

Contribuições das competições M para o aprimoramento dos métodos de previsão de séries temporais

Igor Costa de Freitas

Graduando em Engenharia de Produção - UFMG

E-mail: igorfreitascp@gmail.com

Marcelo Azevedo Costa

Professor Titular do Departamento de

Engenharia de Produção/Universidade Federal de Minas Gerais - UFMG.

E-mail: macosta@ufmg.br

Resumo:

A grande questão que envolve o planejamento de suprimentos é a incerteza da demanda futura. Desde os primeiros ensaios acerca do futuro de vendas, modelos de consumo vêm sendo estudados e aprimorados. O modelo do lote econômico de 1918 é bastante famoso por sua eficácia para problemas simples, mas conforme a complexidade das redes de suprimentos cresce, os problemas de dimensionamento de lote e compra também ficam mais complexos. Para isso surgiu a análise de séries temporais com intuito de lidar com este tipo de problema e fazer previsões mais acuradas. O presente texto visa a abordagem em forma de revisão bibliográfica das publicações sobre as competições de séries temporais, especialmente as competições M, organizadas pelo *Makridakis Open Forecasting Center* desde o fim dos anos 1970 e com sua última edição em 2020. O estudo busca expor o contexto das competições e seus impactos na área de modelagem e previsão de séries temporais.

Palavras-chave: Séries temporais, Competições M, Cadeia de suprimentos, Demanda.

Abstract:

The major problem in the supply chain is to deal with the variance of demand, for this since the beginning of sells future researches, consumption models has been chased, the Economic Production Quantity of 1918 is quite famous by its precision in small problems, but as que complexity of the supply networks grows up the problems of sizing batches e purchases became also too complicated. The time series analysis came with the intention of dealing with this type of data and perform more accurate forecasts. This paper addresses as a bibliographic review the articles about time series competitions, specially the M competitions, founded and organized by the Makridakis Open Forecasting Center since the end of the 1970 and with its last edition 2020. The study aims to expose the competitions context and how it has been enhancing the area of time series forecasting.

Keywords: Time series, M competitions, Supply Chain, Demand.

1. Introdução

Na área de ciência dos dados um tema bastante explorado pelas empresas é a capacidade de prever valores seriais de demanda que podem trazer respostas como a quantidade a ser produzida, comprada, estocada e vários outros desafios da indústria. Essas são as chamadas séries temporais, que podem ser definidas como uma sequência de observações realizadas ao longo do tempo (Ehlers, 2007), sendo a ordem dessa coleta de dados o fator crucial. Análise de Séries Temporais é um campo particularmente interessante tanto em aplicações em problemas industriais, como também nas áreas de macroeconomia, análises climáticas, biológicas, dentre outras várias.

O uso intenso de bases de dados propiciou nos últimos anos uma procura acentuada por estudantes e profissionais já formados por tecnologias de desenvolvimento de ferramentas e modelos. Associado a este fato, surgem plataformas que possibilitam a diferentes usuários a inserção e a disponibilidade imediata de acesso a conteúdos de uso e auxílio a novas linguagens de programação e de manipulação de dados. Plataformas como *StackOverflow*, *Kaggle* e *GitHub* fornecem dados de diversos tipos de problema, códigos para tratamento, visualização, extração, exportação e análise de dados, dentre outras ferramentas, voltados ao desenvolvimento de soluções computacionais, cursos, desafios, competições e potenciais

ofertas de vagas de emprego. Em plataformas como o *Kaggle*, plataforma dedicada à aplicação de técnicas de Ciência dos Dados, inúmeras competições de previsão de séries temporais são realizadas (BOJER e MELDGAARD, 2021). Tais competições são fontes de aprendizado e apresentam conteúdo documentado e de qualidade disponível ao público.

As competições M existem desde o início dos anos 1980, e já foram realizadas cinco edições desde então, o intuito de uma competição M é de busca pela melhoria e estado da arte do campo de previsão de séries. Atualmente as competições são totalmente abertas. Torneios desse tipo contam com uma base de dados pré-selecionada que será analisada pelas equipes participantes. Cada equipe fará suas previsões, utilizando métodos diversos de análise da precisão e diferentes modelos estatísticos e computacionais. Em grande parte das competições, medidas de precisão dos modelos são divulgadas somente ao final das contribuições, quando as equipes vencedoras são identificadas, como foi o caso da competição M5. Um ponto importante em competições de séries temporais é o volume de séries fornecidas. Séries reduzidas podem ser insuficientes para desenvolvimento de modelos complexos de previsão. Para isso a quantidade de dados deve ser massiva, como nas competições M4 e M5 que tiveram, respectivamente, 100,000 séries reais acerca de dados macroeconômicos e 42,000 séries de vendas diárias de produtos do Walmart, as três competições anteriores desafiaram seus participantes em previsões de uma quantidade bem menor de dados (a mesma fonte da quarta edição foi a fonte das três primeiras, mas menores em montante de dados), o que reflete a evolução computacional dos modelos e do compartilhamento de conteúdo do tema e corrobora com a alta busca pela área citada.

São realizadas diversas competições por ano na plataforma Kaggle como tráfego na rede web, estocagem de itens, vendas e competições de dados esportivos. Algumas outras competições já passadas merecem destaque como a Neural Forecasting Competition, conhecida como NN, a NN3, NN5 e NNC que possuem alguns dados e formato baseado no M3, são bastante citadas e geraram um bom conteúdo no início do século (LEMKE e GABRYS, 2010 e CRONE et al. 2011).

2. Objeto de estudo

Uma série temporal nada mais é que uma sequência de valores distribuídos ao longo do tempo de alguma medição de interesse, como a produção de smartphones nos Estados Unidos, a poluição por combustíveis fósseis ao redor do mundo, a quantidade de minério de ferro produzida por uma empresa no ano e dentre outras. O que se deseja em um estudo de

previsão de séries temporais é obter um modelo que seja capaz de gerar uma sequência que mais se aproxima da sequência real e seja capaz de prever futuros valores, é necessário dizer que a precisão é muito importante mas o objeto de maior valor na análise é o quão *ajustada* à curva real a curva de previsão se encontra, isto é, um comportamento parecido entre as duas curvas é almejado em detrimento da precisão absoluta de valores (que pode ser encarada com negativa em algumas situações, o *overfitting*, que indica que possivelmente o modelo terá dificuldades de lidar com mudança repentinas na série). Aqui são descritas as componentes de uma série temporal (Ehlers, 2007):

- Tendência: trata-se do comportamento enviesado ao longo do tempo de uma série, pode ser observado em lançamento de produtos e/ou crescimento de demanda por certo item no mercado ou serviço (streaming, nos últimos anos por exemplo);
- Sazonalidade: Basicamente uma mudança mais brusca no comportamento durante um período menor do que o período total da série (ano, mês, etc), comum em dados agrícolas, vendas em feriados, efeitos do clima e afins;
- Ciclos: Variações no estado da série que possuem padrão superior ao período de medição da série, pode sempre vir à tona com certa periodicidade, crises econômicas, crescimento econômico de um país. Fenômenos climáticos são exemplo de séries que certamente contam com a presença de ciclos;
- Aleatoriedade: São imprevisíveis no comportamento da série, variações abruptas que são provocadas por fatores externos e incomuns, como um atentado terrorista, decisões empresariais e governamentais impulsivas com impacto em grandes cenários.

Assim os modelos se concentram em identificar principalmente os três primeiros tópicos, sua presença e seu peso no comportamento da série, como pode ser descrito computacionalmente e se o modelo foi capaz de entender suas características.

As competições fornecem os dados em um repositório como Google Drive e GitHub para que os participantes investiguem todas as características que as séries demonstram, uma série pode apresentar sazonalidade e tendência ao mesmo tempo, bem como ser cíclica e sazonal. Modelos se concentram em abarcar todas essas possíveis fontes de variação.

Um modelo construído computacionalmente que enxerga as variações deve, antes da análise, ter seus dados de imputação tratados, após este processo crucial, os conjuntos de treino e de teste se tornam a atenção principal do processo de desenvolvimento do modelo, o primeiro será utilizado, bem como o nome, para o treinamento do modelo, é onde o modelo aprende dos dados e/ou tenta aplicar padrões estatísticos que melhor se encaixe neles, o segundo faz parte da validação do modelo com dados que este nunca teve contato, dados novos para o modelo, é a etapa de teste e nela é que se obtém uma resposta sobre a adequação da previsão com os dados de interesse. O tratamento de dados, etapa crucial de análises estatísticas no geral, é um ponto bastante delicado da construção do modelo, pois não se deseja mutilar (MEIRA e OLIVEIRA, 2021) a amostra de dados mas algumas observações devem ser retiradas para que não causem a identificação de um padrão inexistente nos dados, pois valores muito discrepantes caem no fator de aleatoriedade das séries e se tornam empecilhos na previsão pois não fazem parte da realidade comum das séries. Para isso, é comum eliminar *outliers* estatísticos, com uma e meia distância interquartílica ou mais de tolerância para os valores da série, outra opção é uma análise exploratória profunda dos dados que podem gerar respostas mais qualitativas para a compreensão dos dados, em alguns gráficos e relatórios a tendência e sazonalidade pode ser vista à olho nu (pequena parcela).

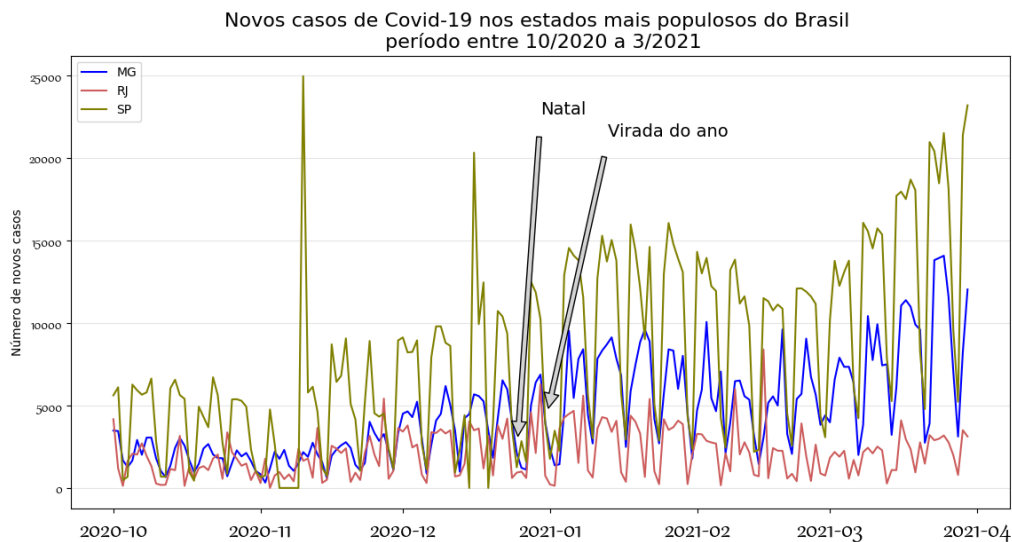


Figura 1: Curva de novos casos diários de COVID-19 e sua tendência de aumento logo após os encontros e festas de fim de ano no Brasil, fator determinante na transmissão do vírus. Elaborado pelo autor. Fonte dos dados: Corona virus Brazil/Kaggle.

Um bom exemplo, é o pico de casos de COVID-19 vivido no Brasil em janeiro e fevereiro de 2021, logo após as festas de fim de ano, tradicionais no país, os números de novos casos ascenderam abruptamente, o que configurou um dos piores período de transmissão do vírus. Foi um cenário em que a tendência em uma série (novos casos diários) pôde ser observada. Outra boa ilustração, é como o preço do barril de petróleo sofre alterações diárias devido a fatores políticos e de incerteza no mercado(LIU, 2020), a história do produto sempre esteve associada às questões governamentais e diretrizes das grandes empresas mundiais. A substância tem um histórico de preços bastante sazonal em várias partes de sua trajetória, e um período de destaque é seu valor nos anos que sucederam a Guerra do golfo, conflito com viés bastante focado na posse do petróleo.

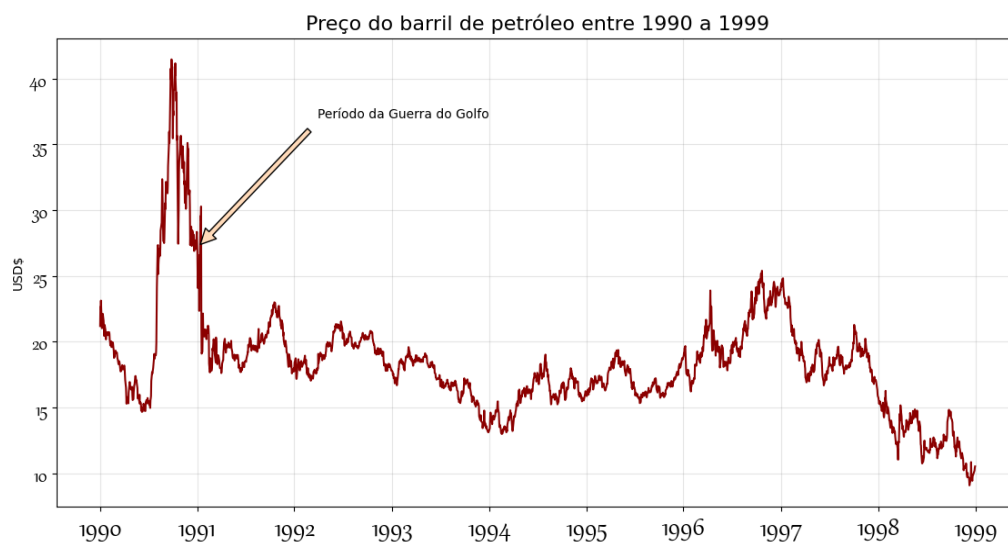


Figura 2: Evolução do preço do barril do petróleo na década de 1990, observa-se uma sazonalidade de baixa nos valores por volta da época de mudança de um ano para outro. Elaborado pelo autor. Fonte dos dados: Brent Oil Price/Kaggle.

O treino e teste são etapas que podem ser feitas separadamente ou em conjunto. No caso de competições, existem regras que descrevem como deve ser feita a validação do modelo e quais tipos não podem ser usados. Um método bastante utilizado é o de validação cruzada, porém para séries temporais a validação cruzada é mais delicada do que para problemas de classificação ou previsão baseada em *features*, em que amostras podem ser retiradas e embaralhadas, nas séries o componente hierárquico é crucial para compreensão dos dados pelos modelos e pelos criadores do modelo, um valor de uma série é único porque além de seu

significado numérico a posição que ele ocupa na sequência também é uma característica que será avaliada pelo modelo. Portanto a dependência direta de valores vizinhos não pode ser desconsiderada, para isso existem métodos que partem a série em séries menores justamente na intenção de não violar a hierarquia das observações e testar o modelo com os dados imediatamente posteriores a esse pequeno recorte.

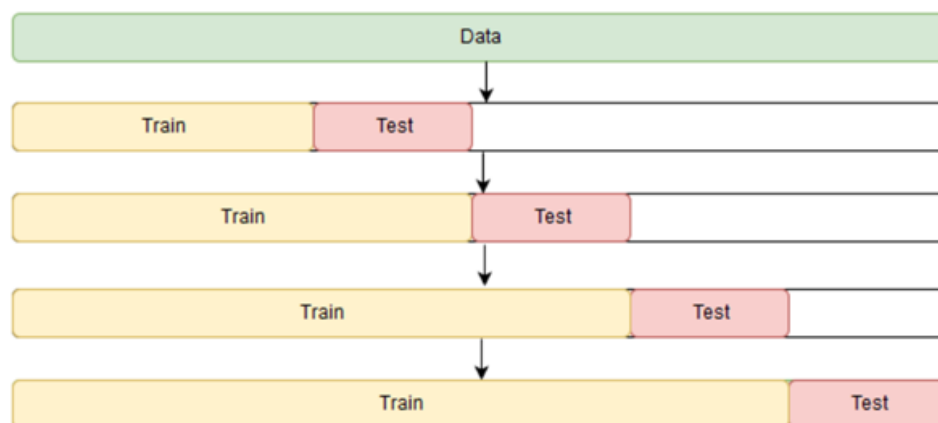


Figura 3: Exemplo de validação cruzada com séries temporais. Fonte: Soumya Shrivastava, 2020.

A etapa também pode ser composta por *subsets* de dados que não possuem valores em comum, apenas parte do conjunto de teste de uma iteração da validação como conjunto de treino de uma outra iteração. Deve-se ter em mente que o valor muitas vezes não têm significado sozinho e sim em qual ponto da trajetória ele se encontra e quais são seus arredores, a investigação de autocorrelação não pode ser realizada quando uma peça está ausente, bem como a previsão, que não pode ser realizada adequadamente caso os dados de treino e teste se encontrem em partes diferentes ao longo da amostra.

Para desenvolvimento de modelos são utilizadas linguagens de programação que permitem extração dos dados, processamento, visualização e análise. É comum na comunidade programadoras a divulgação de códigos de modelos inteiros através das redes de interação, são diversos *scripts* disponíveis de um usuário para outro. O que é mais procurado pelos estudiosos são linguagens que processem os dados com a presença de ferramentas estatísticas e de Machine Learning suficientes para compreender problemas complexos e que não apresente lentidão excessiva na análise, pois grandes modelos são acompanhados de grandes bases de dados e a eficiência computacional permite maior tempo de análise qualitativa e teste de mais possibilidades. A linguagem de programação conta com a presença

de bibliotecas e módulos que fornecem as ferramentas de análise, como uma rede neurais e seus parâmetros (números de nós, tempo limite de análise, número máximos de iterações, etc), esses pacotes de ferramentas existem em todas as tecnologias de sistemas e alguns são considerados diferenciais para escolha de linguagens como R, Python, C, C++, C#, Java e etc.

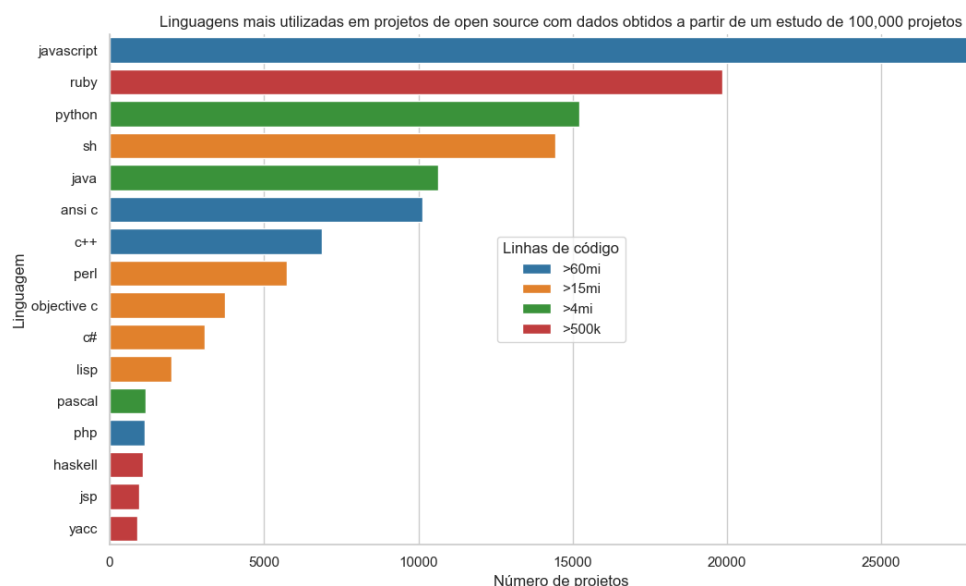


Figura 4: Linguagens de programação mais utilizadas em um estudo que levantou 100,000 projetos de *open source*. Elaborado pelo autor. Fonte: Popularity, interoperability, and impact of programming languages in 100,000 open source projects.

3. As competições M

Foi baseado em suposições de combinação de modelos preliminares que o estudos da previsão de dados sofreu sua maior reviravolta, à época as competições eram realizadas de forma muito restrita e apenas a elite da academia participava dos torneios que eram particularmente bem triviais, apenas estudiosos testando seus modelos uns contra os outros em horizontes de previsão bem extensos, foram os estudos de PhD na universidade de Nottingham de David Reid, juntamente com os estudiosos Paul Newbold e Clive Granger que conduziram uma evolução no estudos de séries temporais com teste de precisão com modelos aplicados a 106 séries(NEWBOLD e GRANGER, 1974) mas também com caráter restrito e seus resultados apresentados ao restante da academia e seus métodos não conquistaram muita popularidade no meio.

“Métodos simples desempenham bem em comparação com os mais complexos e estatisticamente sofisticados”

– Makridakis, Hibon e Moser, 1979

No final dos anos 1970, os pesquisadores Makridakis e Hibon resolveram analisar 111 séries temporais com diversos métodos, seus resultados foram expostos em uma publicação e apresentados para a Sociedade Real de Estatística, nessa publicação revolucionária os autores fazem uma severa crítica aos pensamentos que dominavam o meio de previsão de séries naquele tempo, a crença de que existe um modelo que governa os dados e o dever do cientistas de dados é encontrar e descrever este modelo matematicamente não era mais suficiente para este campo de estudos, foi exposto no texto que modelos de previsão combinados se saíram melhores nos testes porém a sociedade estudiosa do tema desconsiderou esses indicadores e até com certo desdém pelos autores (HYNDMAN, 2020). Devido ao caráter fechado das competições de previsões que haviam ocorrido anteriormente, Makridakis e Hibon iniciaram uma competição aberta envolvendo 1001 séries temporais de dados da indústria, população e economia, que com comprimentos variados, entre 9 e 132 observações, MAKRIDAKIS et al. (1982). A competição pôde concluir que: modelo muito complexos estatisticamente não costumam ter desempenho notavelmente acima de modelos simples; A combinação de mais de um método rendeu melhor na média do que métodos estatísticos solitários; Métodos combinados têm sua performance mais afetada pelo horizonte de previsão; e um fato bastante curioso, o ranking final do torneio é diferente para medidas de acurácia diferentes, o que mostra como modelo podem ter diferentes interpretações além da precisão estritamente numérica.

Alguns anos após essa primeira edição foi realizada a M2 que contou com 29 séries das mesmas áreas do M1, um número bastante inferior ao da primeira competição, porém com mais variáveis explicativas de contexto e maiores conclusões sobre o tema foram tiradas do estudo. A edição que ficou marcada como a virada da maré para os organizadores e estudiosos Makridakis e Hibon e para a organização Makridakis Open Forecasting Center (MOFC) da universidade de Nicosia (UNIC), foi a edição M3 em 1998, que teve como intuito uma tentativa final de adequação da acurácia de modelos combinados, contou com 3003 séries de comprimento entre 14 e 126 observações acerca dos mesmo dados das duas primeiras competições, e assim como nas anteriores essa competição contou com dados não-sazonais, mensal e trimestrais e todos os dados positivos. A evidência de que modelos mais simples

podem superar modelos muito sofisticados e complicados não pôde ser provada, porém os dois melhores métodos do torneio foram o métodos Theta (ASSIMAKOPOULOS e NIKOLOPOULOS, 2000), que era composto por um método equivalente a regressão linear e suavização exponencial, métodos simples, mas não para a época, e o método de um software comercial chamado ForecastPro, que faz uma suavização exponencial e utiliza o método ARIMA para selecionar modelos de previsão através do Bayesian Information Criteria (BIC), métodos não tão simples mas de complexidade moderada em comparação com alguns modelos atuais (MAKRIDAKIS, 1993).

Em 2018, a competição M4 foi realizada com a enorme quantidade de 100 mil séries temporais com diversas mudanças no seu formato, como a presença de dados semanais, diários e horários juntamente com os outros tipos de dados, os competidores puderam submeter intervalos de predição junto com a previsão pontual e a competição focou na reprodução e compartilhamento de códigos, pedindo aos participantes para enviarem seus resultados pela plataforma GitHub, MAKRIDAKIS et al. (2018).

A competição M5, iniciada em 2020 contou com a maior mobilização da história das competições de previsões de séries temporais e bateu o recorde de participantes com 20 mil equipes no torneio, de várias pessoas ou individualmente. Nesta edição o desafio era prever o número de vendas da rede Walmart em uma janela de 28 dias a frente da data final da submissão, os dados são oriundos de lojas de três estados norte-americanos, Califórnia, Wisconsin e Texas, três dos estados mais populosos dos Estados Unidos, nesta base de dados é possível encontrar dados por item, rotulados por departamento, categoria, preço, detalhes da loja e dados de dia da semana, promoções e eventos e dois tipos de torneio foram realizados com essas informações, de precisão e de incerteza. Essa edição, em especial, contou com a presença forte e encorajamento da direção da utilização de métodos de aprendizado de máquinas além da prática já comum nas competições de Makridakis e Hibon de combinação de vários modelos estatísticos, e o foco da competição foi de fato em desenvolver conhecimento na área que possa ser facilmente replicado na indústria de distribuição de bens, MAKRIDAKIS et al. (2021).

4. Metodologia da pesquisa

Inicialmente foi definido que o estudo seria uma revisão sobre publicações com tema focado em séries temporais, um assunto que figura nas principais pautas da área de engenharia e de dados. A finalidade da pesquisa é de apresentar como a área vem evoluindo academicamente e como estas evoluções vêm atingindo o mercado de cadeias de suprimentos e setores que fazem uso de previsão de dados em forma de série.

Para fixar as bases do estudo bases de publicações renomadas e com viés em ciências exatas foram escolhidas para uma busca inicial e pouco filtrada propositalmente, a base *Scopus*, juntamente com a *ScienceDirect*, *Google Scholar* e *Engineering Village*, serviram com ferramenta de pesquisa para os textos. Foram escolhidas porque possuem um bom número de artigos no total cada uma. Nestas bases as palavras chave foram idealizadas sobre o conhecimento prévio da competição M5, buscou-se abordar em quantidade de e quais as chaves os temas que estão sempre em pauta na ciência dos dados, desde os primórdios deste tipo de estudo, para isso foram usadas "Supply planning", "Time series", "Hierarchical sales", "Supply Scheduling" foi testado mas não houveram muitas contribuições em sua utilização, também houve a preocupação com a evolução das técnicas de tratamento e predição de dados, por isso "Business forecasting", "Sales prediction" e "Machine learning" foram escolhidas, pois tratam-se de assuntos que abordam com ferramentas contemporâneas problemas tanto da academia quanto da indústria na prática, por fim "M competitions", o foco da presente análise, finalizou a lista.

Contudo, apenas as palavras cruas não foram suficientes para atingir um número viável, foi então que o estudo adotou os operadores booleanos para combinações dos termos definidos na etapa anterior, como o operador *AND* e *OR*, para melhor filtragem dos textos. As combinações escolhidas foram:

- "M competitions" AND "Business forecasting"
- "M competitions" AND "Business forecasting" OR "hierarchical sales"
- "Machine learning" AND "M competitions"
- "Machine learning" AND "M competitions"
- "Business forecasting" AND "supply planning" OR "supply scheduling"
- "time series" AND "Business forecasting"

Essa combinação de chaves é amplamente utilizada em revisões e foi crucial para convergir o estudo ao campo desejado. Em seguida, com os valores encontrados na combinação foi realizado mais uma escolha, desta vez, uma base e uma combinação foram selecionadas para prosseguir nesta análise, a ScienceDirect e o termo "Machine learning" AND "M competitions", devido à quantidade suficiente e viável para análise, que equilibrou um bom montante de informações com objetividade na construção do estudo. Embora a Google Scholar também tenha sido usada constantemente para consulta de artigos sobre temas preliminares, fonte de citações e referências.

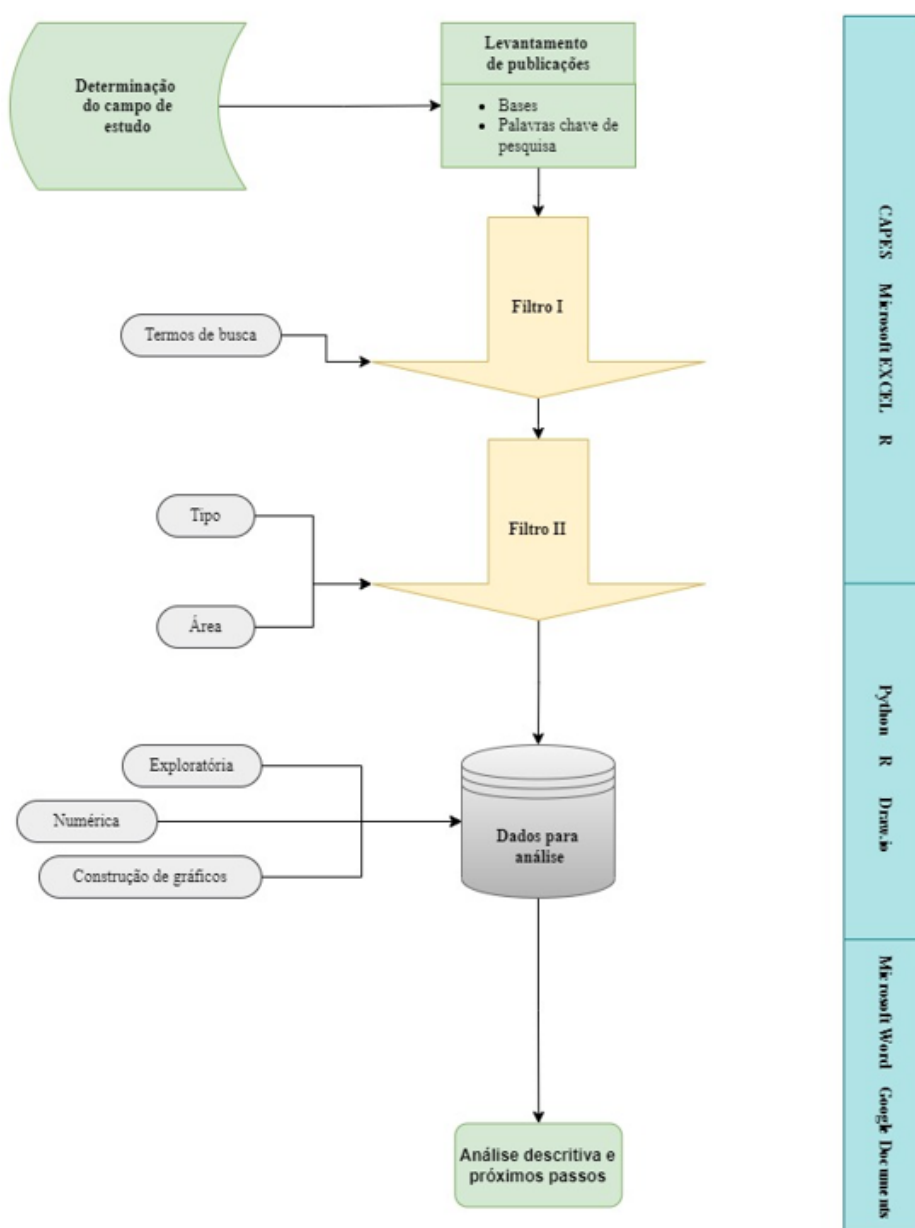


Figura 5: Fluxograma esquemático da pesquisa. Elaborado pelo autor

Definida essa pesquisa, o segundo filtro foi aplicado e os textos foram separados por tipo, foram extraídos da análise as publicações que não se tratavam de artigos de pesquisa, posteriormente os esses artigos não foram filtrados, mas sim classificados por área, são elas: *Decisions Sciences, Business, Management and Accounting, Economics, Econometrics and Finance, Computer Science, Engineering, Mathematics, Energy, Environmental Science, Agricultural and Biological Sciences* e *Neuroscience*. As últimas três foram excluídas da análise por fugir muito do tema, e após os resultados encontrados foram levantados um total final de 49 nesta pesquisa, exceto os já mencionados de análises preliminares e explicativas.

Como ferramentas inicialmente o portal CAPES e as planilhas eletrônicas do Microsoft Excel serviram como repositório de resultados e filtragem simples. As planilhas têm a vantagem de receberem qualquer tipo de formato, desde consultas em linguagem estruturada à *outputs* de programas. A linguagem R também auxiliou em algumas referências e busca autores. Após o levantamento, assim como as competições e análise de dados foi utilizado a linguagem Python para análise exploratória e construção de gráficos, toda essa etapa foi trabalhada dentro do Visual Studio Code, ambiente de desenvolvimento muito popular na ciência de dados e engenharia de software. O Draw.io, ferramenta do google, foi usado para construção do fluxograma do estudo, e as bibliotecas *pandas, matplotlib, seaborn, numpy, datetime, newspaper, collections, functools* serviram de recurso na pesquisa em Python.

5. Resultados

Na pesquisa realizada foram obtidos os seguintes resultados na tabela 1, as palavras selecionadas retornaram grandes valores de resultado no geral, mas se concentraram na ScienceDirect, enquanto o termo Machine Learning marcou presença forte em todas as bases.

Palavras chave	Bases			
	Scopus	ScienceDirect	Engineering village	Google Scholar
Sales prediction	297	59,464	39	4,140
Business forecasting	186	130,795	110	20,000
Supply planning	937	644,400	710	21,200
Machine learning	349,790	319,989	357,729	3.640,000
M competitions	146	840,000	2	545
hierarchical sales	3	42,059	2	93
Time series	263,119	3.967,407	145,043	4.240,000

Tabela 1: Pesquisa inicial com palavras e termos utilizados individualmente. Elaborado pelo autor.

Entretanto devido à grande quantidade de publicações, embora bastante informativa acerca do tema, uma filtragem através da combinações de chaves foi necessária, os resultados foram definitivos para o fechamento de quais textos seriam explorados e qual seria a base selecionada. Abaixo, algumas combinações.

Termo	Nº resultados	Base
"M competitions" AND "Business forecasting"	1	Scopus
"M competitions" AND "Business forecasting" OR "hierarchical sales"	81	ScienceDirect
"Machine learning" AND "M competitions"	87	ScienceDirect
"Machine learning" AND "M competitions"	11	Scopus
"Business forecasting" AND "supply planning" OR "supply scheduling"	2,066	ScienceDirect
"time series" AND "Business forecasting"	33	Scopus

Tabela 2: Algumas combinações. Elaborado pelo autor.

Esse recente clamor pela área pode ser visto na figura 5 logo abaixo, que lembra uma curva exponencial, em que as publicações pesquisadas se concentram nos últimos três anos, o que ilustra um maior interesse por tecnologias virtuais e de desenvolvimento remoto, devido também à pandemia de covid-19. À altura do desenvolvimento deste estudo, as publicações em metade de um ano já superam em número as publicações do mesmo tema em 2019, dois anos e meio atrás, e também já é aproximadamente o dobro de 10 anos atrás.

A discussão do tema vem evoluindo fortemente, porém devido a imensa quantidade de artigos sobre as competições e previsão de séries e a inviabilidade de um revisão de todos eles, a análise concentrada em termos específicos

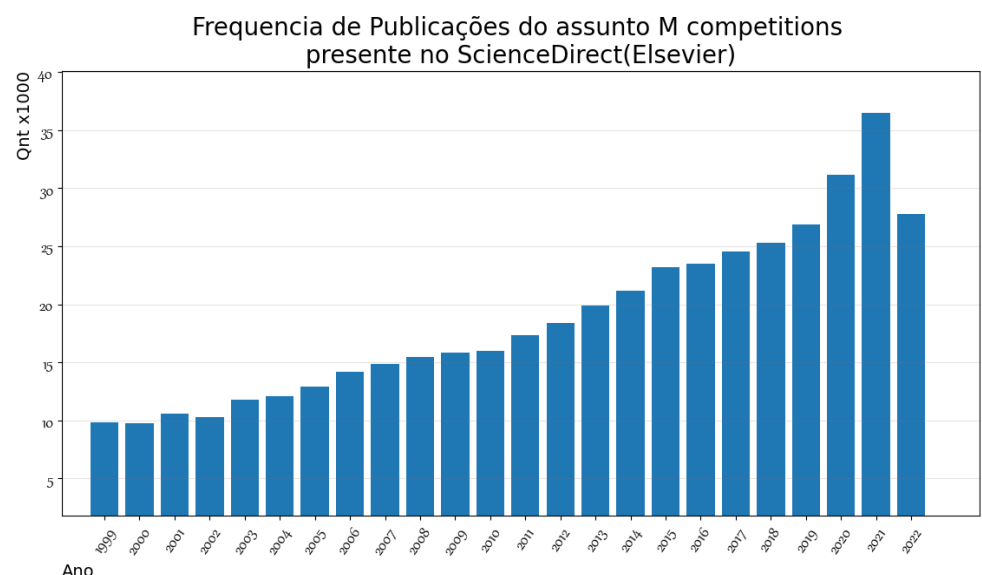


Figura 5: Frequências das publicações no repositório ScienceDirect. Elaborado pelo autor.

Dentre os textos pesquisados, como mencionado, foi dado foco para os artigos de pesquisa propriamente ditos, estes formaram a maioria dos resultados pesquisados com a correspondência "Machine learning" AND "M competitions", o tema discussão chamou atenção por compor mais da metade da parcela de *não* artigos que foi preterida.

O ranking da competição é um ponto que perde importância frente às discussões que elas geram, estes debates e publicações sobre o tema do torneio produzem fontes de aprendizado muito ricas, podem também despertar nos próprios especialistas do tema *insights* e novas ideias de abordagens. A base de dados fica disponível para os interessados no assunto, na plataforma Kaggle e além da base é possível encontrar o guia, as regras, participantes, fóruns de debates, códigos e dentre outras informações acerca da M5, por exemplo.

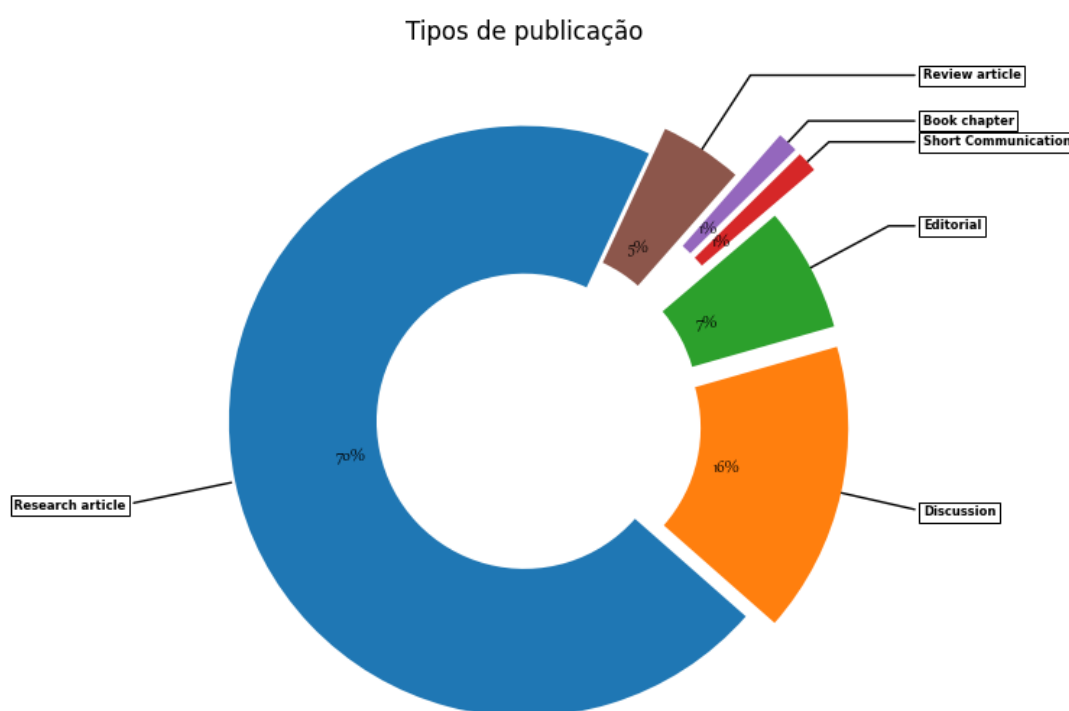


Figura 6: Gráfico de rosca sobre os tipos de publicação. Elaborado pelo autor.

Ainda baseado na ideia de que métodos sofisticados e complexos demais não necessariamente possuem maior precisão do que métodos simplórios, as duas últimas edições das competições trouxeram modelos inovadores e com ampla mescla de métodos de previsão,

estes iniciaram um movimento em torno da utilização de técnicas de aprendizado de máquinas que no começo do século bastante prematuro e mais abstrato, é chamativa a presença deste termo em especial(em inglês, Machine Learning) dentre as maiores aparições de um método no artigos revisados sobre as competições, pois algumas palavras têm frequência naturalmente maior, exemplo: *Data*, *Forecast*, *Competition*, *Methods* e etc. Abaixo é ilustrado a presença de palavras em formato de nuvem, trata-se de termos e palavras bastante conhecidos da área de dados, mas que chamam a atenção pela presença em publicações que abordaram edições diferentes das competições M e de maneiras distintas, demonstrando a importância dos temas na área desde os primeiras previsões de séries em grande escala.

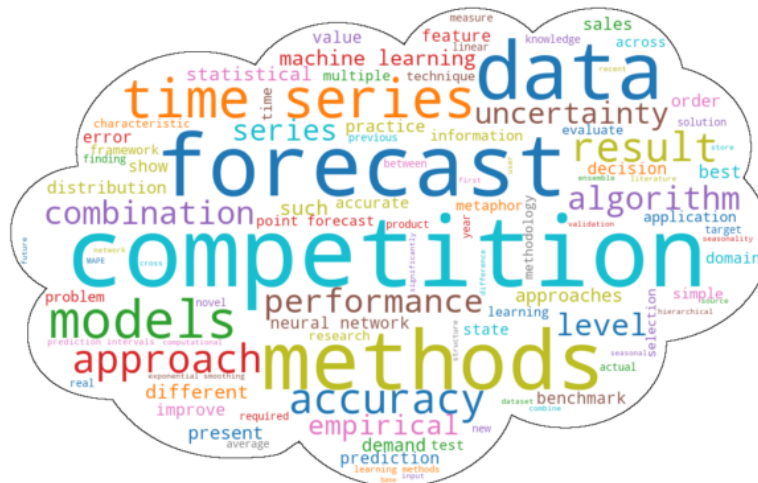


Figura 7: Nuvem de palavras. Elaborado pelo autor.

A crescente popularidade do estudo da Ciência dos dados tem trazido à tona termos como estes, diretamente ligados à computação e às redes de compartilhamento de conteúdo. Grande parte das publicações sobre os modelos das competições estão presentes nas plataformas de compartilhamento de códigos e bases de dados. Diversos modelos competidores exploram a inteligência computacional como trunfo de suas soluções, alguns algoritmos são usados não apenas para prever os dados, mas para prever o erro ou até para selecionar um método dentre vários, portanto a presença da combinação de métodos está bem estabelecida, o que torna um modelo de previsão de uma competição M , sobretudo as mais recentes, uma base para possíveis estudos posteriores dentro do mesmo tema, isto é, os modelos e suas respectivas

publicações podem ser usados como *benchmarking* de outros futuros devido suas características heterogêneas.

É evidente a presença do tema com grande força nos últimos tempos, ainda em menção à tabela 1, muitas dessas publicações servem como ponto de partida para outros estudos. Os artigos de Makridakis, fundador das competições M, estão constantemente em pauta e entre os 10 artigos mais citados da pesquisa, em metade houve sua participação.

Título	Ano	Autor	Nº Citações
<i>The M4 Competition: Results, findings, conclusion and way forward</i>	2018	Spyros Makridakis, Evangelos Spiliotis, Vassilios Assimakopoulos	443
<i>The M4 Competition: 100,000 time series and 61 forecasting methods</i>	2019	Spyros Makridakis, Evangelos Spiliotis, Vassilios Assimakopoulos	414
<i>To combine or not to combine: selecting among forecasts and their combinations</i>	2005	Miche'le Hibon, Theodoros Evgeniou	372
<i>Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction</i>	2011	Sven F. Crone, Michèle Hibon, Konstantinos Nikolopoulos	305
<i>M5 accuracy competition: Results, findings, and conclusions</i>	2021	Spyros Makridakis, Evangelos Spiliotis, Vassilios Assimakopoulos	119
<i>Exploring the sources of uncertainty: Why does bagging for time series forecasting work?</i>	2018	Fotios Petropoulos, Rob J. Hyndman, Christoph Bergmeir	88
<i>Kaggle forecasting competitions: An overlooked learning opportunity</i>	2020	Casper Solheim Bojer, Jens Peder Meldgaard	76
<i>The M5 uncertainty competition: Results, findings and conclusions</i>	2021	Spyros Makridakis, Evangelos Spiliotis, Vassilios Assimakopoulos, Zhi Chen, Anil Gaba, Ilia Tsetlin, Robert L. Winkler	35
<i>A novel ranking procedure for forecasting approaches using Data Envelopment Analysis</i>	2016	Ali Emrouznejad, Bahman Rostami-Tabar, Konstantinos Petridis	30
<i>The M5 competition: Background, organization, and implementation</i>	2021	Spyros Makridakis, Evangelos Spiliotis, Vassilios Assimakopoulos	23

Tabela 3: Lista dos 10 artigos relacionados às competições M mais citados de todo levantamento. Elaborado pelo autor.

Um bom exemplo de pesquisas derivadas das competições que discutem métodos e propõe mudança em alguns, é o estudo de HIBON e EVGENIOU(2005), que seguindo o raciocínio de

combinações de modelos traz um estudo com as séries do M3 que explora modelos combinados contra modelos individuais, o interessante é que não apresentam desempenhos muito diferentes, porém demonstram em quatro hipóteses que os métodos combinados têm a vantagem de oferecer precisão e risco menos variáveis ao longo da série em comparação com os individuais. Bem como o artigo de SPILIOTIS et al. (2021) que embora não figure na tabela 2, avalia métodos do M5 na intenção de aplicá-los a problemas de cadeia de suprimentos que lidam com muita incerteza e necessitam de previsões apuradas para operar com menor custo possível. O estudo utiliza uma comparação direta entre os métodos de previsão do M5 com abordagens estatísticas e de aprendizado de máquinas mais sofisticadas. Resultados indicam que os modelos devem ser ajustados para cada situação específica, como para qual quantil de interesse deseja-se a previsão ou os aspectos da série, métodos mais rápidos costumam desempenhar melhor do que os sofisticados, um texto que certamente contribui para a ponte entre academia e mercado com métodos de muito potencial.

Outros textos de bastante relevância são os que abordam as duas competições realizadas na M5, de acurácia e de incerteza, ambos por MAKRIDAKIS et al. (2020, 2021), com autores em comum mas sobre diferentes tipos de medições. A competição de acurácia do M5 consistiu na submissão de 30,490 pontos de previsão para dados das vendas do Walmart e o texto apresenta os resultados do desempenho dos modelos no quesito acurácia, discute as implicações que esta competição gerou e pode gerar. O segundo texto da competição debate a incerteza associada aos métodos de previsão e como prevê-la para a distribuição correta que representa 42840 registros de vendas unitárias; A competição de incerteza normalmente requer um conhecimento estatístico mais profundo para implementação e combinação de distribuições consagradas enquanto que a competição de acurácia tende a favorecer métodos de aprendizado de máquinas híbridos que têm grande poder de fogo devido a adequação mais suave aos dados.

Um exemplo de texto que aborda as competições do Kaggle de forma geral é o de BOJER e MELDGAARD (2021), que utiliza abordagens diferentes de validação e expõe os efeitos das características dos dados na árvores de decisão, debate o cross-learning e como os métodos tipicamente estatísticos tiveram muitos sucesso em competições passadas, mas atualmente são os algoritmos de *gradient-boosting* e de redes neurais que dominam o mundo das previsões, isso se deve ao fato da presença de mais *features* nas bases, isto é, mais informações e variáveis que possuem peso na atividade que se quer prever. Embora a presente análise foque esforços nas competições M5, o texto impele muitos membros da

comunidade a dar mais atenção à plataforma como ferramenta de aprendizado e rica fonte de dados.

São alguns exemplos do alcance que as competições de séries temporais possuem, movimentam a comunidade cientista de dados e programadora em torno de novos desafios e implementações mais eficientes. Um evento desta envergadura em qualquer área acadêmica tem bastante prestígio justamente por fomentar melhorias contínuas em seus estudos, também por produzir material de qualidade com forte amparo técnico e presença das principais tecnologias de ponta para resolução de problemas. Estes aspectos têm ainda mais peso por se tratar da área de engenharia e computação, que têm visto seu horizonte de alcance cada vez mais extenso.

6. Conclusão

A comparação dos dados com o petróleo é bastante coerente, como no texto de HIRSCH(2013), em que ele demonstra a usabilidade de dados na previsão e prevenção de crises econômicas, pandemias e dados demográficos em geral, as informações no nível disponível hoje em dia tem potencial inimaginável, porém foram os métodos de análise dos dados que sofreram a maior transformação nos últimos tempos, a própria competição M5 contando com mais de vinte mil submissões de diversos tipos de combinações de diversos modelos foi um marco positivo para a área. Essa vasta disponibilidade de métodos em questão de número e de acessibilidade tem agregado bastante o setor industrial de fato com previsões adequadas à sua realidade que podem ser moldadas e alteradas conforme desejado, como o modelo de inteligência computacional apresentados por Mihaela Oprea(2020), em que pode ser capaz de resolver problemas sismológicos, de energia e hidrologia, e que visa ser implementado com ferramenta de algum software para ser amplamente utilizado em previsões da cadeia de suprimentos e infraestrutura.

A *feature engineering* é bastante explorada e desenvolvida em competições, é uma forma de entender e processar dados de maneira eficiente e que produza um modelo fiel ao cenário, as variáveis de interesse e explicativas são parte delicadas da construção de um método preditivo e tem recebido grande aporte não apenas dos torneios, como cada vez mais fala-se no tema nas comunidades acadêmicas e da internet, pois muito pode ser feito com um bom encaixe entre dados e modelo. YeonJun In et al. (2018) expõem isso com seu modelo *LightGBM* voltado

à feature engineering em detrimento de um pouco de acurácia, mas ganha em velocidade de processamento e performa à ponto de figurar no topo ranking de acurácia.

É visível que quaisquer que sejam os dados hoje em dia, existe um tipo de análise capaz de extrair conclusões, ou no caso de séries temporais, prevê-las, a cada dia mais pessoas interagem por meio de competições de previsões de dados e debates sobre o assunto, as competições de Makridakis foram de suma importância para o desenvolvimento da área e de novas competições, a previsão de dados pode ser considerada como comum nos próximos anos devido à uma comunidade vasta que agrega entre profissionais e interessados de outras áreas do conhecimento, que não as ciências exatas. A área tem sido uma das mais prestigiadas no mercado por parte das empresas e dos candidatos e isso é fruto de uma conectividade e compartilhamento de estudos que as competições M fomentaram, juntamente com a principal matéria prima do nosso tempo: os Dados.

7. Referências

EHLERS, Ricardo S. Análise de séries temporais. **Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná**, v. 1, p. 1-118, 2007.

BOJER, Casper Solheim; MELDGAARD, Jens Peder. Kaggle forecasting competitions: An overlooked learning opportunity. **International Journal of Forecasting**, v. 37, n. 2, p. 587-603, 2021.

LIU, Siyao et al. Self-similar behaviors in the crude oil market. **Energy**, v. 211, p. 118682, 2020.

NEWBOLD, Paul; GRANGER, Clive WJ. Experience with forecasting univariate time series and the combination of forecasts. **Journal of the Royal Statistical Society: Series A (General)**, 137.2: 131-146, 1974.

MAKRIDAKIS, Spyros; HIBON, Michele; and MOSER, Claus. Accuracy of forecasting: An empirical investigation. **Journal of the Royal Statistical Society. Series A (General)**, 142(2):97–145, 1979.

MAKRIDAKIS, Spyros et al. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. **Journal of Forecasting**, 1(2):111–153, 1982

ASSIMAKOPOULOS, Vassilis; NIKOLOPOULOS, Konstantinos. The theta model: a decomposition approach to forecasting. **International journal of forecasting**, v. 16, n. 4, p. 521-530, 2000.

MAKRIDAKIS, Spyros et al. The M2-competition: A real-time judgmentally based forecasting study. **International Journal of forecasting**, 9.1: 5-22, 1993.

MAKRIDAKIS, Spyros; SPILIOTIS, Evangelos; ASSIMAKOPOULOS, Vassilios. The M4 Competition: Results, findings, conclusion and way forward. **International Journal of Forecasting**, v. 34, n. 4, p. 802-808, 2018.

MAKRIDAKIS, Spyros; SPILIOTIS, Evangelos; ASSIMAKOPOULOS, Vassilios. The M5 competition: Background, organization, and implementation. **International Journal of Forecasting**, 2021.

HIBON, Michele; EVGENIOU, Theodoros. To combine or not to combine: selecting among forecasts and their combinations. **International journal of forecasting**, 21.1: 15-24, 2005.

SPILIOTIS, Evangelos et al. Product sales probabilistic forecasting: An empirical evaluation using the M5 competition data. **International Journal of Production Economics**, v. 240, p. 108237, 2021.

MAKRIDAKIS, Spyros; SPILIOTIS, Evangelos; ASSIMAKOPOULOS, Vassilios. M5 accuracy competition: Results, findings, and conclusions. **International Journal of Forecasting**, 2022.

MAKRIDAKIS, Spyros; SPILIOTIS, Evangelos; ASSIMAKOPOULOS, Vassilios. The M5 Uncertainty competition: Results, findings and conclusions. **International Journal of Forecasting**, 2021.

HIRSCH, Dennis D. The glass house effect: Big Data, the new oil, and the power of analogy. **Me. L. Rev.**, 66: 373, 2013.

OPREA, Mihaela. A general framework and guidelines for benchmarking computational intelligence algorithms applied to forecasting problems derived from an application domain-oriented survey. **Applied Soft Computing**, v. 89, p. 106103, 2020.

IN, YeonJun; JUNG, Jae-Yoon. Simple averaging of direct and recursive forecasts via partial pooling using machine learning. **International Journal of Forecasting**, 2021.

HASELBECK, Florian et al. Machine Learning Outperforms Classical Forecasting on Horticultural Sales Predictions. **Machine Learning with Applications**, v. 7, p. 100239, 2022.

ORD, J. Keith. The uncertainty track: Machine learning, statistical modeling, synthesis. **International Journal of Forecasting**, 2021.

MAKRIDAKIS, Spyros; SPILIOTIS, Evangelos; ASSIMAKOPOULOS, Vassilios. The M4 Competition: 100,000 time series and 61 forecasting methods. **International Journal of Forecasting**, v. 36, n. 1, p. 54-74, 2020.

HIBON, Michele; EVGENIOU, Theodoros. To combine or not to combine: selecting among forecasts and their combinations. **International journal of forecasting**, v. 21, n. 1, p. 15-24, 2005.

CRONE, Sven F.; HIBON, Michele; NIKOLOPOULOS, Konstantinos. Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. **International Journal of forecasting**, v. 27, n. 3, p. 635-660, 2011.

PETROPOULOS, Fotios; HYNDMAN, Rob J.; BERGMEIR, Christoph. Exploring the sources of uncertainty: Why does bagging for time series forecasting work?. **European Journal of Operational Research**, v. 268, n. 2, p. 545-554, 2018.

WELLENS, Arnoud P.; UDENIO, Maxi; BOUTE, Robert N. Transfer learning for hierarchical forecasting: Reducing computational efforts of M5 winning methods. **International Journal of Forecasting**, 2021.

EMROUZNEJAD, Ali; ROSTAMI-TABAR, Bahman; PETRIDIS, Konstantinos. A novel ranking procedure for forecasting approaches using Data Envelopment Analysis. **Technological Forecasting and Social Change**, v. 111, p. 235-243, 2016.

MAKRIDAKIS, Spyros; SPILIOTIS, Evangelos; ASSIMAKOPOULOS, Vassilios. Predicting/hypothesizing the findings of the M5 competition. **International Journal of Forecasting**, 2021.

LI, Xixi; BAI, Yun; KANG, Yanfei. Exploring the social influence of the Kaggle virtual community on the M5 competition. **International Journal of Forecasting**, 2021.

MAKRIDAKIS, Spyros; HYNDMAN, Rob J.; PETROPOULOS, Fotios. Forecasting in social settings: The state of the art. **International Journal of Forecasting**, v. 36, n. 1, p. 15-28, 2020.

MA, Shaohui; FILDES, Robert. Forecasting third-party mobile payments with implications for customer flow prediction. **International Journal of Forecasting**, v. 36, n. 3, p. 739-760, 2020.

MEIRA, Erick; OLIVEIRA, Fernando Luiz Cyrino; JEON, Jooyoung. Treating and Pruning: New approaches to forecasting model selection and combination using prediction intervals. **International Journal of Forecasting**, v. 37, n. 2, p. 547-568, 2021.

DE REZENDE, Rafael et al. A white-boxed ISSM approach to estimate uncertainty distributions of Walmart sales. **International Journal of Forecasting**, 2021.

JAGANATHAN, Srihari; PRAKASH, P. K. S. A combination-based forecasting method for the M4-competition. **International Journal of Forecasting**, v. 36, n. 1, p. 98-104, 2020.

TALAGALA, Thiyanga S.; LI, Feng; KANG, Yanfei. FFORMPP: Feature-based forecast model performance prediction. **International Journal of Forecasting**, v. 38, n. 3, p. 920-943, 2022.

TALEB, Nassim Nicholas. On the statistical differences between binary forecasts and real-world payoffs. **International Journal of Forecasting**, v. 36, n. 4, p. 1228-1240, 2020.

SPILIOTIS, Evangelos et al. Product sales probabilistic forecasting: An empirical evaluation using the M5 competition data. **International Journal of Production Economics**, v. 240, p. 108237, 2021.

SPILIOTIS, Evangelos; NIKOLOPOULOS, Konstantinos; ASSIMAKOPOULOS, Vassilios. Tales from tails: On the empirical distributions of forecasting errors and their implication to risk. **International Journal of Forecasting**, v. 35, n. 2, p. 687-698, 2019.

PRAK, Dennis; ROGETZER, Patricia. Timing intermittent demand with time-varying order-up-to levels. **European Journal of Operational Research**, 2022.

JEON, Yunho; SEONG, Sihyeon. Robust recurrent network model for intermittent time-series forecasting. **International Journal of Forecasting**, 2021.

MA, Shaohui; FILDES, Robert. The performance of the global bottom-up approach in the M5 accuracy competition: A robustness check. **International Journal of Forecasting**, 2021.

HEWAMALAGE, Hansika; BERGMEIR, Christoph; BANDARA, Kasun. Recurrent neural networks for time series forecasting: Current status and future directions. **International Journal of Forecasting**, v. 37, n. 1, p. 388-427, 2021.

KIM, Sungil; KIM, Heeyoung. A new metric of absolute percentage error for intermittent demand forecasts. **International Journal of Forecasting**, v. 32, n. 3, p. 669-679, 2016.

LAINER, A. David; WOLFINGER, Russell D. Forecasting with gradient boosted trees: augmentation, tuning, and cross-validation strategies: Winning solution to the M5 Uncertainty competition. **International Journal of Forecasting**, 2022.

LOLLI, Francesco et al. Single-hidden layer neural networks for forecasting intermittent demand. **International Journal of Production Economics**, v. 183, p. 116-128, 2017.

NASIOS, Ioannis; VOGKLIS, Konstantinos. Blending gradient boosted trees and neural networks for point and probabilistic forecasting of hierarchical time series. **International Journal of Forecasting**, 2022.

CHEN, Zhi et al. Evaluating quantile forecasts in the M5 uncertainty competition. **International Journal of Forecasting**, 2022.

WILSON, Tom; GROSSMAN, Irina; TEMPLE, Jeromey. Evaluation of the best M4 competition methods for small area population forecasting. **International Journal of Forecasting**, 2021.

ANDRAWIS, Robert R.; ATIYA, Amir F.; EL-SHISHINY, Hisham. Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition. **International journal of forecasting**, v. 27, n. 3, p. 672-688, 2011.

ALROOMI, Azzam et al. Fathoming empirical forecasting competitions' winners. **International Journal of Forecasting**, 2022.

KANG, Yanfei et al. Déjà vu: A data-centric forecasting approach through time series cross-similarity. **Journal of Business Research**, v. 132, p. 719-731, 2021.

BEKIROGLU, Korkut; GULAY, Emrah; DURU, Okan. A multi-method forecasting algorithm: Linear unbiased estimation of combine forecast. **Knowledge-Based Systems**, v. 239, p. 107990, 2022.

SPILIOTIS, Evangelos et al. Hierarchical forecast reconciliation with machine learning. **Applied Soft Computing**, v. 112, p. 107756, 2021.

WANG, Xiaozhe; SMITH-MILES, Kate; HYNDMAN, Rob. Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series. **Neurocomputing**, v. 72, n. 10-12, p. 2581-2594, 2009.

MARSDEN, Chris; MEYER, Trisha; BROWN, Ian. Platform values and democratic elections: How can the law regulate digital disinformation?. **Computer Law & Security Review**, v. 36, p. 105373, 2020.

SUJJAVIRIYASUP, Thoranin. A new class of MODWT-SVM-DE hybrid model emphasizing on simplification structure in data pre-processing: A case study of annual electricity consumptions. **Applied Soft Computing**, v. 54, p. 150-163, 2017.

GAO, Ruobin et al. Time series forecasting based on echo state network and empirical wavelet transformation. **Applied Soft Computing**, v. 102, p. 107111, 2021.

SU, Chang; HUANG, Shuman; CHEN, Yijiang. Automatic detection and interpretation of nominal metaphor based on the theory of meaning. **Neurocomputing**, v. 219, p. 300-311, 2017.

PRESTWICH, S. D. Tuning forecasting algorithms for Black Swans. **IFAC-PapersOnLine**, v. 52, n. 13, p. 1496-1501, 2019.

BEKIROGLU, Korkut et al. Predictive analytics of crude oil prices by utilizing the intelligent model search engine. **Applied Energy**, v. 228, p. 2387-2397, 2018.

HYNDMAN, Rob J. A brief history of forecasting competitions. **International Journal of Forecasting**, v. 36, n. 1, p. 7-14, 2020.

BEZERRA, Manoel Ivanildo Silvestre. Apostila de Análise de séries temporais. **UNESP: Curso de estatística. São Paulo, SP**, 2006.

LEMKE, Christiane; GABRYS, Bogdan. Meta-learning for time series forecasting in the NN GC1 competition. In: **International Conference on Fuzzy Systems**. IEEE, p. 1-5, 2010.

Figura 1: Fonte dos dados: Corona virus Brazil. Kaggle. Disponível em: <https://www.kaggle.com/datasets/unanimad/corona-virus-brazil>. Acesso em: 08/07/2022.

Figura 2: Fonte dos dados: Brent Oil Prices. Kaggle. Disponível em: <https://www.kaggle.com/datasets/mabusalah/brent-oil-prices>. Acesso em: 09/07/2022.

Figura 2: SHRIVASTAVA, Soumya. Cross validation in time series. Medium, 2018. Disponível em: <https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4>. Acesso em: 09/07/2022.

Figura 4: Fonte dos dados: BISSYANDÉ, Tegawendé F. et al. Popularity, interoperability, and impact of programming languages in 100,000 open source projects. In: **2013 IEEE 37th annual computer software and applications conference**. IEEE, 2013. p. 303-312.