributos descritivos para o treinamento do modelo preditivo.

Característica	A	B	C	D	E	F	G
Dados brutos	✓					✓	✓
Consumo total diário		✓					
Magnitude do pico semanal			✓		✓		
Posição do pico semanal				✓	✓		
Catch22						✓	
Rocket							✓

Tabela 5.1: Combinações de características utilizadas nos experimentos de predição.

Por fim, foram conduzidos experimentos para coletar e ordenar métricas resultantes da construção e avaliação dos modelos com diferentes conjuntos de características considerando os dados de 100 consumidores, construindo modelos de predição individuais por consumidor e também um modelo global contendo todos os clientes. Além disso, foram realizados experimentos visando identificar o impacto do tamanho dos exemplos nos resultados, o impacto do horizonte de predição nos resultados e a comparação entre modelos individuais treinados para cada consumidor com um modelo global treinado com dados de diferentes consumidores. É importante destacar que para a verificação das questões de pesquisa, é necessário gerar novas versões dos conjuntos de dados considerando modificações no processo de construção das matrizes de características discutidas no Capítulo 4.

5.4 Métodos de avaliação

Para a comparação dos diferentes métodos e abordagens exploradas neste projeto, foi considerado um conjunto de dados com 100 consumidores selecionados dos dados fornecidos pela LADPU, conforme anteriormente discutido no Capítulo 4. A comparação de desempenho dos métodos leva em consideração os resultados obtidos em quatro diferentes medidas de avaliação para tarefas de regressão (no caso da predição da magnitude) e quatro medidas para classificação (no caso da predição do dia da semana - posição), conforme apresentado a seguir.

- MAE (*Mean Absolute Error*): o erro médio absoluto é o somatório da diferença absoluta entre o valor real (x_i) e o valor predito (y_i) de duas observações no tempo i , em uma série com D observações. A definição desta medida é apresentada na Equação 5.1, em que o valor 0 representa uma predição livre de erros.

$$MAE = \sum_{i=1}^D |x_i - y_i| \quad (5.1)$$

- MSE (*Mean Squared Error*): o erro médio quadrático é o somatório da diferença elevada ao quadrado entre o valor real (x_i) e o predito (y_i) de duas observações, em uma série de D amostras. A medida é definida de acordo com a Equação 5.2.

$$MSE = \sum_{i=1}^D (x_i - y_i)^2 \quad (5.2)$$

Da mesma maneira que o MAE, duas séries terão seus valores próximos quando o MSE for próximo de zero.

- MAPE (*Mean Absolute Percentual Error*): o erro médio absoluto percentual é o somatório da diferença absoluta entre o valor real (x_i) e o predito (y_i), dividido pelo valor real (x_i), em uma série de D amostras. O MAPE expressa em termos relativos o desvio entre duas observações. Da mesma maneira que MAE e MSE, quanto mais próximo de 0 estiver o valor do MAPE, mais próximas duas séries estarão, indicando predições próximas dos valores esperados. No denominador da equação do MAPE aplica-se um coeficiente e , que é um valor arbitrariamente pequeno utilizado para que se evite uma divisão por zero, caso o valor real seja zero. Nesta situação, o valor do MAPE pode se tornar arbitrariamente elevado. A medida MAPE é definida de acordo com a Equação 5.3.

$$MAPE = \frac{100}{D} \sum_{i=1}^D \left| \frac{x_i - y_i}{\max(e, x_i)} \right| \quad (5.3)$$

- R^2 score (Coeficiente de determinação): é uma medida que expressa a quantidade de variância dos dados, indicando o quão próximo os dados estão de uma linha de regressão ajustada. Seu valor varia entre 0 e 1, onde o mais próximo de 1 indica que as predições observadas estão próximas de seus valores reais. Em uma série de D amostras, onde x_i representa o valor real, y_i o valor predito e \hat{x} a média dos valores reais, o R^2 score é calculado da seguinte maneira (Equação 5.4):

$$R^2 = 1 - \frac{\sum_{i=1}^D (x_i - y_i)^2}{\sum_{i=1}^D (x_i - \hat{x})^2} \quad (5.4)$$

- **Acurácia:** é uma medida que determina a taxa de acertos de um modelo de classificação. Em uma série de D amostras, um modelo que produz VP predições verdadeiramente positivas e VN predições verdadeiramente negativas, a acurácia é calculada da seguinte maneira:

$$\text{Acurácia} = \frac{VP + VN}{D} \quad (5.5)$$

- **Precisão:** é uma proporção entre amostras verdadeiramente positivas e amostras classificadas positivamente pelo modelo. Em uma série de D amostras, um modelo que produz VP predições verdadeiramente positivas e FP predições falsamente positivas, a precisão é calculada da seguinte maneira:

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (5.6)$$

- **Recall:** é uma proporção entre amostras classificadas corretamente pelo modelo e as verdadeiramente positivas. Em uma série de D amostras, um modelo que produz VP predições verdadeiramente positivas e FN predições falsamente negativas, o recall é calculado da seguinte maneira:

$$\text{Recall} = \frac{VP}{VP + FN} \quad (5.7)$$

- **F1-score:** é a média harmônica da precisão e do recall. Em virtude disto, para um modelo de classificação obter um elevado valor de F1-score, é necessário que tanto a precisão como o recall sejam altos, sendo assim uma excelente medida para avaliar o desempenho de modelos de classificação. O F1-score é calculado da seguinte maneira:

$$\text{F1-score} = 2 \times \frac{\text{precisão} \times \text{recall}}{\text{precisão} + \text{recall}} \quad (5.8)$$

Estas medidas são frequentemente encontradas nos trabalhos relacionados com esta pesquisa e em trabalhos que avaliam o desempenho de modelos de regressão e classificação. Desta maneira, esta escolha é conveniente para a comparação dos resultados obtidos por esta pesquisa com trabalhos relacionados.

5.5 Considerações finais

O protocolo experimental definido pelo trabalho foi baseado em análises preliminares, que apontaram os algoritmos que obtiveram melhor desempenho para a aplicação. Além disso, se faz necessário explorar diferentes combinações de características, assim como analisar a influência da quantidade de medidas de consumo

consideradas para o treinamento dos modelos preditivos. As diferentes combinações de características foram sintetizadas em grupos para facilitar o entendimento e a rastreabilidade dos resultados. Nos capítulos seguintes, são discutidos com detalhes os resultados obtidos pelo protocolo, apresentando diversas análises e exemplos que atestam a eficiência do método de regressão extrínseca *multi-target*.

Capítulo 6

Avaliação Experimental

6.1 Considerações iniciais

Neste capítulo, são apresentados os resultados obtidos para o problema de predição de picos considerando dados de consumo de energia e as abordagens anteriormente discutidas. Os resultados consideram o desempenho de algoritmos convencionais ao abordar o problema de predição da magnitude e da posição do pico de maneira independente. A análise destes resultados é importante para estabelecer metas de desempenho a serem superadas com desenvolvimento deste projeto, utilizando algoritmos de regressão extrínseca *multi-target*. Os resultados apresentados neste capítulo representam apenas uma amostra do conjunto de consumidores selecionados para esta pesquisa, detalhados na Seção 4.3. No entanto, os resultados completos envolvendo todos os consumidores podem ser consultados no material complementar disponível online¹.

6.2 Descrição do protocolo experimental

A tarefa de predição abordada neste trabalho visa identificar a magnitude e a posição do valor máximo nas observações futuras de uma série temporal. No caso da aplicação explorada, o objetivo é identificar com uma semana de antecedência qual será a magnitude e o dia da semana do pico de consumo diário de energia de um cliente.

Após observações preliminares, verificou-se que a tarefa de predição da posição do pico obteve melhores resultados se tratada como uma tarefa de classificação, passando a ser considerada desta maneira, enquanto foi mantida a tarefa de predição da magnitude como uma tarefa de regressão. Desse modo, a configuração *multi-target* do problema considera uma variável numérica e outra categórica.

¹<https://sites.google.com/view/rodkruger-predicao-picos>

Por se tratar da predição de dois valores, uma abordagem simples para este problema é considerar duas tarefas de predição independentes, utilizando um modelo para prever a magnitude e outro para prever a posição do pico. Assim, em um primeiro momento, foi realizada uma avaliação considerando esta abordagem e diferentes algoritmos, rasos e profundos. No decorrer do documento, esta abordagem é referenciada como *single-target*.

Uma outra abordagem considerada foi avaliar métodos de regressão *multi-target*, que exploram a correlação entre as variáveis de saída para melhorar os resultados. Nesta abordagem foram avaliados algoritmos rasos e profundos como regressor-base em cada método *multi-target*, para avaliar o desempenho do método *multi-target* conforme o modelo do regressor-base.

Com relação ao conjunto de características, diferentes conjuntos foram avaliados, conforme anteriormente discutido na Seção 5.3 e ilustrado na Tabela 5.3. Cada combinação de características foi avaliada em todos os métodos de regressão explorados. Além disso, o intervalo de medidas de consumo considerado para extrair as características de cada combinação foi de 1 a 4 semanas, de modo que para cada combinação foi gerado uma matriz que considera 1, 2, 3 ou 4 semanas de medidas de consumo.

Na Tabela 6.1, são apresentados os algoritmos selecionados para os experimentos, assim como os parâmetros utilizados na melhor rodada. A escolha destes algoritmos foi baseada nas soluções propostas por trabalhos relacionados em diferentes domínios de aplicação.

6.2.1 Linhas de base

Além dos métodos baseados em aprendizagem de máquina, também foram definidos dois métodos de *baseline* para a comparação dos resultados. Os métodos se baseiam exclusivamente em informações do passado das séries temporais para realizar predições futuras, sem a necessidade da construção de modelos preditivos. A seguir, é apresentada uma descrição destes métodos:

- Baseline 1 (semana anterior): considera que o pico (magnitude e dia da semana) da semana seguinte será o mesmo observado na semana anterior. Desse modo, considera que os padrões de consumo dos consumidores se repetem em dias consecutivos.
- Baseline 2 (mês anterior): similar ao Baseline 1, mas considera que o pico da semana seguinte será exatamente o mesmo observado na semana correspondente do mês anterior. Assim, leva em consideração a repetição de um padrão de consumo em diferentes meses.

Abordagem	Tipo de modelo	Algoritmo	Parâmetros
Single-target	Raso	Linear Regression	
		SVM	kernel RBF C = 1.0 epsilon = 0.2
		XGBoost	learning rate = 0.3 max depth = 6
		ANN	hidden layer = (84, 42, 21, 10, 5) learning rate = 0.001 activation function = ReLU
		Voting	1 - ANN hidden layer = (84, 42, 21, 10, 5); learning rate = 0.001; activation function = ReLU 2 - XGBoost learning rate = 0.3; max depth = 6 3 - SVM kernel RBF; C = 1.0; epsilon = 0.2
Single-target	Profundo	LSTM	units = 64 fully connected layer = (48, 24, 12)
		biLSTM	units = 64 fully connected layer = (48, 24, 12)
		F-CNN	fully connected layer = (48, 24, 12)
		Inception	fully connected layer = (48, 24, 12)
Multi-target	n/a	ERC SST DSTARS	Regressores base utilizados - Linear Regression - SVM - XGBoost - ANN - Voting - LSTM - biLSTM - F-CNN - Inception

Tabela 6.1: Algoritmos utilizados nos experimentos e seus parâmetros.

6.2.2 Método de validação

Dentre os métodos de validação disponíveis na literatura, este trabalho escolheu o *holdout* temporal. Os conjuntos de dados foram separados em treino e teste, utilizando os dados observados em 2017 e 2018 para o treinamento dos modelos e os dados de 2019 na etapa de testes. Foram considerados os valores obtidos pelos modelos nos exemplos de teste de acordo com as medidas MAE, MSE, RMSE e MAPE para a comparação dos resultados de regressão e as medidas acurácia, precisão, recall e f1-score para a tarefa de classificação. Tais medidas foram previamente definidas na Seção 5.4.

6.3 Análise geral dos resultados

Encontrar a configuração que maximiza o desempenho dos regressores e classificadores na tarefa de predição de picos depende de diversos fatores, como a quantidade de semanas consideradas para a construção das matrizes de características, quais características são utilizadas, o algoritmo de regressão ou classificação, sendo ele de natureza *single-target*, baseado em *deep learning* ou um modelo *multi-target*. Por essa razão, a avaliação experimental para este problema considerou um protocolo amplo que avalia todas as combinações possíveis destas variáveis, visando encontrar a melhor configuração.

Nas seções seguintes, são apresentados maiores detalhes de cada uma destas variáveis, esclarecendo os critérios de avaliação adotados pelo protocolo experimental, além de detalhar as razões para escolher a melhor solução em cada uma das variáveis em avaliação.

6.3.1 Janela de semanas

Analisar a influência da quantidade de semanas no desempenho de predição é fundamental para definir um conjunto de características que maximiza o desempenho do modelo. Por essa razão, cada algoritmo foi avaliado em matrizes de características que consideram janelas de 1, 2, 3 e 4 semanas de consumo de cada um dos 100 clientes em avaliação por este trabalho.

Conforme explicado na Seção 4.3, as leituras de consumo ocorrem a cada 15 minutos. Ou seja, diariamente são obtidas 96 medidas de consumo, que totalizam 672 medidas semanalmente. Desse modo, os modelos de predição foram avaliados em matrizes de características construídas a partir de 672, 1344, 2016 e 2688 medidas de consumo, ocorridas no intervalo de 1 à 4 semanas.

A análise não considera a influência no desempenho de predição que diferentes características exerceram e se concentra exclusivamente em determinar o tamanho de janela que proporciona os melhores resultados. Na Seção 6.3.2 é discutida a influência de diferentes características no desempenho. Além disso, esta análise não separa a influência no desempenho de diferentes modelos-base utilizados em regressores *multi-target*. Esse aspecto é discutido na Seção 6.3.3.

Para ambas as variáveis-alvo e, de um modo geral, a menor janela de uma semana obteve o melhor desempenho, inclusive nos experimentos que empregaram algoritmos baseados em *deep learning*, que é esperado desempenho melhor com uma quantidade maior de dados, como por exemplo aqueles que possuem medidas de consumo de 4 semanas. Os gráficos das Figuras 6.1 e 6.2 ilustram as métricas para predição da magnitude e da posição, respectivamente.

Dentre os algoritmos avaliados para predição da magnitude, o Linear Regression foi o melhor algoritmo *single-target*, o biLSTM o melhor algoritmo de *deep learning*

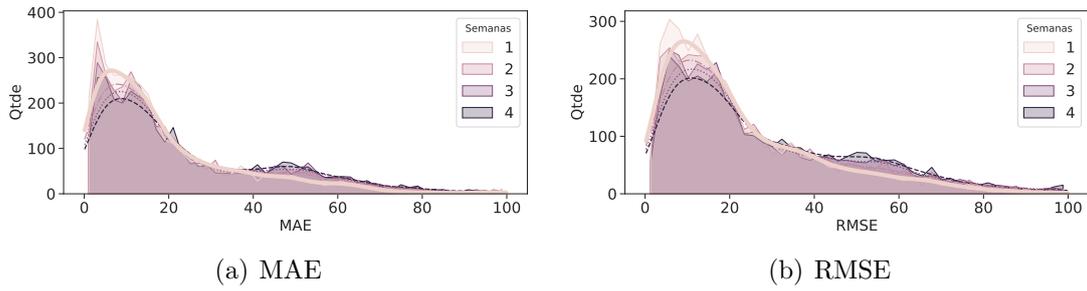


Figura 6.1: Distribuição das métricas MAE e RMSE conforme a janela de semanas, para predição da magnitude.

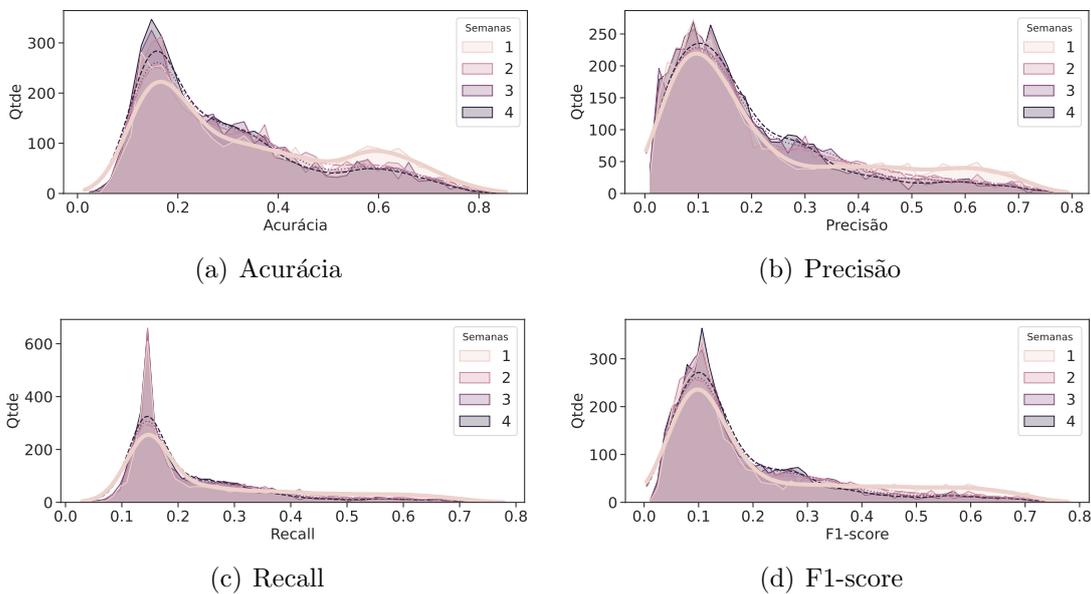


Figura 6.2: Distribuição das métricas acurácia, precisão, recall e f1-score conforme a janela de semanas, para predição da posição.

e o ERC o melhor algoritmo *multi-target*. A Figura 6.3 ilustra o erro médio absoluto obtido pela avaliação. Já os algoritmos avaliados para predição da posição, o XGBoost foi o melhor algoritmo *single-target*, o LSTM o melhor de *deep learning* e o SST o melhor *multi-target*. A Figura 6.4 ilustra a distribuição da acurácia obtida na avaliação.

A análise individual do desempenho de cada algoritmo também demonstrou que, na maioria dos casos, os melhores resultados foram obtidos quando os modelos de predição foram treinados com matrizes construídas a partir de medidas de consumo de 1 semana. Por esse motivo, os demais experimentos deste trabalho foram realizados utilizando somente a janela de 1 semana, para simplificar a análise dos resultados.

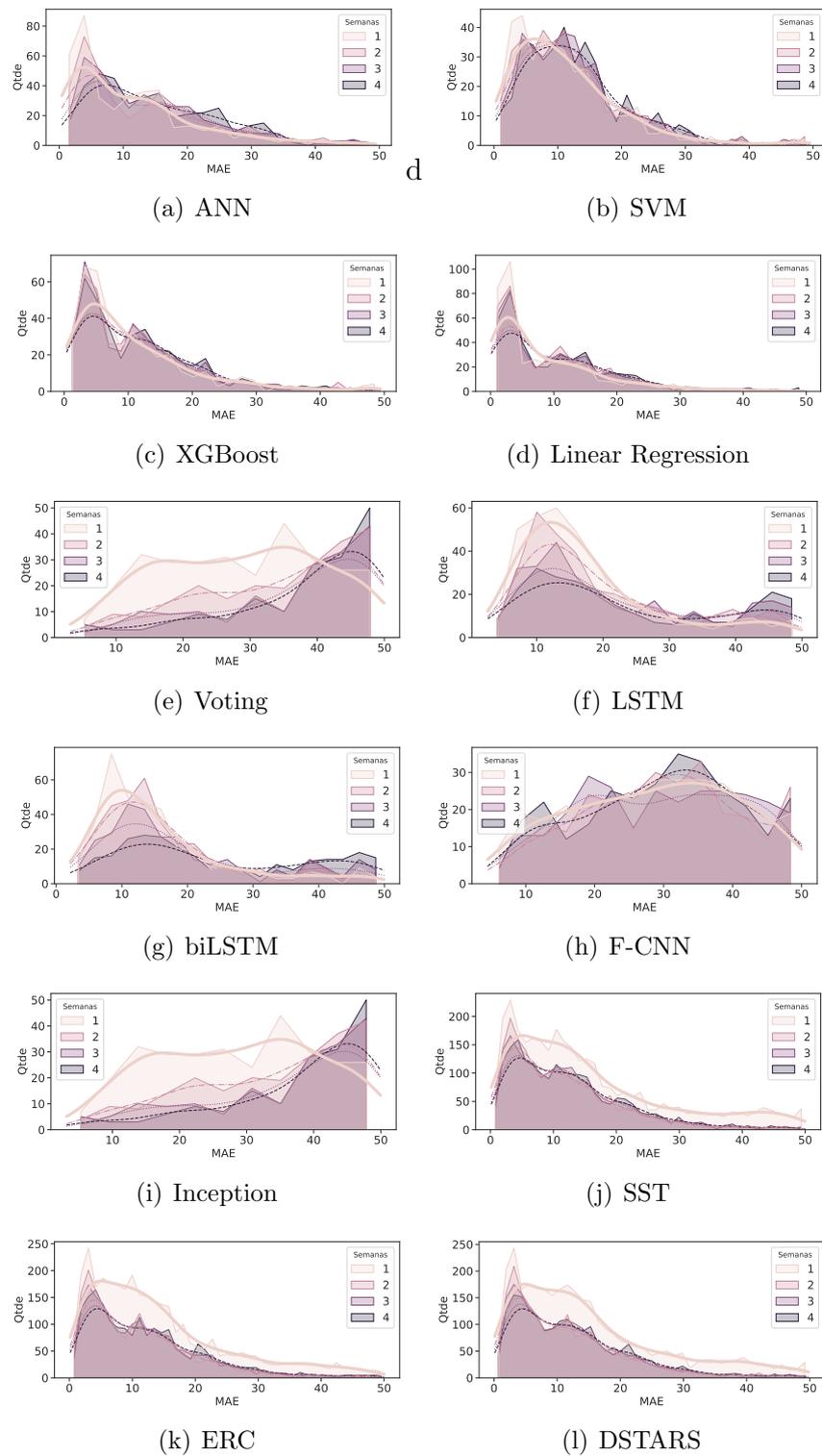


Figura 6.3: Distribuição do erro médio absoluto por algoritmo, conforme a janela de semanas, para predição da magnitude.

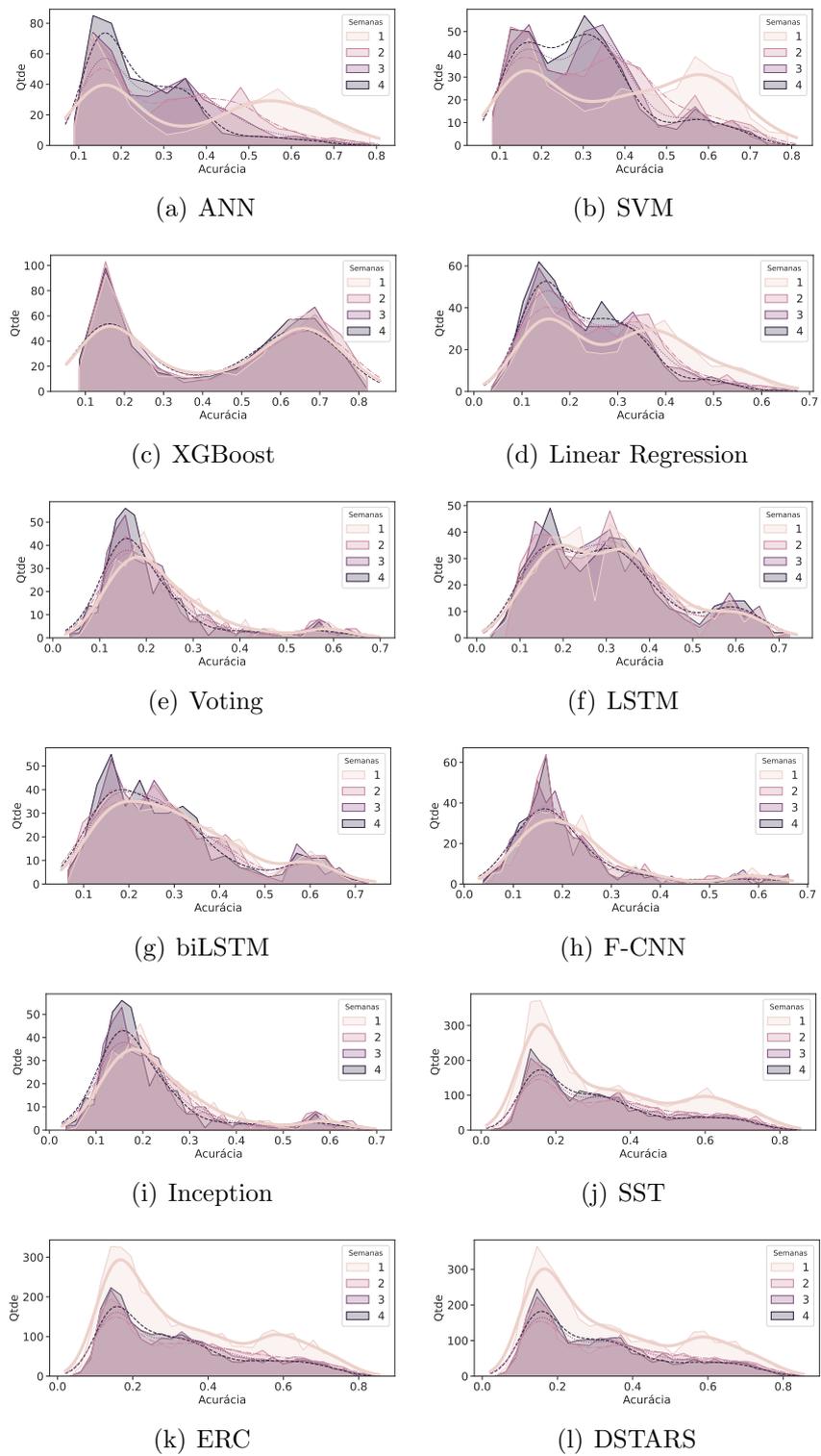


Figura 6.4: Distribuição da acurácia por algoritmo, conforme a janela de semanas, para predição da posição.

6.3.2 Matriz de características

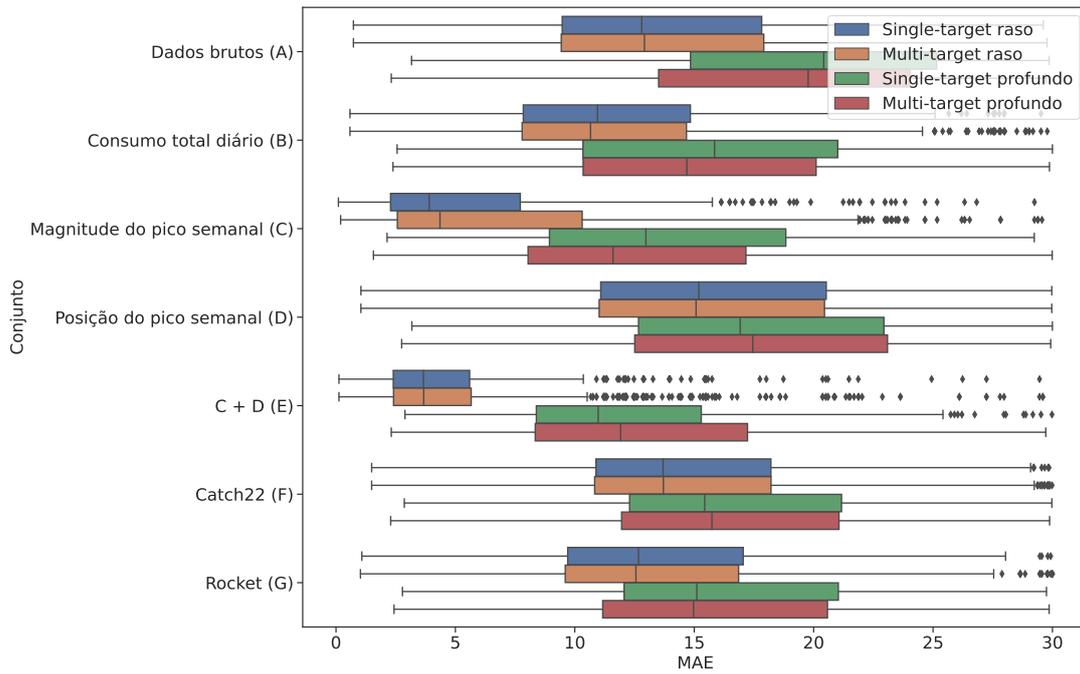
Diferentes combinações de conjuntos de características foram avaliadas no trabalho, conforme apresentado na Tabela 5.3. Da mesma maneira, conforme apresentado na Seção 6.3.1, a melhor janela de semanas avaliada foi de uma semana e, portanto, os conjuntos de características avaliados consideram valores de consumo de apenas uma semana de medições.

A análise não avalia a influência do algoritmo no desempenho de predição em cada conjunto de características avaliado. Entretanto, a análise dos resultados é conduzida considerando os resultados obtidos pelas quatro categorias de algoritmos explorados: *i*) single-target (raso), *ii*) single-target (profundo), *iii*) multi-target (raso) e *iv*) multi-target (profundo). Desse modo, os resultados apresentados representam a média obtida por diferentes algoritmos dentro de uma mesma categoria e considerando 100 consumidores. Uma avaliação detalhada que considera o desempenho de cada algoritmo individualmente é apresentada na Seção 6.3.3.

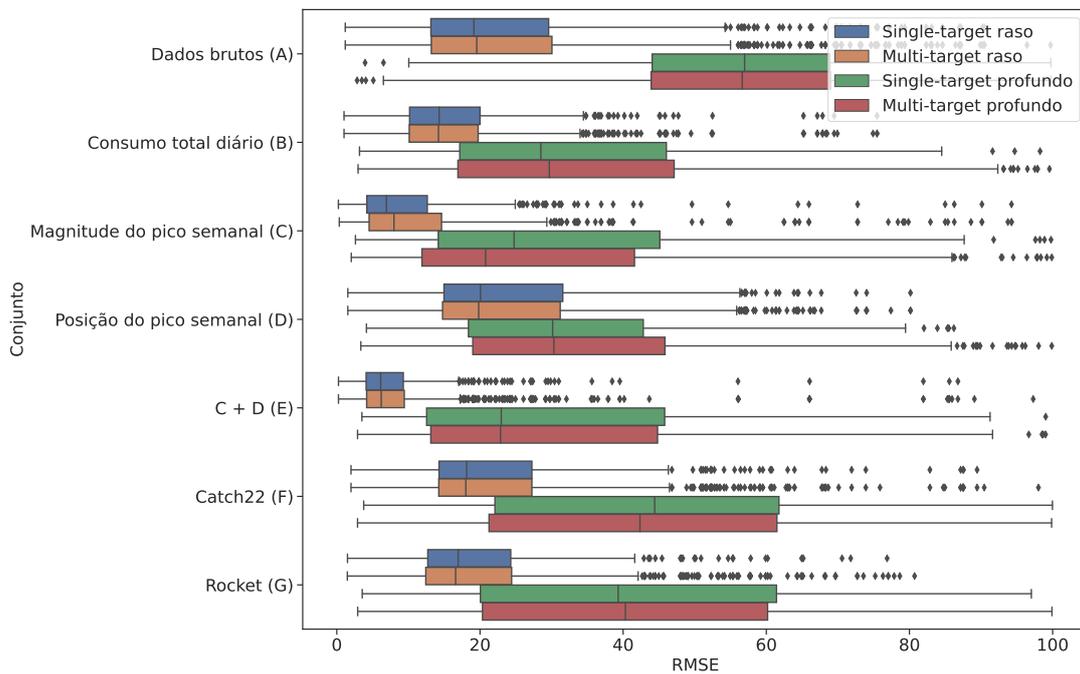
Os gráficos das Figuras 6.5 e 6.6 apresentam o erro médio absoluto e a acurácia da predição, respectivamente. De um modo geral, a predição da magnitude tem melhor desempenho quando se utilizam as características do conjunto do tipo E, que é formado pela magnitude do pico e a posição do pico da semana anterior. Já a predição da posição do pico tem melhor desempenho quando se utiliza o conjunto do tipo D, que utiliza somente a posição do pico da semana anterior.

A matriz do tipo A, construída unicamente a partir dos dados brutos das séries temporais, apresentou resultados inferiores para ambas as tarefas de predição (magnitude e posição), justificando a necessidade de construir matrizes a partir destes dados, buscando melhorar os resultados. Por este motivo, o trabalho explorou a redução de dimensionalidade da matriz do tipo A com o Catch2, gerando a matriz do tipo F, e o Rocket gerando a matriz do tipo G. Ambas as abordagens não geraram melhorias significativas nos resultados, produzindo resultados inferiores aos obtidos pelas matrizes do tipo E e D na predição da magnitude e posição, respectivamente.

Os resultados apresentam uma diferença significativa de desempenho entre o melhor caso e os demais, em cada uma das variáveis-alvo. Desse modo, utilizar o mesmo conjunto de características para prever as duas variáveis-alvo implicaria em uma perda de desempenho e, por essa razão, o trabalho utilizou conjuntos de características diferentes para cada variável-alvo, considerando a matriz do tipo E (magnitude do pico semanal + posição do pico semanal) para predição da magnitude e a matriz do tipo D (posição do pico semanal) para predição da posição. Por esta razão, as seções posteriores fornecem métricas de avaliação considerando estes tipos de matriz de melhor caso em cada uma das variáveis-alvo.

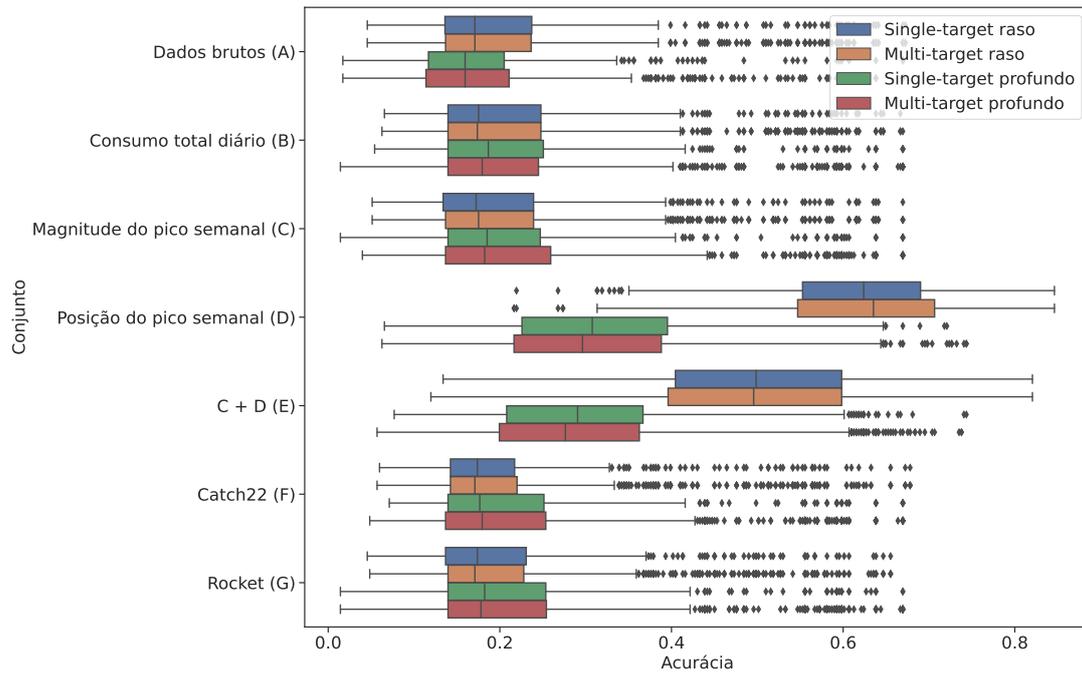


(a) MAE

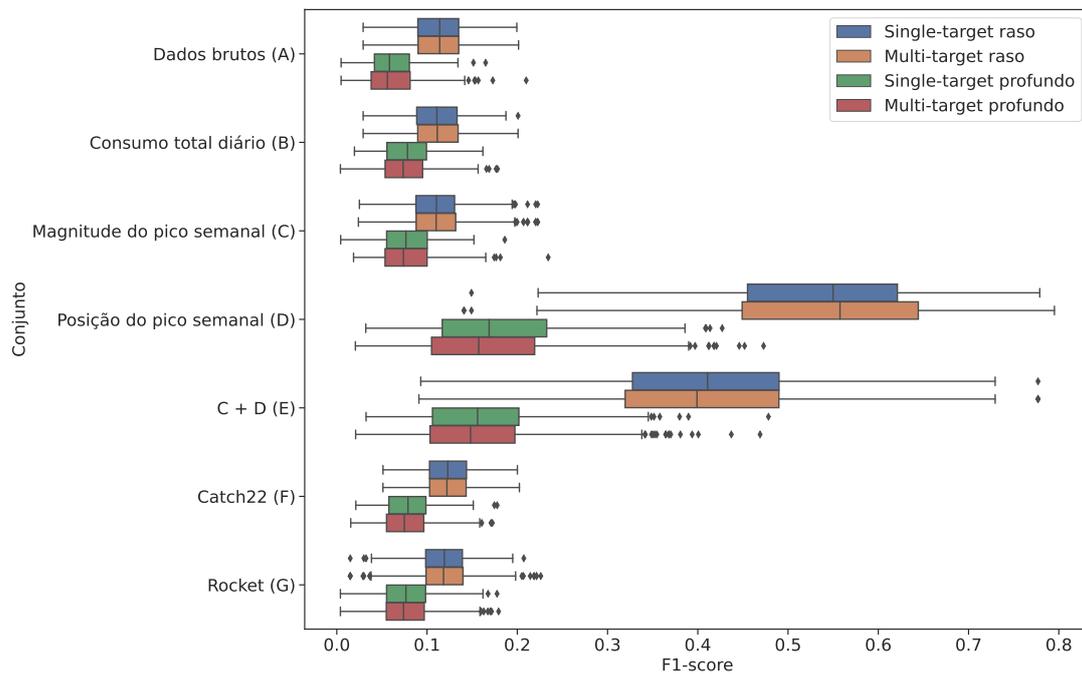


(b) RMSE

Figura 6.5: Distribuição das métricas MAE e RMSE conforme a matriz de características para predição da magnitude e considerando as diferentes abordagens de algoritmos de predição.



(a) Acurácia



(b) F1-score

Figura 6.6: Distribuição da acurácia e f1-score conforme a matriz de características para predição da posição e considerando as diferentes abordagens de algoritmos de predição.

6.3.3 Algoritmos

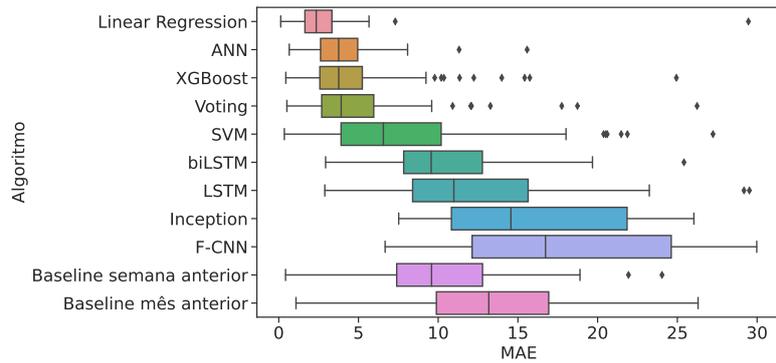
Conforme apresentado na Tabela 6.1, diferentes algoritmos foram avaliados, incluindo modelos rasos e profundos, considerando métodos *single-target* e *multi-target*. Uma das questões de pesquisa deste projeto é verificar se modelos *multi-target* melhoram o desempenho quando exploram a correlação existente entre as variáveis-alvo do problema. Da mesma forma que as análises anteriores, esta análise se restringiu a matrizes de características construídas com janelas de consumo de uma semana, mas, agora restrito ao tipo E (magnitude do pico semanal + posição do pico semanal) para predição da magnitude e ao tipo D (posição do pico semanal) para predição da posição, pelas razões explicadas na Seção 6.3.2.

É importante ressaltar que métodos *multi-target* efetuam a predição de múltiplas variáveis ao mesmo tempo e, desse modo, é coerente utilizar a mesma matriz de características para predição destas variáveis. Por esse motivo, os resultados apresentados para os métodos *multi-target* utilizam somente a matriz do tipo E (magnitude do pico semanal + posição do pico semanal) para predição da magnitude e posição, enquanto os resultados dos métodos *single-target* utilizam a matriz do tipo D (posição do pico semanal) para predição da posição.

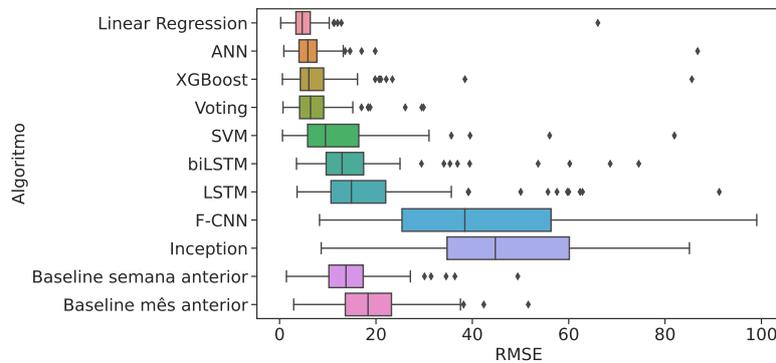
Os resultados dos algoritmos *single-target* estão ilustrados nas Figuras 6.7 e 6.8, para predição da magnitude e posição, respectivamente. Nessas figuras, as barras dos gráficos estão ordenadas pelo melhor resultado, de cima para baixo, com as duas últimas barras exibindo os resultados das linhas de base, definidos na Seção 6.2.1. Na predição da magnitude, o Linear Regression foi o algoritmo que obteve o melhor desempenho geral dentre todos algoritmos, enquanto o biLSTM foi o melhor entre os algoritmos profundos. Já na predição da posição, o Voting foi o algoritmo de melhor resultado entre os rasos e o biLSTM para os profundos.

Os resultados obtidos pela avaliação de métodos *multi-target* estão ilustrados nas Figuras 6.9 e 6.10, para magnitude e posição, respectivamente. Para facilitar a visualização do impacto de cada método no desempenho, a sequência das linhas é a mesma das análises feitas para os métodos *single-target*, ilustrados pelas Figuras 6.7 e 6.8. Na predição da magnitude, o Linear Regression combinado a métodos *multi-target* permaneceu tendo o melhor desempenho, sem apresentar grandes variações para sua versão *single-target*. Dentre os algoritmos profundos, o biLSTM permaneceu sendo a melhor opção, também sem apresentar grandes variações para a sua versão *single-target*. Já para predição da posição, o XGBoost combinado a métodos *multi-target* obteve os melhores resultados, embora não tenha superado os melhores resultados da versão *single-target*, obtidos pelo Voting e até mesmo da versão do XGBoost. Dentre os algoritmos profundos, o biLSTM se manteve como melhor opção.

De uma maneira geral, os métodos *multi-target* pouco influenciaram a predição da magnitude, tendo apresentado maiores variações nos resultados obtidos para predição da posição. Além disso, a maioria destas variações apontam para resultados



(a) MAE



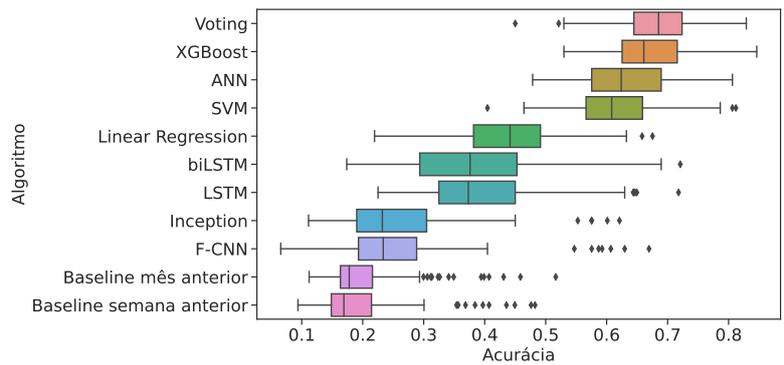
(b) RMSE

Figura 6.7: Distribuição das métricas MAE e RMSE conforme o algoritmo para predição da magnitude, utilizando métodos *single-target*.

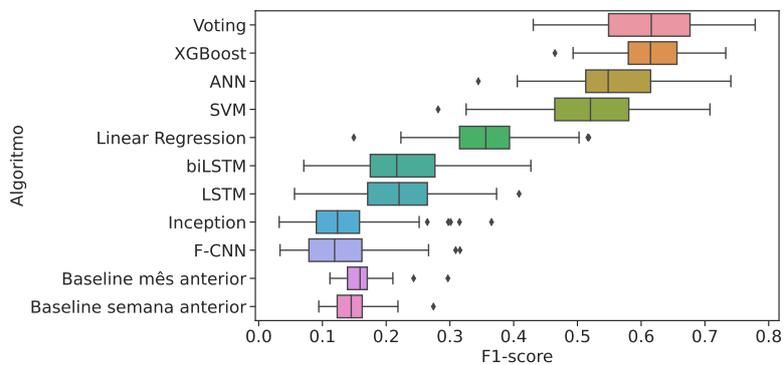
inferiores às versões *single-target*. Isto é explicado pelo fato dos métodos *multi-target* utilizarem a mesma matriz de características para predição de todas as variáveis e, como pode ser verificado pela Figura 6.6, a matriz do tipo E (magnitude do pico semanal + posição do pico semanal) foi apenas a segunda melhor em acurácia para predição da posição. Deve-se considerar também que o tipo do método *multi-target* também pouco influenciou, apresentando maiores variações quando combinados a algoritmos profundos, como foi o caso do F-CNN e o Inception que tiveram o desempenho melhorado quando combinados ao método ERC, embora não tenham superado suas versões *single-target*.

6.3.4 Melhor configuração

Até este ponto, foram discutidos resultados considerando diferentes configurações do problema, como o tamanho da janela para treinamento dos modelos e a combinação de diferentes características. Assim, é possível realizar uma análise detalhada considerando a melhor configuração para o problema.



(a) Acurácia

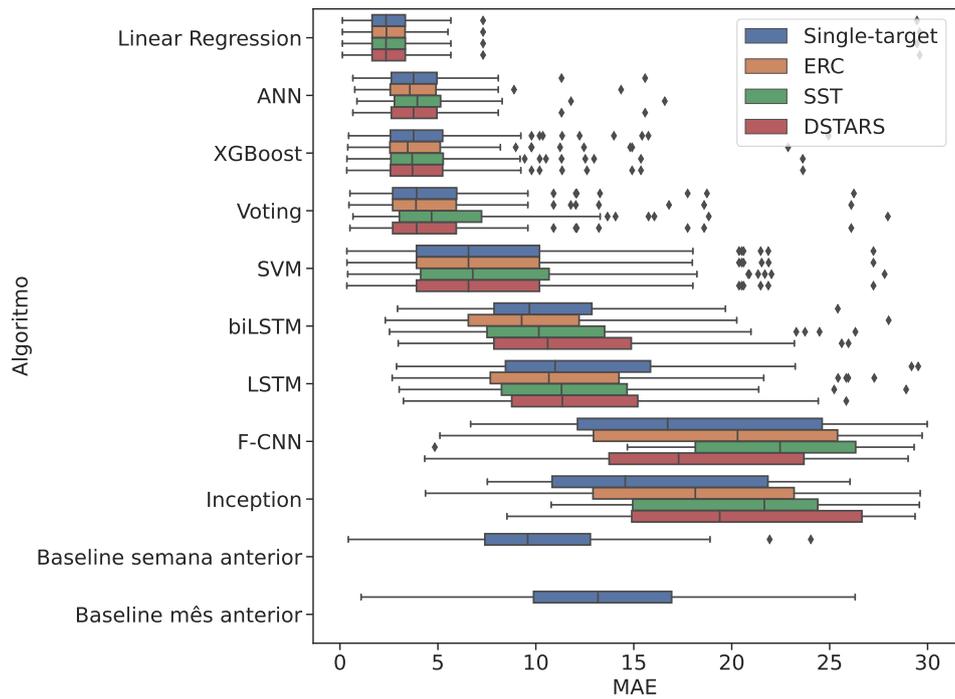


(b) F1-score

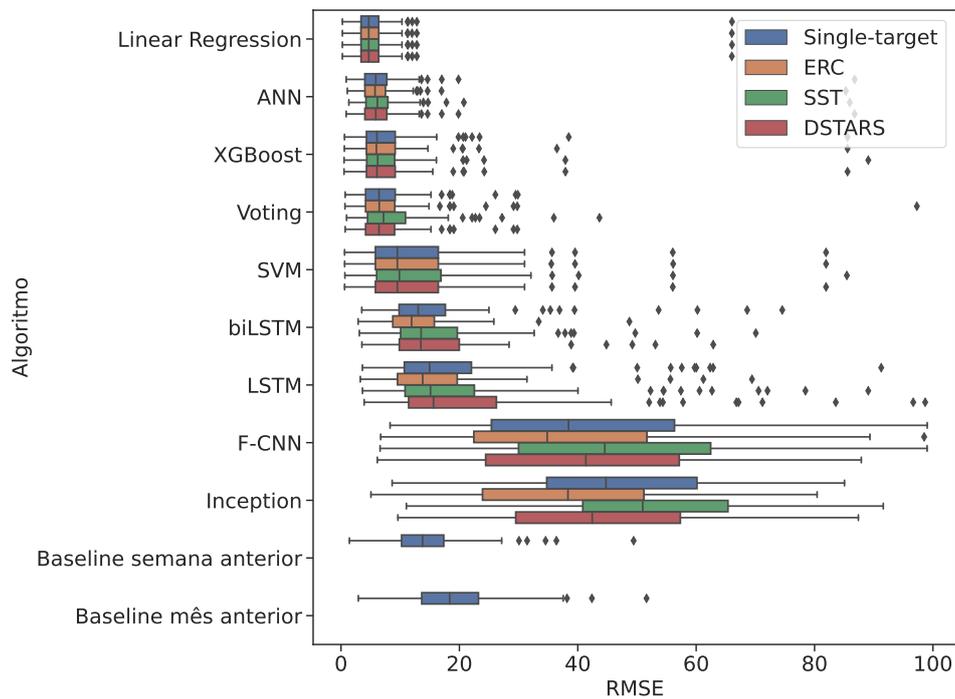
Figura 6.8: Distribuição da acurácia e f1-score conforme o algoritmo para predição da posição, utilizando métodos *single-target*.

Dentre as combinações de janelas de consumo, a janela de uma semana obteve os melhores resultados e foi a escolhida para detalhar as demais análises. Já o tipo de matriz de característica teve uma relação com a variável-alvo, sendo a matriz do tipo E (magnitude do pico semanal + posição do pico semanal) ideal para predição da magnitude e a matriz do tipo D (posição do pico semanal) ideal para predição da posição. Sendo assim, uma análise final, juntamente com o melhor algoritmo de cada categoria (*single-target* raso, *single-target* profundo e *multi-target*), é um critério suficiente para determinar o melhor modelo preditivo para cada variável-alvo da aplicação proposta por este trabalho.

Para a predição da magnitude, a matriz de características do tipo E (magnitude do pico semanal + posição do pico semanal) combinada com os algoritmos Linear Regression, biLSTM e SST utilizando Linear Regression como modelo base, formam uma configuração final de um modelo preditivo que entrega os melhores desempenhos em cada categoria de algoritmo. Na Figura 6.11, percebe-se um desempenho próximo do algoritmo Linear Regression com o método *multi-target* SST que utiliza o mesmo algoritmo *single-target* como regressor-base.

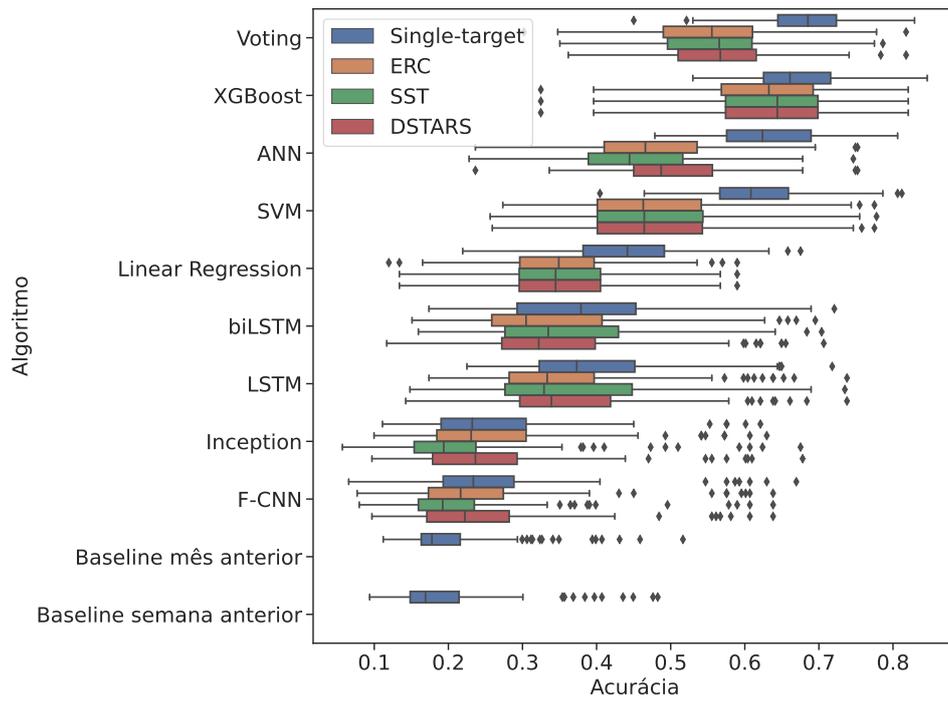


(a) MAE

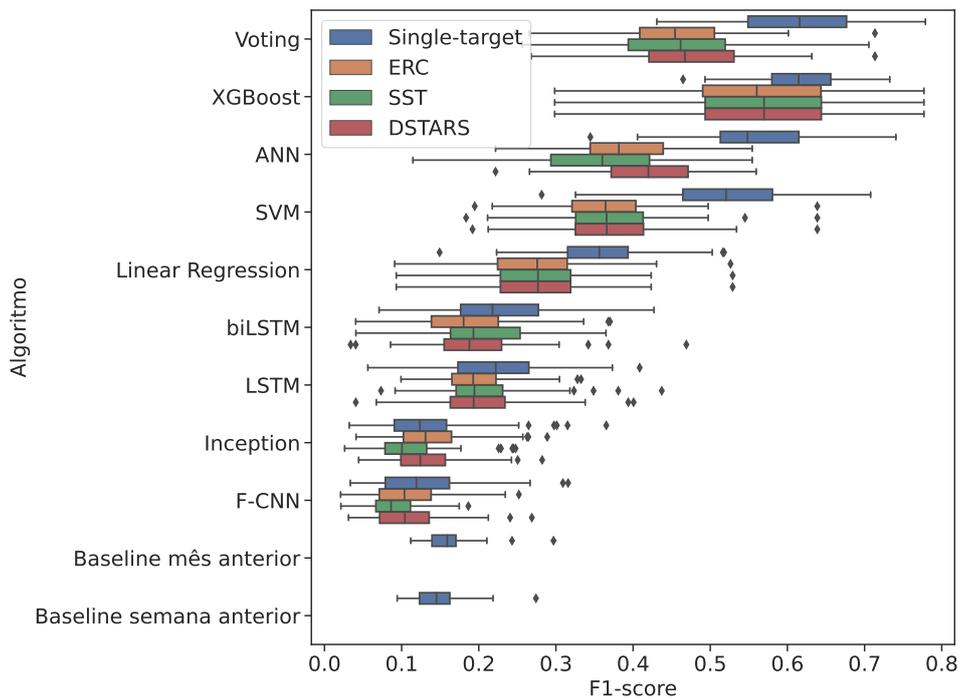


(b) RMSE

Figura 6.9: Distribuição das métricas MAE e RMSE conforme o algoritmo para predição da magnitude, utilizando métodos *multi-target*.



(a) Acurácia



(b) F1-score

Figura 6.10: Distribuição da acurácia e f1-score conforme o algoritmo para predição da posição, utilizando métodos *multi-target*.

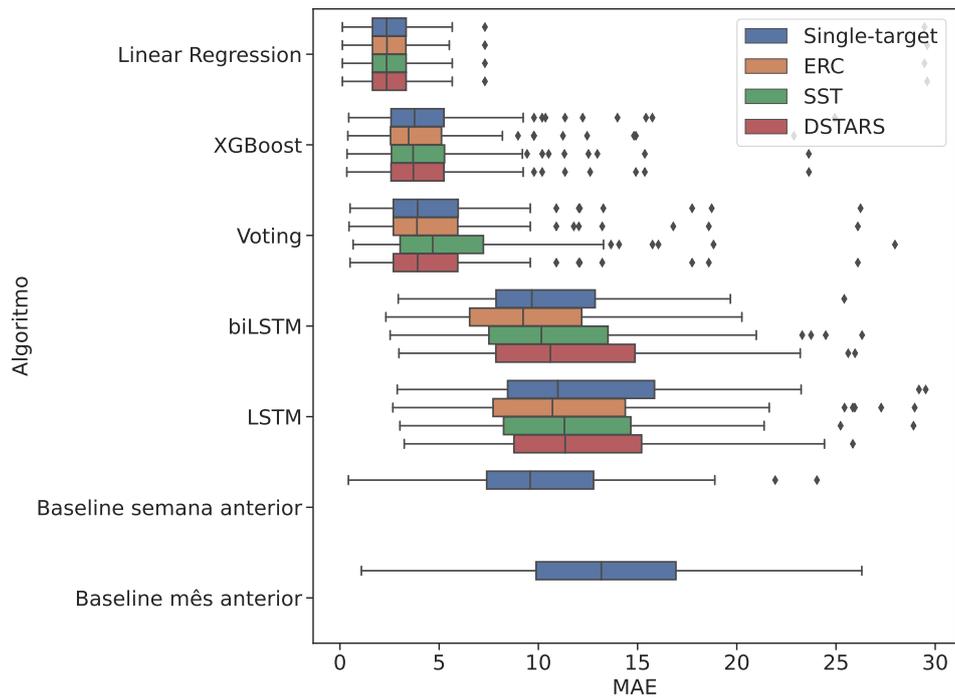
Para confirmar a escolha do melhor algoritmo, foi aplicado o teste não-paramétrico de Wilcoxon (Wilcoxon et al., 1970), com nível de confiança igual a 0.05 e pós teste de Holm com correção alpha ($\alpha = 5\%$) (Benavoli et al., 2016) sobre os erros médios absolutos obtidos pelo experimento, calculados para os 100 consumidores deste trabalho. A partir deste teste, foi gerado um gráfico de diferença crítica, ilustrado na Figura 6.12. É possível verificar que as diferenças encontradas entre as melhores opções não são estatisticamente significativas. Desse modo, pode-se afirmar que o modelo preditivo de melhor desempenho para predição da magnitude é o Linear Regression raso, ou a combinação dele com métodos *multi-target*, utilizando a matriz de característica do tipo E (magnitude do pico semanal + posição do pico semanal) construída a partir de medidas de 1 semana.

Já para a predição da posição, o Voting, o biLSTM e o DSTARS utilizando o XGBoost como modelo-base, aliado a matriz de características do tipo D (posição do pico semanal), configuraram o melhor arranjo do modelo preditivo. Os resultados dos principais algoritmos desta categoria estão ilustrados na Figura 6.13. Da mesma forma que para magnitude, o teste de Wilcoxon foi executado, com nível de confiança de 0.05, sobre a acurácia obtida na construção dos modelos de regressão para 100 consumidores. Não houve diferenças de desempenho estatisticamente significativas para predição da posição, conforme apresentado na Figura 6.14. Desta maneira, pode-se concluir que a melhor configuração de predição de posição utiliza o Voting raso, seguido pelo XGBoost raso combinado a algum método *multi-target*, aliado a matriz de característica do tipo D (posição do pico semanal) construída a partir de medidas de 1 semana.

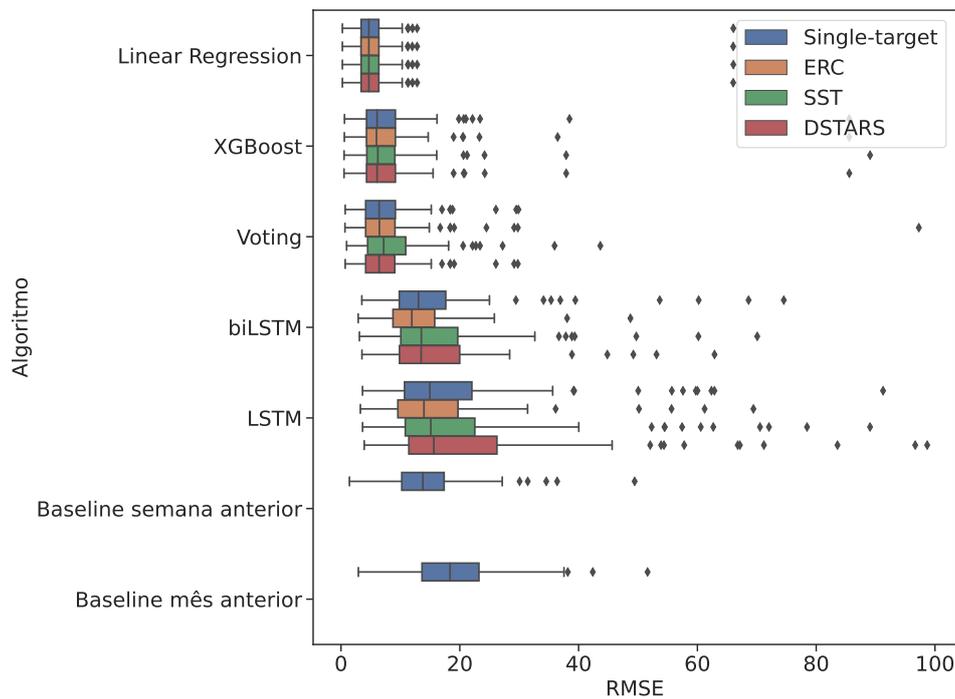
6.4 Modelos individuais vs. modelo global

Os resultados analisados até esta seção consideram que os modelos de predição são construídos a partir de dados de consumo individuais por cliente. Esta abordagem tem a vantagem de identificar melhor o padrão de consumo de cada cliente, porém, também possui as seguintes desvantagens: *i*) dado um novo cliente, a companhia de energia necessita de alguns anos até possuir dados históricos que permitam treinar um modelo de predição deste cliente, e *ii*) nos primeiros anos, o volume de dados existentes para um novo cliente é muito pequeno para treinar modelos de *deep learning*.

Buscando encontrar alternativas para superar estas desvantagens, o trabalho analisou o desempenho de modelos de predição construídos a partir de dados de consumo de todos os clientes, para cada um dos algoritmos em avaliação e em cada método de regressão *multi-target*. O experimento considerou a matriz de característica de melhor desempenho para magnitude e posição, sendo respectivamente o tipo E (magnitude do pico semanal + posição do pico semanal) e tipo D (posição do pico semanal), construídas a partir de dados de consumo de 1 semana.



(a) MAE



(b) RMSE

Figura 6.11: Distribuição das métricas MAE e RMSE dos melhores algoritmos para predição da magnitude.

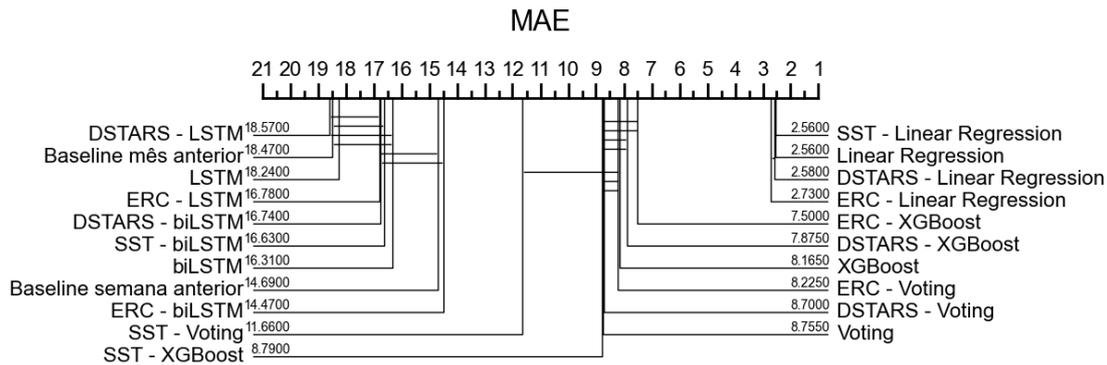


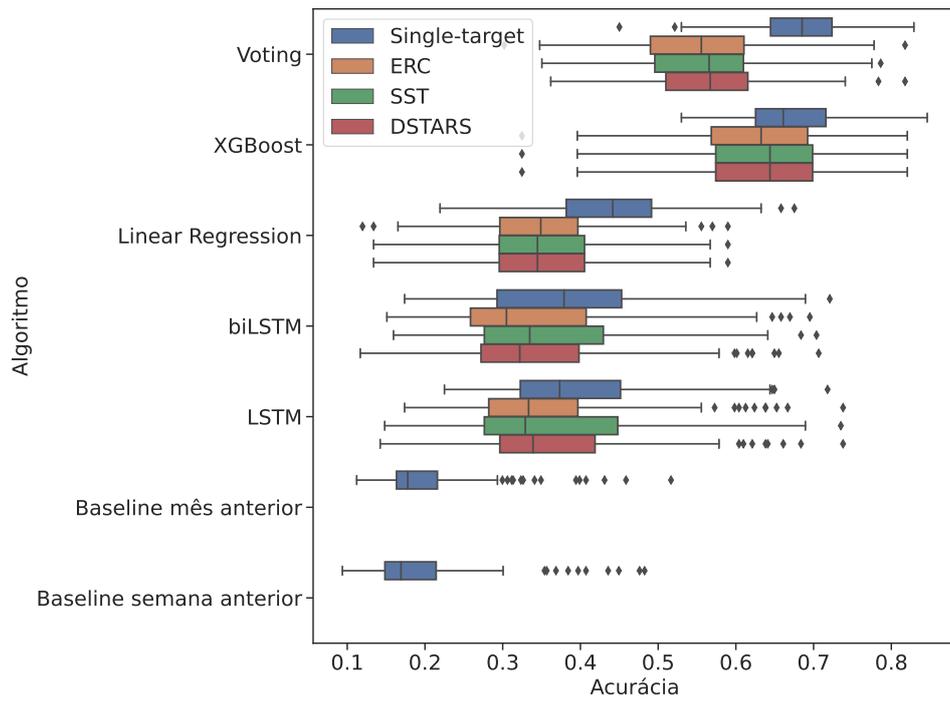
Figura 6.12: Diferença crítica dos melhores algoritmos para predição da magnitude.

A Figura 6.15 ilustra erro médio absoluto de cada algoritmo, comparando as versões de modelos construídos a partir de dados de um único consumidor com modelos construídos com todos os consumidores. Para a maioria dos casos, o desempenho de predição da magnitude reduziu com a introdução de um modelo global, exceto para os algoritmos Inception e F-CNN que, por serem baseados em *deep learning*, costumam ter melhor desempenho em conjuntos de dados maiores, que é o caso de um conjunto que engloba todos os clientes. Porém, o LSTM e o biLSTM não tiveram melhoras de desempenho no modelo global. Além disto, o Linear Regression permaneceu tendo o melhor desempenho entre todos os algoritmos, embora também tenha tido uma redução de desempenho se comparado com sua versão construída a partir de dados de consumos individuais para cada cliente.

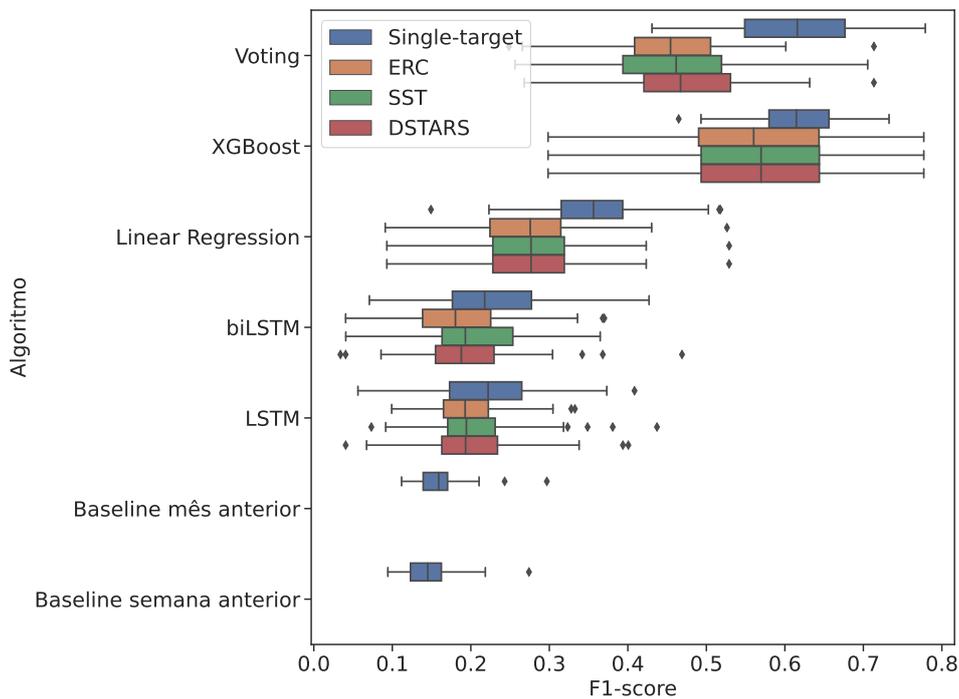
Na predição da posição, os modelos globais obtiveram desempenho inferior nos métodos *single-target* que utilizam algoritmos rasos, mas superior na maioria dos algoritmos de *deep learning* e métodos *multi-target*, indicando que para esta variável-alvo o arranjo de um modelo global favorece o desempenho. Ao mesmo tempo, também se nota uma distribuição mais irregular dos resultados, embora a mediana de desempenho *multi-target* seja superior. A Figura 6.16 ilustra resultados obtidos.

6.5 Análise individual dos resultados

Anteriormente, foram conduzidas análises que consideram os resultados gerais obtidos por cada uma das abordagens em um conjunto de 100 consumidores. Visando uma análise mais detalhada destes resultados, essa seção apresenta uma análise que considera os resultados obtidos pelos modelos de predição de maneira individual para diferentes consumidores. Devido ao alto número de consumidores que foram avaliados neste trabalho (100 consumidores), optou-se por conduzir essa análise em um subconjunto de 10 consumidores que apresentaram os melhores resultados.



(a) Acurácia



(b) F1-score

Figura 6.13: Distribuição da acurácia e f1-score dos melhores algoritmos para predição da posição.

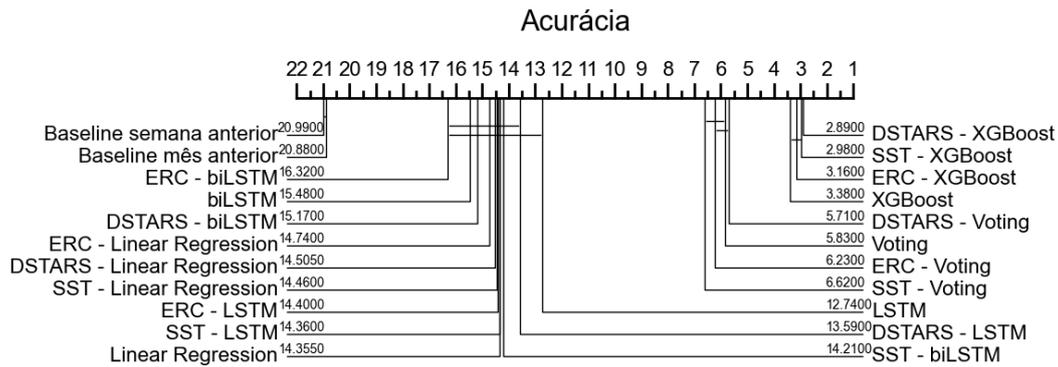


Figura 6.14: Diferença crítica dos melhores algoritmos avaliados para predição da posição.

Com o objetivo de comparar as diferentes abordagens exploradas neste projeto, as análises individuais levam em consideração os algoritmos que apresentaram os melhores resultados em cada uma das três abordagens avaliadas anteriormente. Para magnitude, são detalhados as métricas do Linear Regression, Voting, biLSTM e SST, com a matriz de características do tipo E (magnitude do pico semanal + posição do pico semanal). Já para posição, são detalhadas as métricas do XGBoost, Voting, biLSTM e SST, com a matriz de características do tipo D (posição do pico semanal). Além disso, considerou-se janelas de treinamento com uma semana de consumo.

6.5.1 Predição da magnitude do pico

Considerando a tarefa de predição da magnitude do pico, os resultados obtidos pelo diferentes regressores de acordo com as medidas MAE, MSE, RMSE e MAPE são apresentados nas Tabelas 6.2, 6.3, 6.4 e 6.5, respectivamente. Nestas tabelas, a primeira coluna representa o identificador (ID) de cada consumidor avaliado. Além disso, o melhor resultado para cada consumidor é destacado em negrito. Por fim, as últimas linhas indicam a média e desvio-padrão de cada método de regressão considerando os 10 consumidores.

Ao considerar o erro absoluto médio para cada consumidor, pode-se notar que o método do Linear Regression frequentemente obtém o melhor desempenho dentre os algoritmos em avaliação. É importante salientar que há uma diferença insignificante entre o SST e o Linear Regression *single-target*, denotando que os conjuntos de dados de cada consumidor não possuem correlação entre as variáveis-alvo suficiente para proporcionar um ganho na predição nos modelos *multi-target*.

Para melhor compreensão dos resultados, a Figura 6.17 apresenta gráficos de dispersão que ilustram as predições do algoritmo Linear Regression para os consumidores 40488, 18870, 27994 e 27439. Neste gráfico, cada ponto indica uma predição,

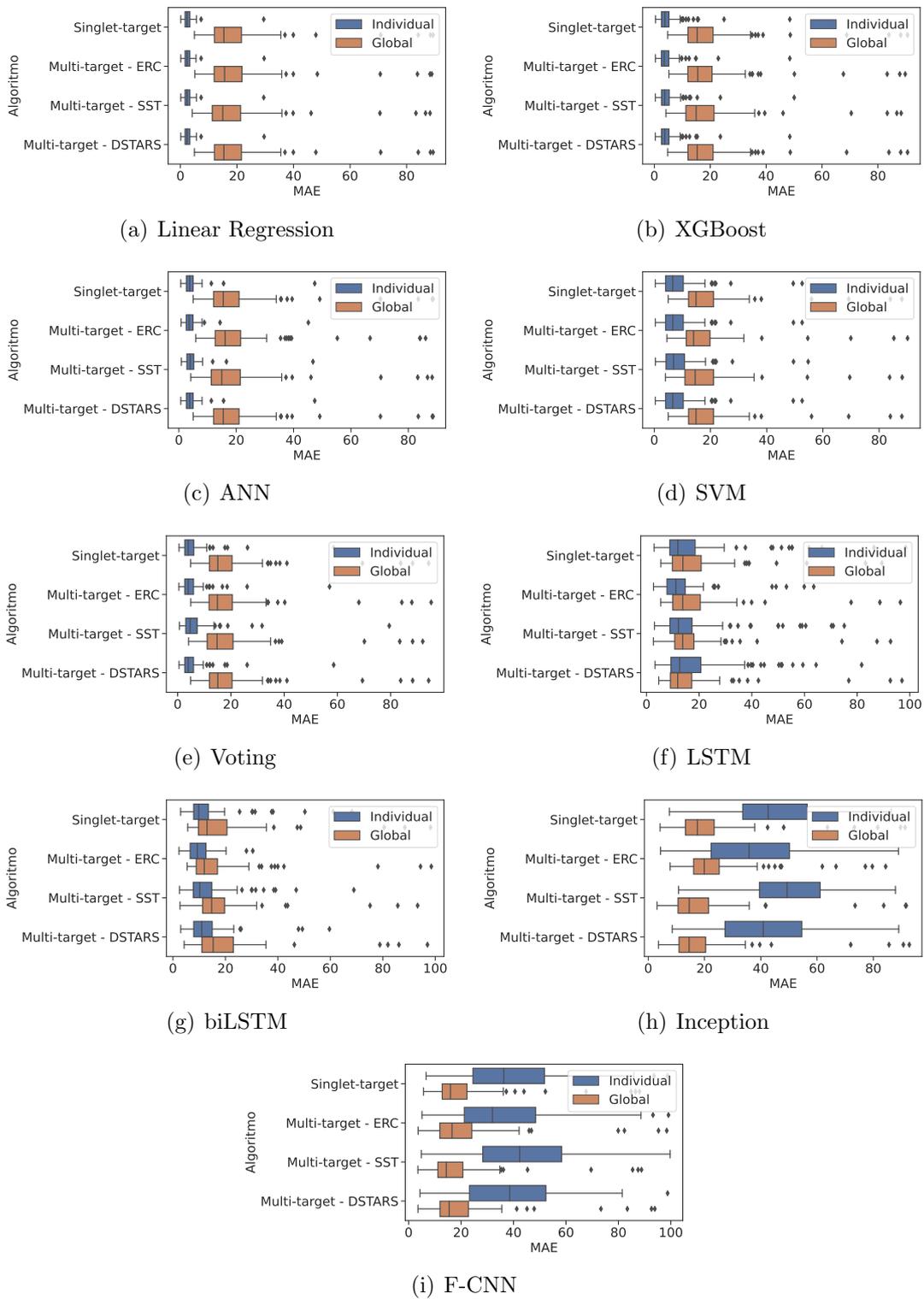


Figura 6.15: Distribuição do MAE dos algoritmos comparados a sua versão de treinamento a partir de dados de todos os consumidores.

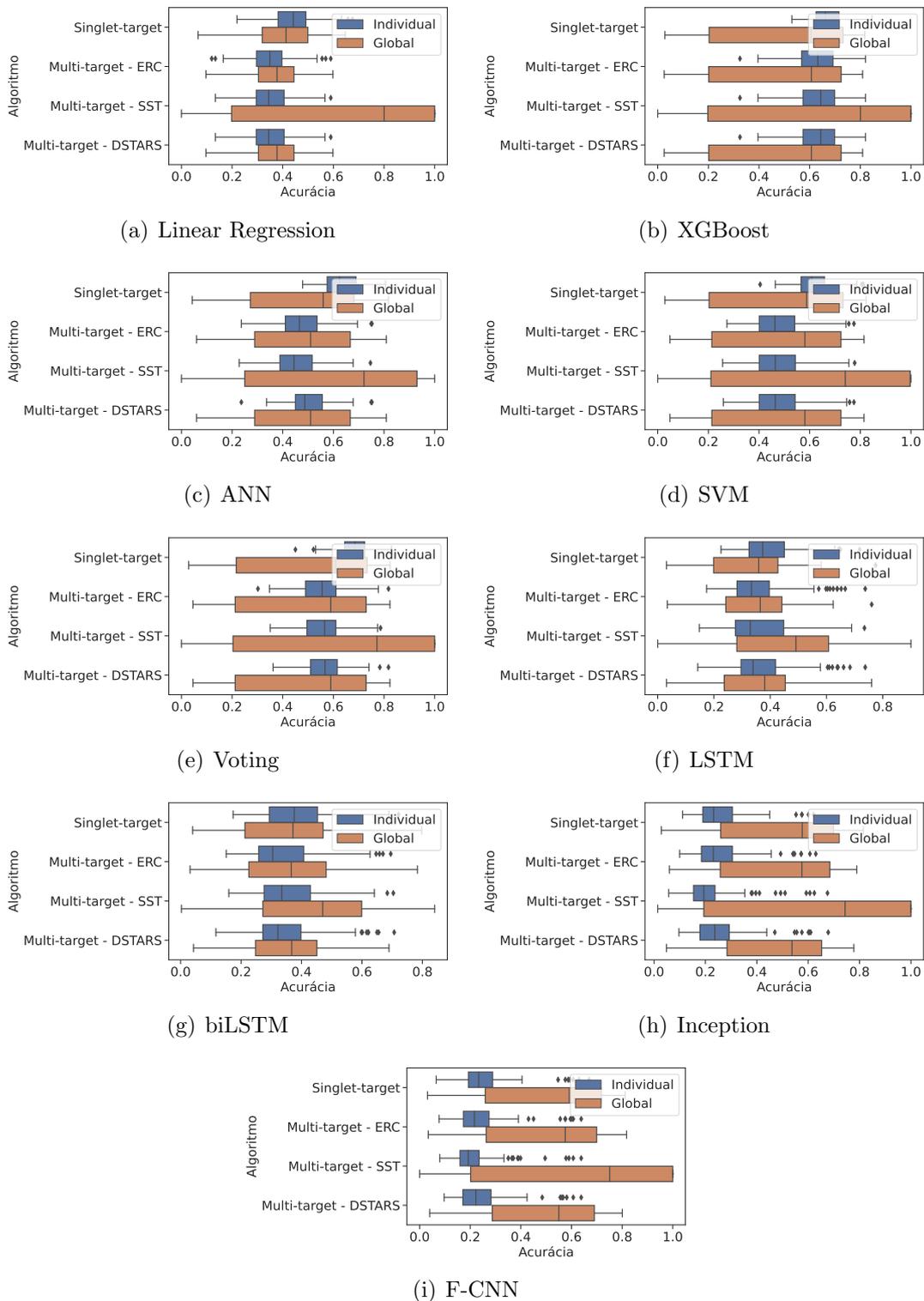


Figura 6.16: Distribuição da acurácia dos algoritmos comparados a sua versão de treinamento a partir de dados de todos os consumidores.

ID	Bsl. 1 (sem.)	Bsl. 2 (mês)	Linear Reg.	Voting	biLSTM	SST
40488	0,427	1,081	0,123	0,510	7,870	0,123
18870	1,158	2,637	0,207	2,717	7,744	0,207
27994	1,263	3,571	0,318	0,598	4,603	0,318
27439	4,267	4,636	0,903	0,945	6,559	0,903
27966	3,677	4,776	0,977	1,870	9,670	0,977
7212	4,625	5,743	1,046	1,386	2,939	1,046
14312	5,095	7,925	1,119	1,668	7,231	1,119
28961	4,432	5,172	1,144	1,434	7,579	1,144
20985	5,499	8,878	1,165	1,579	8,149	1,165
2220	5,713	9,338	1,174	2,471	6,742	1,174
Média	3,616	5,376	0,818	1,518	6,909	0,818
Desvio padrão	1,844	2,545	0,404	0,686	1,805	0,404

Tabela 6.2: Resultados de MAE para a tarefa de predição da magnitude do pico de consumo de energia elétrica.

ID	Bsl. 1 (sem.)	Bsl. 2 (mês)	Linear Reg.	Voting	biLSTM	SST
40488	2,017	8,577	0,055	0,503	64,555	0,055
18870	5,280	39,437	0,128	12,459	68,613	0,128
27994	11,991	45,068	2,521	3,128	43,361	2,521
27439	40,798	46,246	2,336	2,355	54,6	2,336
27966	30,084	49,554	2,475	9,193	101,372	2,475
7212	39,395	69,747	3,069	3,848	12,297	3,069
14312	60,418	135,443	5,511	7,081	70,638	5,511
28961	33,960	47,869	5,317	5,785	71,722	5,317
20985	50,674	129,640	5,990	6,433	86,480	5,99
2220	81,784	208,462	4,097	9,997	58,153	4,097
Média	35,64	78,004	3,15	6,078	63,179	3,15
Desvio padrão	23,781	57,595	1,987	3,544	22,896	1,987

Tabela 6.3: Resultados de MSE para a tarefa de predição da magnitude do pico de consumo de energia elétrica.

ID	Bsl. 1 (sem.)	Bsl. 2 (mês)	Linear Reg.	Voting	biLSTM	SST
40488	1,420	2,929	0,234	0,709	8,035	0,234
18870	2,298	6,280	0,357	3,530	8,283	0,357
27994	3,463	6,713	1,588	1,768	6,585	1,588
27439	6,387	6,800	1,528	1,535	7,389	1,528
27966	5,485	7,039	1,573	3,032	10,068	1,573
7212	6,277	8,351	1,752	1,962	3,507	1,752
14312	7,773	11,638	2,347	2,661	8,405	2,347
28961	5,828	6,919	2,306	2,405	8,469	2,306
20985	7,119	11,386	2,447	2,536	9,299	2,447
2220	9,043	14,438	2,024	3,162	7,626	2,024
Média	5,509	8,249	1,616	2,33	7,767	1,616
Desvio padrão	2,300	3,155	0,734	0,806	1,691	0,734

Tabela 6.4: Resultados de RMSE para a tarefa de predição da magnitude do pico de consumo de energia elétrica.

ID	Bsl. 1 (sem.)	Bsl. 2 (mês)	Linear Reg.	Voting	biLSTM	SST
40488	0,010	0,025	0,003	0,011	0,167	0,003
18870	0,019	0,038	0,002	0,028	0,081	0,002
27994	0,021	0,061	0,005	0,010	0,075	0,005
27439	0,091	0,098	0,021	0,022	0,145	0,021
27966	0,057	0,074	0,015	0,027	0,145	0,015
7212	0,10	0,122	0,026	0,035	0,072	0,0260
14312	0,121	0,185	0,023	0,032	0,136	0,023
28961	0,097	0,114	0,024	0,029	0,155	0,024
20985	0,094	0,155	0,019	0,025	0,127	0,019
2220	0,121	0,197	0,037	0,084	0,229	0,037
Média	0,073	0,107	0,017	0,03	0,133	0,017
Desvio padrão	0,041	0,056	0,011	0,02	0,046	0,011

Tabela 6.5: Resultados de MAPE para a tarefa de predição da magnitude do pico de consumo de energia elétrica.

em que a sua localização no eixo horizontal indica o valor esperado e a localização no eixo vertical indica o valor predito. Assim, pontos localizados na diagonal principal do gráfico indicam predições livres de erros. Nestes gráficos, também é apresentado o coeficiente de determinação (R^2score).

Para ilustrar a eficiência dos modelos de regressão construídos, os gráficos da Figura 6.18 apresentam as séries de picos de magnitude reais e preditas pelos modelos construídos pelo algoritmo Linear Regression, juntamente com o coeficiente de determinação (R^2score). É possível perceber uma grande proximidade da série predita e um alto valor de R^2score , indicando um modelo de regressão de desempenho muito bom.

6.5.2 Predição da posição do pico

Os resultados obtidos por diferentes classificadores e métodos de *baseline* para a tarefa de predição da posição do pico, considerando as medidas acurácia, precisão, recall e f1-score, são apresentados nas Tabelas 6.6, 6.7, 6.8 e 6.9, respectivamente.

Em termos de acurácia, o algoritmo Voting foi o que apresentou os melhores resultados. Porém, da mesma maneira que para magnitude, não houveram diferenças significativas de desempenho para o XGBoost, combinado ou não a métodos *multi-target*, indicando que há falta de correlação entre as variáveis-alvo do problema.

A Figura 6.19 apresenta matrizes de confusão que ilustram as predições do algoritmo XGBoost para os consumidores 1318, 12636, 27975 e 43355. A partir disto, é possível perceber que há um grande número de amostras na classe Domingo, indicando um problema desbalanceado. Porém, o modelo de classificação construído lidou bem com esta característica e apresentou resultados satisfatórios em todas as classes, considerando as medidas calculadas pelo f1-score de cada consumidor, apresentados na tabela 6.9.

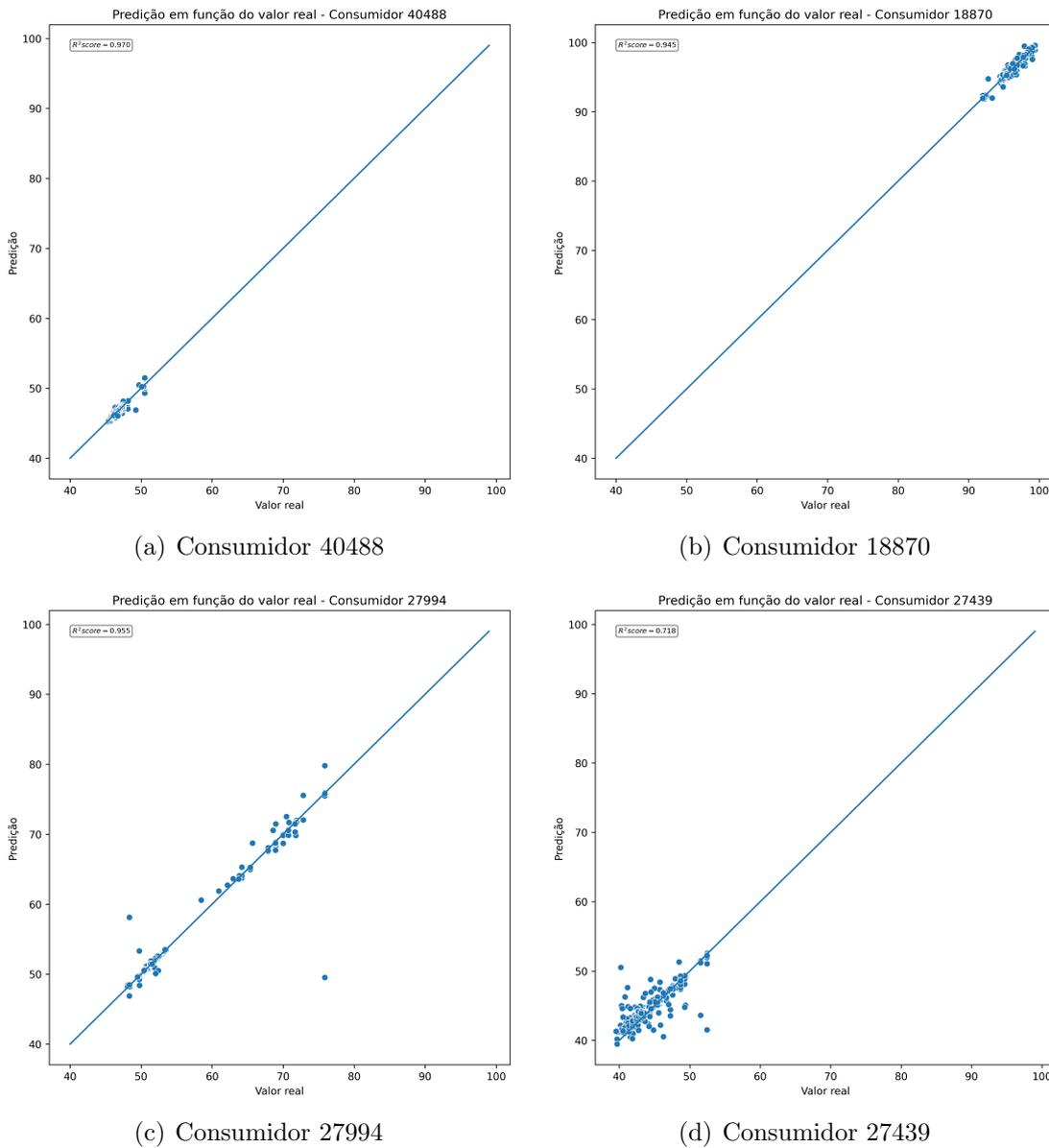
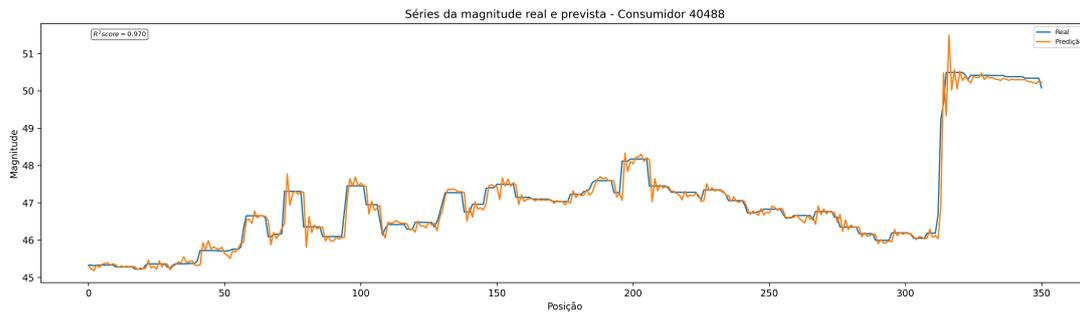
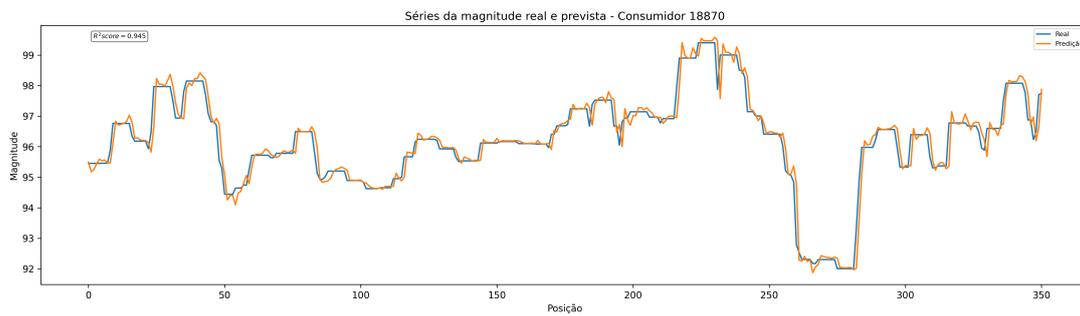


Figura 6.17: Gráficos de dispersão das previsões da magnitude do pico do algoritmo Linear Regression para quatro consumidores.

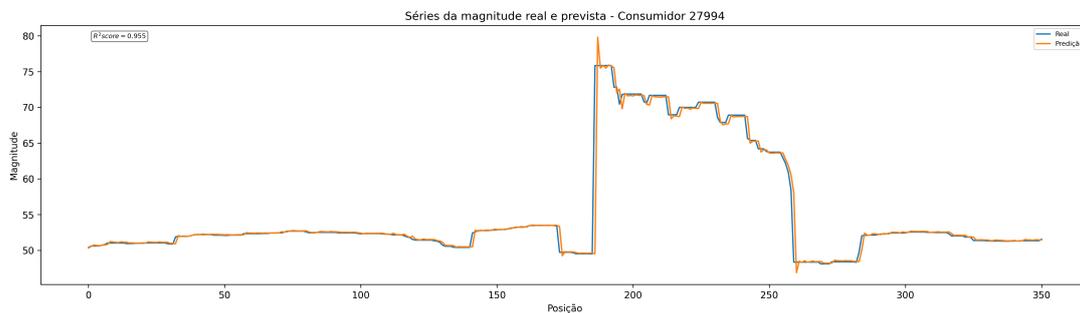
A eficiência dos modelos de classificação construídos estão ilustrados na Figura 6.20. Da mesma maneira que para magnitude, é possível perceber uma série de previsões bastante próxima a série verdadeira, juntamente com um elevado valor de R^2 score, também indicando um modelo de classificação de bom desempenho.



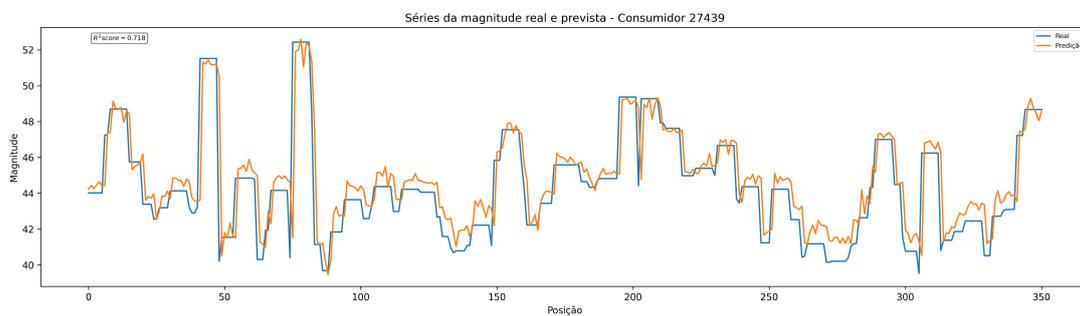
(a) Consumidor 40488



(b) Consumidor 18870



(c) Consumidor 27994



(d) Consumidor 27439

Figura 6.18: Séries temporais das previsões da magnitude do pico do algoritmo Linear Regression para quatro consumidores.

ID	Bsl. 1 (sem.)	Bsl. 2 (mês)	XGBoost	Voting	biLSTM	SST
1318	0,354	0,341	0,846	0,803	0,593	0,821
12636	0,476	0,517	0,826	0,829	0,721	0,815
27975	0,384	0,399	0,823	0,826	0,624	0,678
43355	0,357	0,349	0,775	0,766	0,587	0,746
28310	0,211	0,224	0,786	0,758	0,496	0,764
33879	0,186	0,169	0,752	0,806	0,402	0,681
31543	0,222	0,215	0,761	0,778	0,553	0,761
27431	0,397	0,394	0,769	0,775	0,689	0,772
1209	0,407	0,312	0,789	0,775	0,607	0,752
48264	0,293	0,313	0,766	0,724	0,575	0,704
Média	0,329	0,323	0,789	0,784	0,585	0,749
Desvio padrão	0,092	0,097	0,03	0,031	0,086	0,047

Tabela 6.6: Resultados de Acurácia para a tarefa de predição da posição do pico de consumo de energia elétrica.

ID	Bsl. 1 (sem.)	Bsl. 2 (mês)	XGBoost	Voting	biLSTM	SST
1318	0,145	0,143	0,709	0,526	0,214	0,57
12636	0,112	0,159	0,702	0,703	0,177	0,554
27975	0,163	0,174	0,715	0,730	0,258	0,622
43355	0,172	0,186	0,69	0,683	0,165	0,678
28310	0,147	0,179	0,714	0,657	0,166	0,7
33879	0,156	0,147	0,741	0,787	0,271	0,715
31543	0,135	0,132	0,706	0,699	0,274	0,724
27431	0,168	0,180	0,607	0,598	0,302	0,581
1209	0,170	0,114	0,525	0,491	0,240	0,396
48264	0,126	0,143	0,68	0,671	0,125	0,64
Média	0,149	0,156	0,679	0,654	0,219	0,618
Desvio padrão	0,019	0,023	0,061	0,087	0,056	0,094

Tabela 6.7: Resultados de Precisão para a tarefa de predição da posição do pico de consumo de energia elétrica.

ID	Bsl. 1 (sem.)	Bsl. 2 (mês)	XGBoost	Voting	biLSTM	SST
1318	0,145	0,145	0,579	0,469	0,248	0,507
12636	0,112	0,159	0,584	0,592	0,226	0,464
27975	0,162	0,174	0,689	0,679	0,252	0,613
43355	0,172	0,187	0,574	0,565	0,259	0,487
28310	0,147	0,180	0,705	0,653	0,237	0,668
33879	0,156	0,146	0,678	0,777	0,297	0,678
31543	0,135	0,131	0,649	0,732	0,312	0,712
27431	0,168	0,181	0,541	0,558	0,314	0,565
1209	0,170	0,115	0,475	0,420	0,278	0,368
48264	0,126	0,144	0,609	0,514	0,23	0,566
Média	0,149	0,156	0,608	0,596	0,265	0,563
Desvio padrão	0,019	0,023	0,069	0,108	0,031	0,103

Tabela 6.8: Resultados de Recall para a tarefa de predição da posição do pico de consumo de energia elétrica.

ID	Bsl. 1 (sem.)	Bsl. 2 (mês)	XGBoost	Voting	biLSTM	SST
1318	0,145	0,144	0,603	0,450	0,222	0,509
12636	0,112	0,159	0,605	0,593	0,198	0,49
27975	0,162	0,174	0,695	0,688	0,220	0,61
43355	0,172	0,186	0,600	0,582	0,195	0,5
28310	0,147	0,180	0,706	0,651	0,187	0,679
33879	0,156	0,146	0,690	0,779	0,237	0,673
31543	0,135	0,131	0,670	0,705	0,277	0,709
27431	0,168	0,180	0,546	0,540	0,300	0,558
1209	0,170	0,115	0,493	0,431	0,250	0,375
48264	0,126	0,144	0,626	0,513	0,157	0,579
Média	0,149	0,156	0,623	0,593	0,224	0,568
Desvio padrão	0,019	0,023	0,065	0,107	0,041	0,098

Tabela 6.9: Resultados de F1-score para a tarefa de predição da posição do pico de consumo de energia elétrica.

6.6 Respostas às questões de pesquisa

- **RQ1:** A combinação de algoritmos que consideram a correlação entre os atributos de entrada (regressão extrínseca) e a correlação entre os atributos-alvo (regressão *multi-target*), pode levar a resultados que superam abordagens de regressão mais simples que não consideram tais correlações?
- **Resposta:** os métodos *multi-target* geraram resultados levemente superiores aos métodos *single-target*, porém, os resultados se mostraram ser estatisticamente não significativos. A baixa correlação entre as variáveis-alvo contribuiu para este resultado, fazendo com que métodos como o DSTARS se comportassem de maneira muito similar a um *single-target*.
- **RQ2:** Dentre os diferentes algoritmos de regressão e regressão *multi-target*, qual a combinação de algoritmos de ambas as abordagens leva aos melhores resultados preditivos para a tarefa de predição de picos?
- **Resposta:** o algoritmo Linear Regression, combinado ou não ao método *multi-target* SST, superou os demais métodos na maioria dos testes para predição da magnitude, inclusive considerando diferentes tipos de matrizes de características. O Voting se mostrou ser o algoritmo mais eficiente para predição da posição, e em segundo lugar o XGBoost, também combinado ou não ao método *multi-target* SST. Porém, diferentemente da magnitude, a predição da posição mostrou sofrer influência do tipo da matriz de característica, mostrando resultados diferentes de acordo com os atributos considerados para o treinamento do modelo.
- **RQ3:** Qual o impacto de se utilizar diferentes combinações de características extraídas das séries temporais? Por exemplo, utilizar a soma do consumo diário, a magnitude do pico, a posição do pico, ou uma combinação destes fatores?

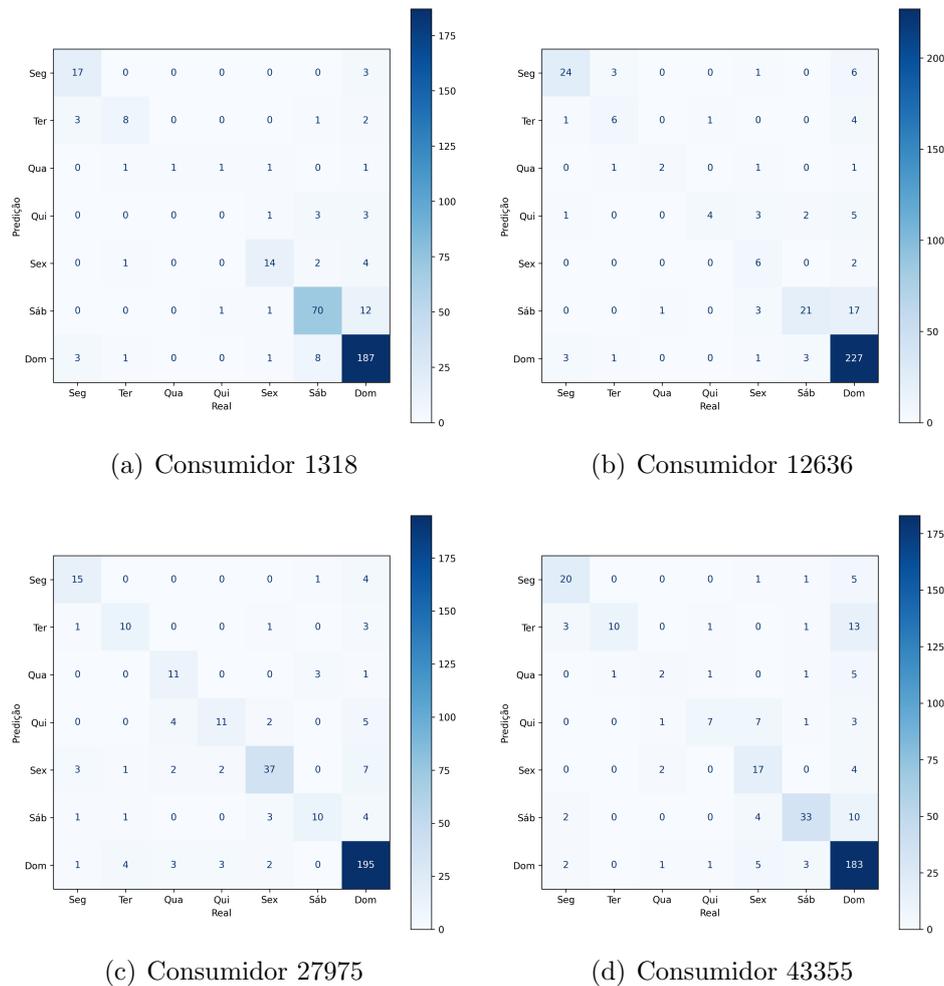
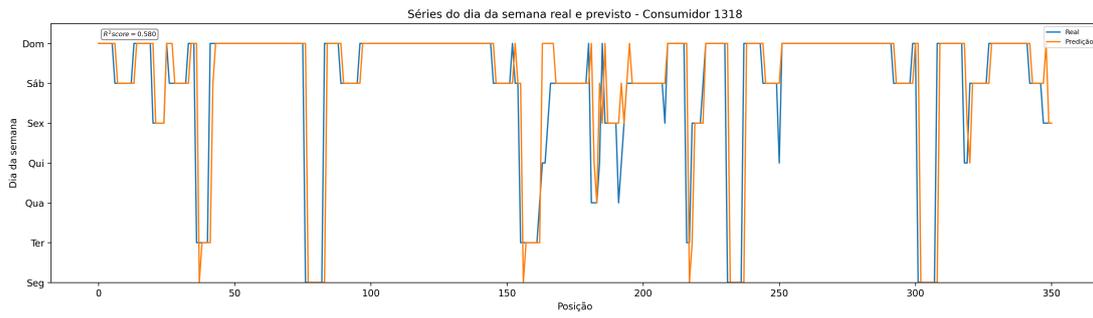
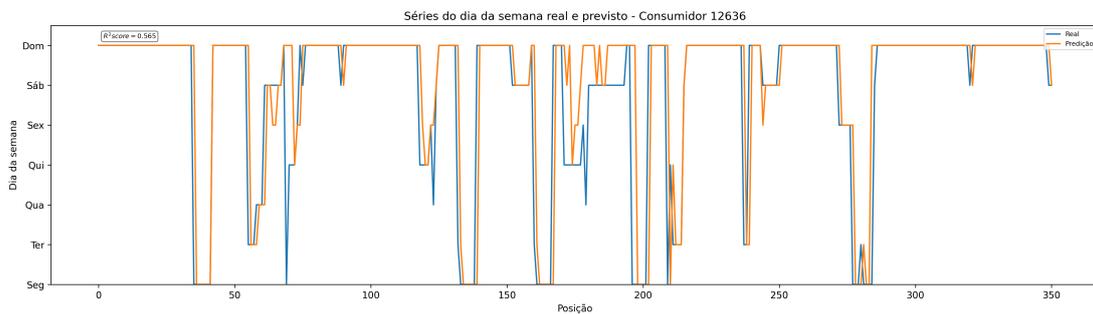


Figura 6.19: Matriz de confusão das predições da posição do pico do algoritmo XGBoost para quatro consumidores.

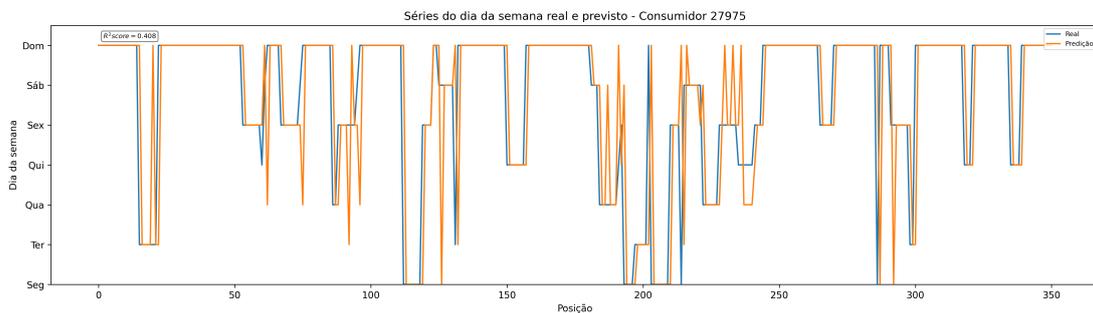
- **Resposta:** a partir da avaliação de diferentes combinações de características para predição da magnitude e da posição, ficou evidente que a magnitude sofre menos influência da matriz de característica que a posição. A predição da posição se mostrou ser mais desafiadora, passando por uma avaliação maior que considerasse também a tarefa como regressão ou classificação. Em todas as avaliações realizadas, os resultados mostraram que a predição da posição é eficiente quando se utiliza a matriz do tipo D (posição do pico semanal) em uma tarefa de classificação. As demais combinações resultaram em resultados bastante ruins e não foram considerados.
- **RQ4:** Qual o impacto do tamanho das instâncias utilizadas no treinamento do modelo preditivo (por exemplo, dados de consumo de 1 dia, 1 semana, 2 semanas ou 1 mês) nos resultados de predição?



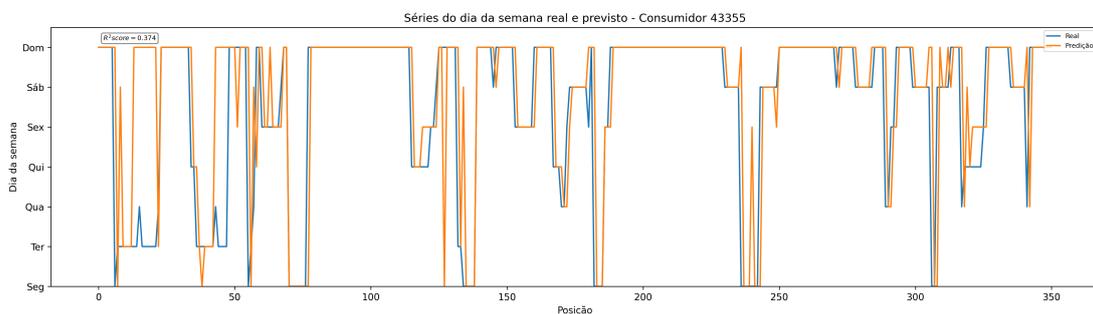
(a) Consumidor 1318



(b) Consumidor 12636



(c) Consumidor 27975



(d) Consumidor 43355

Figura 6.20: Série temporal das previsões da posição do pico do algoritmo XGBoost para quatro consumidores.

- **Resposta:** os resultados mostraram que o desempenho da predição se degrada na medida que o tamanho da instância aumenta. Inclusive os métodos *single-target* profundos, que é esperado se beneficiarem de instâncias maiores, não apresentaram melhora no desempenho com instâncias maiores. Por essa razão que se concluiu que o melhor tamanho de janela é de 1 semana.
- **RQ5:** Para o problema de predição de picos, a melhor abordagem é a construção de um modelo individual para cada consumidor ou um modelo global com dados de diferentes consumidores?
- **Resposta:** os resultados comprovaram que, para predição da magnitude, a abordagem de um modelo global compromete o desempenho, mas favorece a predição da posição principalmente quando se utilizam algoritmos baseados em *deep learning* ou métodos *multi-target*.

6.7 Considerações finais

A partir da análise dos resultados, é possível concluir que a janela de uma semana produziu os melhores resultados de predição. Além disso, utilizar matrizes de características diferentes para cada variável-alvo melhorou os resultados e, desta maneira, o projeto escolheu construir modelos preditivos independentes para cada variável-alvo.

Em linhas gerais, algoritmos *multi-target* não apresentaram resultados superiores a modelos *single-target*, indicando que a falta de correlação entre as variáveis-alvo prejudicou o desempenho destes algoritmos. De qualquer maneira, o desempenho obtido nas tarefas de regressão e classificação foram bastante satisfatórias, no momento que o projeto escolheu utilizar combinações de características e algoritmos distintos para predição de cada variável-alvo.

Em específico para predição da posição, utilizar uma matriz de características global com todos os consumidores produziu resultados superiores, melhorando significativamente o desempenho da maioria dos algoritmos, principalmente aqueles que combinam métodos *multi-target*. Porém, o mesmo comportamento não existiu para predição da magnitude, realçando a limitação dos métodos *multi-target* em prever, com desempenho superior, ambas as variáveis alvo da aplicação.

Modelos de *deep learning* apresentaram desempenho insatisfatório em virtude do baixo número de amostras de cada consumidor, quando foram construídos a partir de matrizes de características individuais por consumidor. A análise com modelos globais obteve resultados melhores, principalmente quando se combinou algoritmos baseados em *deep learning* com métodos *multi-target*.

Embora não apresentado com detalhes nesta seção para simplificar a análise dos dados, na média nenhum algoritmo *multi-target* apresentou diferenças significativas

de desempenho para nenhum dos consumidores avaliados pelo projeto, apontando que a aplicação utilizada para avaliar o projeto reproduz variáveis-alvo sem correlação. Para reforçar esta hipótese, chegou a ser explorado um outro conjunto de dados para prever a magnitude de pico e posição de casos de Covid19, que geraram resultados bastante similares aos apresentados por este projeto, denotando que problemas de predição de pico podem se tratar de problemas que não possuem correlação entre as variáveis-alvo.

Capítulo 7

Conclusões

No contexto deste trabalho, diversas técnicas foram exploradas para estabelecer um método capaz de prever picos em séries temporais, utilizando somente os dados brutos das séries, também conhecido por regressão extrínseca. Diferentes abordagens, explicadas no protocolo experimental, foram testadas para encontrar a melhor combinação de conjunto de características e algoritmo que maximize o desempenho dos modelos de regressão ou classificação. Embora os resultados obtidos sejam satisfatórios, algumas limitações foram encontradas, que serão discutidas nas seções a seguir.

7.1 Considerações iniciais

Neste capítulo são apresentados limitações encontradas durante este trabalho de pesquisa, que podem servir de base para trabalhos futuros relacionados ao tema de regressão em séries temporais. Além disto, a aplicação explorada pelo trabalho, que consiste na predição da magnitude e da posição de um pico em uma série temporal que expressa o consumo de energia de um certo cliente, tem a característica de trabalhar com apenas duas variáveis-alvo, que se mostrou no decorrer deste trabalho ser uma limitação na aplicação de algoritmos *multi-target*.

7.2 Limitações

Embora os resultados obtidos pelo trabalho sejam satisfatórios, algumas limitações foram encontradas no decorrer do projeto, que servem de base para trabalhos futuros.

No capítulo 6 foi apresentado que, para elevar o desempenho de predição de todas as variáveis-alvo, foi escolhido utilizar matrizes de características distintas para cada

variável. Isso por si invalida um conceito da regressão *multi-target* que é explorar a correlação entre as variáveis-alvo, para elevar o desempenho da predição, pois, uma vez que se use conjuntos de características distintos por variável, se retira da arquitetura do modelo a correlação entre as variáveis.

Outro aspecto importante é que a aplicação explorada pelo trabalho possui somente duas variáveis-alvo, significando que há apenas uma combinação possível de correlações a serem exploradas pelos algoritmos *multi-target*. Isso explica o fato de, na maioria dos casos, os algoritmos *multi-target* executarem suas predições praticamente como se fossem *single-target*, devido a ausência de correlação entre a magnitude e a posição dos picos. Com isso, podemos afirmar que algoritmos de regressão ou classificação *multi-target* se beneficiam mais quando o número de variáveis-alvo é maior, fornecendo um número maior de combinações a serem exploradas, que podem elevar o desempenho da predição.

7.3 Trabalhos futuros

Este trabalho de pesquisa percebeu que há atividades que podem ser executadas em trabalhos futuros, com a finalidade de aprimorar técnicas de regressão extrínseca em séries temporais.

Um trabalho não realizado pela pesquisa foi de combinar grupos de clientes, de maneira a encontrar um padrão de consumo em um determinado grupo de clientes, e desenvolver um modelo de predição que atenda a um número maior de clientes ao invés de apenas um.

Outro ponto importante também não realizado foi explorar com mais detalhes se é frequente o fato de não existir correlação entre a magnitude e a posição do pico em séries temporais de consumo de energia. O trabalho se concentrou no conjunto de dados fornecido pela LADPU, sem explorar outros conjuntos de consumo de energia. Então, é algo que pode ser explorado por trabalhos futuros, até mesmo para confirmar se o melhor modelo de predição para esta aplicação realmente são os algoritmos *multi-target*.

Referências

- Aghabozorgi, S., Shirkhorshidi, A. S., e Wah, T. Y. (2015). Time-series clustering—a decade review. *Information Systems*, 53:16–38.
- Aho, T., Ženko, B., Džzeroski, S., Elomaa, T., e Brodley, C. (2012). Multi-target regression with rule ensembles. *Journal of Machine Learning Research*, 13(8).
- Bagnall, A., Lines, J., Hills, J., e Bostrom, A. (2015). Time-series classification with cote: the collective of transformation-based ensembles. *IEEE Transactions on Knowledge and Data Engineering*, 27(9):2522–2535.
- Benavoli, A., Corani, G., e Mangili, F. (2016). Should we really use post-hoc tests based on mean-ranks? *The Journal of Machine Learning Research*, 17(1):152–161.
- Borchani, H., Varando, G., Bielza, C., e Larranaga, P. (2015). A survey on multi-output regression. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(5):216–233.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Chemetova, S., Santos, P., e Ventim-Neves, M. (2016). Load peak forecasting in different load patterns situations. In *2016 10th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, páginas 148–151. IEEE.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., et al. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4.
- Cortes, C. e Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.
- Dai, S., Meng, F., Dai, H., Wang, Q., e Chen, X. (2021). Electrical peak demand forecasting—a review. *arXiv preprint arXiv:2108.01393*.
- Dechter, R. (1986). Learning while searching in constraint-satisfaction problems.
- Dempster, A., Petitjean, F., e Webb, G. I. (2020). Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5):1454–1495.

- Dowling, C. P., Kirschen, D., e Zhang, B. (2018). Coincident peak prediction using a feed-forward neural network. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, páginas 912–916. IEEE.
- Drucker, H., Burges, C. J., Kaufman, L., Smola, A., e Vapnik, V. (1996). Support vector regression machines. *Advances in neural information processing systems*, 9.
- Dundar, M., Krishnapuram, B., Bi, J., e Rao, R. B. (2007). Learning classifiers when the training data is not iid. In *IJCAI*, volume 2007, páginas 756–61.
- Fang, K., Pan, M., e Shen, C. (2018). The value of smap for long-term soil moisture estimation with the help of deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, 57(4):2221–2233.
- Feng, D., Fang, K., e Shen, C. (2020). Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resources Research*, 56(9):e2019WR026793.
- Fulcher, B. D. e Jones, N. S. (2017). hctsa: A computational framework for automated time-series phenotyping using massive feature extraction. *Cell systems*, 5(5):527–531.
- Gajowniczek, K., Nafkha, R., e Ząbkowski, T. (2017). Electricity peak demand classification with artificial neural networks. In *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, páginas 307–315. IEEE.
- Gers, F. A., Schraudolph, N. N., e Schmidhuber, J. (2002). Learning precise timing with lstm recurrent networks. *Journal of machine learning research*, 3(Aug):115–143.
- Hadavandi, E., Shahrabi, J., e Shamshirband, S. (2015). A novel boosted-neural network ensemble for modeling multi-target regression problems. *Engineering Applications of Artificial Intelligence*, 45:204–219.
- Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., Webb, G. I., Idoumghar, L., Muller, P.-A., e Petitjean, F. (2020). Inception-time: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, 34(6):1936–1962.
- Jain, A. K., Mao, J., e Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3):31–44.
- Kocev, D. e Ceci, M. (2015). Ensembles of extremely randomized trees for multi-target regression. In *International Conference on Discovery Science*, páginas 86–100. Springer.

- Kocev, D., Vens, C., Struyf, J., e Džeroski, S. (2007). Ensembles of multi-objective decision trees. In *European conference on machine learning*, páginas 624–631. Springer.
- Kruger, R., Mueen, A., e Souza, V. M. A. (2023). Load peak forecasting on energy consumption using smart meters. In *Submetido ao IEEE International Conference on Machine Learning and Applications (ICMLA)*, páginas 1–7. IEEE.
- Le, X.-H., Nguyen, D.-H., Jung, S., Yeon, M., e Lee, G. (2021). Comparison of deep learning techniques for river streamflow forecasting. *IEEE Access*, 9:71805–71820.
- Lines, J., Davis, L. M., Hills, J., e Bagnall, A. (2012). A shapelet transform for time series classification. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, páginas 289–297.
- Lines, J., Taylor, S., e Bagnall, A. (2016). Hive-cote: The hierarchical vote collective of transformation-based ensembles for time series classification. In *2016 IEEE 16th international conference on data mining (ICDM)*, páginas 1041–1046. IEEE.
- Loh, W.-Y. (2011). Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1(1):14–23.
- Lubba, C. H., Sethi, S. S., Knaute, P., Schultz, S. R., Fulcher, B. D., e Jones, N. S. (2019). catch22: Canonical time-series characteristics. *Data Mining and Knowledge Discovery*, 33(6):1821–1852.
- Mastelini, S. M., Santana, E. J., Cerri, R., e Barbon Jr, S. (2020). Dstars: A multi-target deep structure for tracking asynchronous regressor stacking. *Applied Soft Computing*, 91:106215.
- Medsker, L. R. e Jain, L. (2001). Recurrent neural networks. *Design and Applications*, 5:64–67.
- Melamud, O., Goldberger, J., e Dagan, I. (2016). context2vec: Learning generic context embedding with bidirectional lstm. In *Proceedings of the 20th SIGNLL conference on computational natural language learning*, páginas 51–61.
- Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J., e Talagala, T. S. (2020). Fforma: Feature-based forecast model averaging. *International Journal of Forecasting*, 36(1):86–92.
- Mueen, A., Keogh, E., Zhu, Q., Cash, S., e Westover, B. (2009). Exact discovery of time series motifs. In *Proceedings of the 2009 SIAM international conference on data mining*, páginas 473–484. SIAM.
- Mughees, N., Mohsin, S. A., Mughees, A., e Mughees, A. (2021). Deep sequence to sequence bi-lstm neural networks for day-ahead peak load forecasting. *Expert Systems with Applications*, 175:114844.

- O'Shea, K. e Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.
- Palshikar, G. et al. (2009). Simple algorithms for peak detection in time-series. In *Proceedings of the 1st International Conference of Advanced Data Analysis, Business Analytics and Intelligence*, volume 122, páginas 1–14.
- Parmezan, A. R. S., Souza, V. M. A., e Batista, G. E. A. P. A. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information sciences*, 484:302–337.
- Phan, D., Bonnet, S., Guillemaud, R., Castelli, E., e Thi, N. P. (2008). Estimation of respiratory waveform and heart rate using an accelerometer. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, páginas 4916–4919. IEEE.
- Prasad, A., Roy, S., Sarkar, A., Panja, S. C., e Patra, S. N. (2022). Prediction of solar cycle 25 using deep learning based long short-term memory forecasting technique. *Advances in Space Research*, 69(1):798–813.
- Quinlan, J. R. (1993). C4. 5: Programs for machine learning.
- Ratanamahatana, C. A., Lin, J., Gunopulos, D., Keogh, E., Vlachos, M., e Das, G. (2005). Mining time series data. In *Data mining and knowledge discovery handbook*, páginas 1069–1103. Springer.
- Russell, S. e Norvig, P. (2002). Artificial intelligence: a modern approach.
- Sammut, C. e Webb, G. I. (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.
- Santos, P. J., Chemetova, S. R., Pires, A. J., e Neves, M. V. (2020). The importance of daily load forecasting for main electric substation in a spot market context. In *2020 IEEE 14th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, volume 1, páginas 50–54.
- Soman, A., Trivedi, A., Irwin, D., Kosanovic, B., McDaniel, B., e Shenoy, P. (2020). Peak forecasting for battery-based energy optimizations in campus microgrids. In *Proceedings of the Eleventh ACM International Conference on Future Energy Systems*, páginas 237–241.
- Souza, V., Estrada, T., Bashir, A., e Mueen, A. (2020). Ladpu smart meter data. Dryad. <https://doi.org/10.5061/dryad.m0cfxpp2c>.
- Souza, V. M. A., Silva, D. F., e Batista, G. E. A. P. A. (2014). Extracting texture features for time series classification. In *International Conference on Pattern Recognition*, páginas 1425–1430. IEEE.

- Spyromitros-Xioufis, E., Tsoumakas, G., Groves, W., e Vlahavas, I. (2016). Multi-target regression via input space expansion: treating targets as inputs. *Machine Learning*, 104(1):55–98.
- Tan, C. W., Bergmeir, C., Petitjean, F., e Webb, G. I. (2021). Time series extrinsic regression. *Data Mining and Knowledge Discovery*, páginas 1–29.
- Vanderstar, G., Musilek, P., e Nassif, A. (2018). Solar forecasting using remote solar monitoring stations and artificial neural networks. In *2018 IEEE Canadian conference on electrical & computer engineering (CCECE)*, páginas 1–4. IEEE.
- Wilcoxon, F., Katti, S., Wilcox, R. A., et al. (1970). Critical values and probability levels for the wilcoxon rank sum test and the wilcoxon signed rank test. *Selected tables in mathematical statistics*, 1:171–259.
- Xu, S., An, X., Qiao, X., Zhu, L., e Li, L. (2013). Multi-output least-squares support vector regression machines. *Pattern Recognition Letters*, 34(9):1078–1084.
- Xue-Song, Z., Li-Qiang, C., e You-Jie, M. (2010). Research on smart grid technology. In *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*, volume 3, páginas V3–599. IEEE.
- Ye, L. e Keogh, E. (2009). Time series shapelets: a new primitive for data mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, páginas 947–956.
- Yegnanarayana, B. (2009). *Artificial neural networks*. PHI Learning Pvt. Ltd.