

SÉRIES TEMPORAIS E VARIÁVEIS QUALITATIVAS: REVISÃO SISTEMÁTICA DE LITERATURA¹

TIME SERIES AND QUALITATIVE VARIABLES: SYSTEMATIC LITERATURE REVIEW

Viviane de Senna* E-mail: vivianedsenna@hotmail.com
Adriano Mendonça Souza* E-mail: amsouza.sm@gmail.com
Renan Mitsuo Ueda* E-mail: renan.mitsuo@hotmail.com

Universidade Federal de Santa Maria (UFSM), Santa Maria, Rio Grande do Sul (RS), Brasil.

Resumo: As variáveis qualitativas possuem características de influenciar variáveis quantitativas. As séries temporais são exemplo de variáveis quantitativas que podem ser modificadas por alterações abruptas de variáveis qualitativas. O objetivo desta pesquisa é identificar uma lacuna de conhecimento entre modelos de previsão de séries temporais e variáveis qualitativas por meio de uma revisão sistemática de literatura aplicada em artigos científicos publicados em bases internacionais. Foram encontrados 37 artigos que unem séries temporais e variáveis qualitativas no período de 1985 a 2020, nos quais novos modelos foram criados utilizaram as variáveis qualitativas como entradas de informação para a geração de previsões. A lacuna encontrada foi em aproximados 57% dos casos a capacidade de melhoria dos modelos criados e a aplicação em outros sistemas, e em 19% a possibilidade de melhorar a previsão de outras variáveis do sistema. Os resultados indicaram que a inclusão de variáveis qualitativas em modelos quantitativos auxilia na qualidade da precisão das previsões geradas.

Palavras-chave: Séries Temporais. Variáveis Qualitativas. Modelos de Previsão. Revisão Sistemática. Modelagem.

Abstract: Qualitative variables have characteristics of influencing quantitative variables. Time series are an example of quantitative variables that can be modified by abrupt changes in qualitative variables. The objective of this research is to identify a knowledge gap between time series forecasting models and qualitative variables through a systematic review of applied literature in scientific articles published in international databases. We found 37 articles that unite time series and qualitative variables in the period from 1985 to 2020, in which new models were created that used qualitative variables as information inputs for the generation of forecasts. The gap found was in approximately 57% of cases the ability to improve the models created and their application in other systems and in 19% the possibility of improving the prediction of other system variables. The results indicated that the inclusion of qualitative variables in quantitative models helps in the quality of the precision of the generated forecasts.

Keywords: Time Series. Qualitative Variables. Forecast Models. Systematic Review. Modeling.

¹ o texto é inédito, não foi financiado, e é parte integrante de tese de doutorado em andamento.

1 INTRODUÇÃO

Fatos, acontecimentos, eventos, opiniões podem influenciar as pessoas e o mercado econômico, mas a capacidade de interferências que variáveis qualitativas possuem é difícil de mensurar. As variáveis qualitativas ou categóricas descrevem características nominais ou ordinais não quantificáveis. São oriundas de uma pesquisa não estruturada com o objetivo de alcançar um entendimento das razões e motivações subjacentes que resultam em uma compreensão inicial (MALHOTRA, 2006).

Séries temporais são coleções de observações feitas sequencialmente ao longo do tempo aplicadas aos métodos econométricos utilizadas para entender as relações entre variáveis econômicas. As séries podem sofrer influência de fontes distintas, como de variáveis qualitativas. Modelos como alisamento exponencial simples, linear de Brow de Holt-Winters, são automáticos, ou seja, não exigem intervenção para a aplicação. Os modelos não automáticos exigem intervenção para decidir qual o mais adequado para produzir a melhor previsão, como os métodos Box-Jenkins (MORETTIN; TOLOI, 2004). Existem métodos apropriados para modelar séries temporais e pesquisas que visam integrá-los com as variáveis qualitativas.

Para verificar como os estudos colaboram para o desenvolvimento do conhecimento científico sobre séries temporais e variáveis qualitativas a revisão sistemática de literatura é o método adequado. Pela revisão sistemática é possível analisar e sintetizar informações de estudos existentes, gerar novos resultados (THOMÉ; SCAVARDA; SCAVARDA, 2016) e encontrar lacunas a serem desenvolvidas. O problema a ser resolvido é qual o conhecimento gerado que integra variáveis qualitativas e modelos de séries temporais?

A importância de conhecer os estudos que já foram efetuados com relação às modelagens é na identificação de como as variáveis qualitativas foram tratadas nos modelos, os tipos de modelos quantitativos explorados nos procedimentos de união de variáveis e, a partir do conhecimento do que já foi proposto é possível avançar no incremento de modelos precisos, complexos e de fácil aplicabilidade. A identificação de autores que já são considerados referência por suas pesquisas acadêmicas desenvolvidas e citadas potencializa a qualidade de novas pesquisas sobre a

temática, bem como serve de impulsionador para o incremento de novos conhecimentos a partir do que já foi criado e testado.

O objetivo do estudo é identificar uma lacuna de conhecimento entre modelos de previsão de séries temporais e variáveis qualitativas através de uma revisão sistemática de literatura aplicada em artigos científicos publicados em bases internacionais. Identificar como o conhecimento sobre determinada temática evoluiu pode contribuir para o desenvolvimento de pesquisas distintas das já efetuadas. A ciência avança em velocidade rápida acompanhando a tecnologia, por isso se faz necessário seguir essa evolução para que a pesquisa seja colaborativa, aplicada a realidade e atinja a expectativa condizente com meio vivenciado. O uso de modelos de séries temporais para a geração de previsão de séries econômicas serve de embasamento para a tomada de decisões de gestores em diversas áreas mercadológicas, que são influenciadas por variáveis categóricas inesperadas com a capacidade de alterar o cenário vigente.

O estudo está estruturado em três tópicos, esta introdução, no tópico 2 a metodologia utilizada, os resultados e as discussões, no tópico 3 as considerações finais e, por fim as referências.

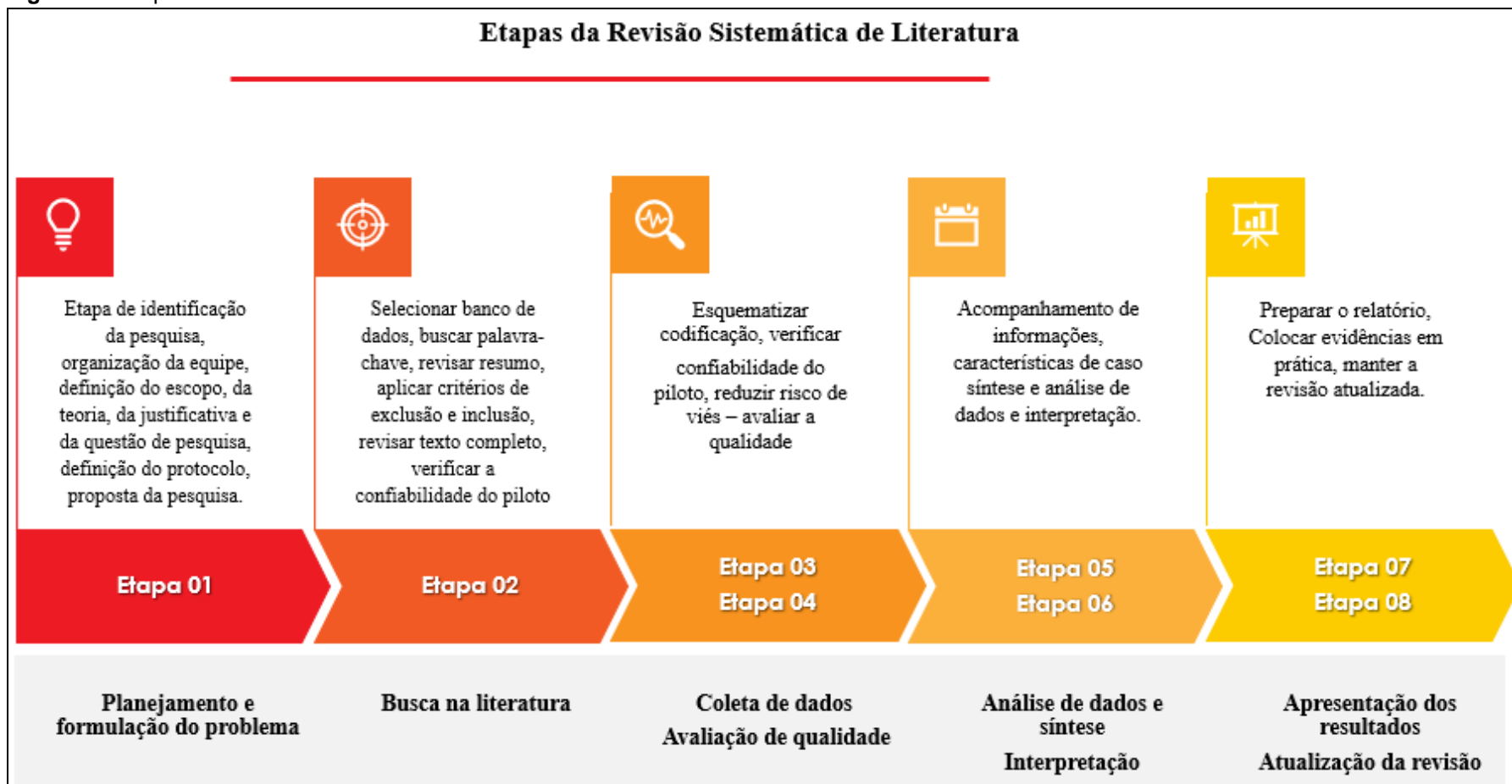
2 DESENVOLVIMENTO

2.1 Metodologia

A revisão sistemática foi baseada no protocolo de Thomé, Scavarda e Scavarda (2016) que foi desenvolvido para efetuar revisões sistemáticas na área de gestão de operações. Foi embasado no protocolo de Cochrane, amplamente utilizado na medicina e na psicologia (COLLABORATION, 2014; HIGGINS; GREEN, 2008).

O protocolo prevê oito etapas que tem por função tentar reduzir os vieses da pesquisa (THOMÉ, SCAVARDA, SCAVARDA, 2016). Na Figura 1 são apresentadas mais informações sobre essas etapas.

Figura 1 – Etapas do método de Revisão Sistemática de Literatura



Fonte: Adaptado de Thomé; Scavarda; Scavarda (2016).

A Etapa 1 descrita na Figura 1, consta a Introdução anterior à metodologia. A Etapa 2 se refere a seleção do banco de dados fonte da pesquisa, as bases foram definidas pelos critérios rigorosos para a indexação das revistas e por serem internacionais de ampla abrangência. A pesquisa foi realizada na *Scopus – Elsevier* e *Web of Science*, devido a quantidade e diversidade de artigos vinculados e pela compatibilidade da exportação dos resultados com o *software* livre RStudio.

A data da pesquisa foi 21/01/2022, o período de coleta das informações das bases foi aberto. A *String* foi confeccionada com as palavras ("*qualitative variable*" AND "*time series*") OR ("*qualitative variable*" AND *forecast**) OR ("*qualitative variable*" AND "*Box*Jenkins*") OR ("*qualitative variable*" AND ARIMA). A *String* busca as palavras nos títulos, palavras-chave e resumos dos artigos. Como critério de inclusão, foram selecionados artigos científicos revisados por pares, excluídos os livros e relatórios e periodicidade livre.

Efetuada a coleta, conforme a Etapa 3, para a análise utilizou-se o *software* RStudio o pacote Bibliometrix (ARIA; CUCCURULLO, 2017). A Etapa 4 inicia com a leitura inicial do título e resumo, os artigos mantidos relacionam de alguma forma as séries temporais e variáveis qualitativas, foram excluídos os duplicados na coleta. Após a leitura dos artigos foram coletados o objetivo do trabalho, métodos e modelos utilizados ou propostos e os resultados encontrados. As Etapas 5 a 8 são desenvolvidas a partir dos resultados e discussões.

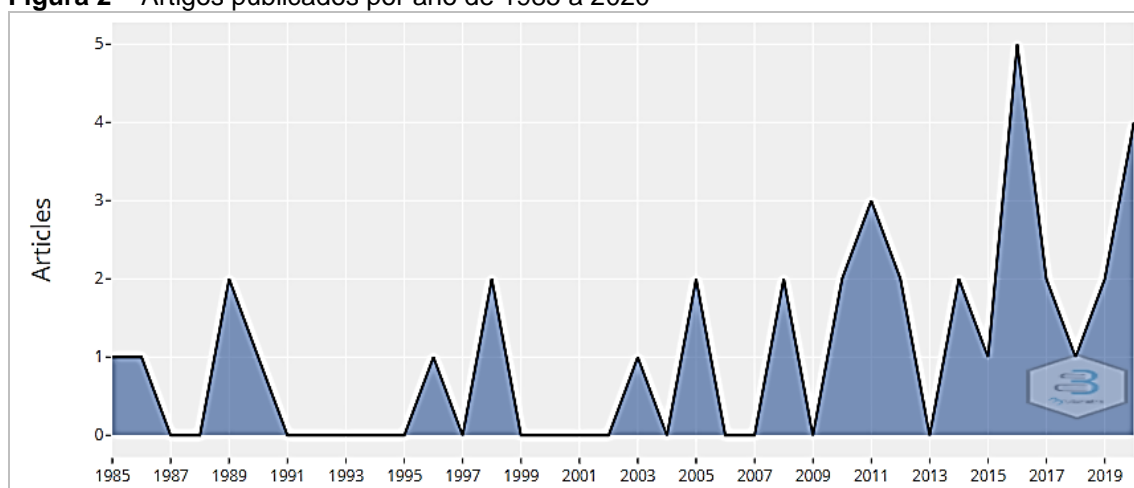
2.2 Resultados e Discussões

O protocolo (THOMÉ; SCAVARDA; SCAVARDA, 2016) define oito etapas, a Etapa 1 foi atendida durante a definição do escopo do estudo, planejamento do objetivo, problema de pesquisa e metodologia. A segunda Etapa é contemplada no tópico 2 em que são abordados as variáveis qualitativas e os modelos de séries temporais. Na etapa três foi efetuada a seleção dos bancos de dados *Scopus* e *Web of Science*, a coleta e exportação dos resultados da pesquisa da *String*. As bases possuem dois tipos de palavras-chave, as definidas como autor, em que são selecionadas as palavras definidas pelos autores dos artigos e as palavras-chave *Plus* que são definidas pelas bases.

O resultado da coleta identificou 37 artigos na *Scopus* e 6 na *Web of Science*, que já estavam contidos no resultado da *Scopus*. Os 37 artigos foram escritos por 104 autores, 11 possuíam um único autor responsável pela publicação e 93 publicaram em conjunto, com média de 2,8 autores por artigo. As palavras-chave dos autores somaram 188 e as Palavras-chave *Plus* somam 345. Foram identificadas 1.353 referências distintas que serviram de base para teórica para os estudos. O primeiro registro data de 1985 e os últimos foram 2020.

A etapa 4 inicia com a análise das publicações por países descrita na Figura 2.

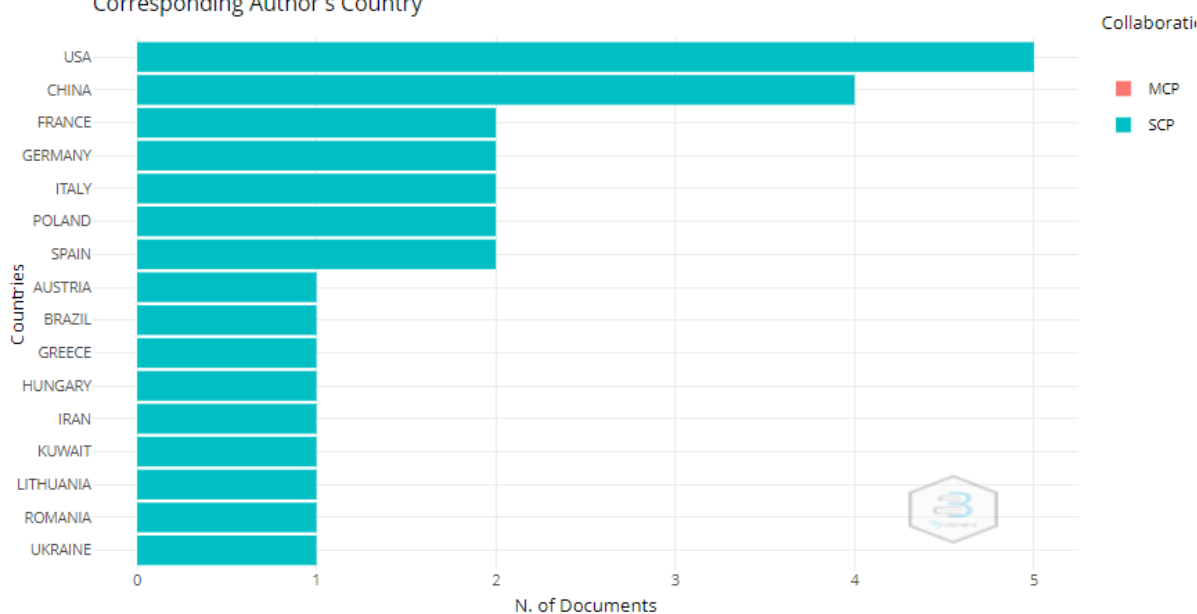
Figura 2 – Artigos publicados por ano de 1985 a 2020



Fonte: Autores (2022).

A Figura 2 indica que em 2016 foram publicados 5 artigos e em 2020 foram 4, os dois anos com maior produção. Foram indicadas publicações durante o período de 35 anos, entretanto mais de 50% delas foram divulgadas nos últimos 8 anos, ou seja, de 2012 a 2020. Entre os anos de 1991 à 1995 e 1999 à 2002 não foram feitas publicações de artigos nos periódicos indexados nas bases pesquisadas. A Figura 3 indica os países de origem dos autores correspondentes.

Figura 3 – Artigos publicados por país considerando o autor correspondente de 1985 a 2020
Corresponding Author's Country



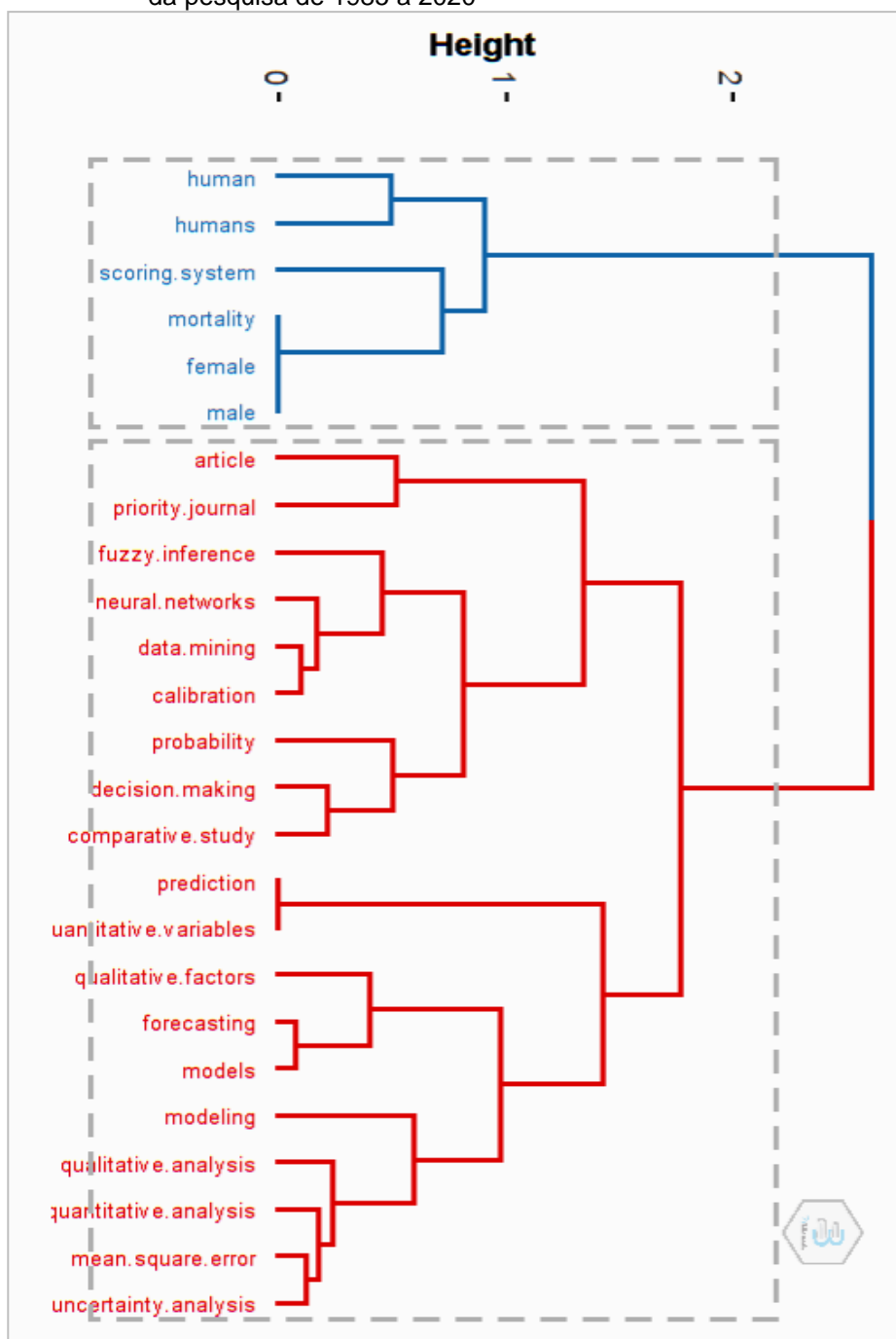
Fonte: Autores (2022).

SCP: Single Country Publications, **MCP:** Multiple Country Publication.

A Figura 3 apresenta a nacionalidade dos autores, a legenda identifica que os artigos representados em azul foram produzidos por autores do mesmo país, e em vermelho os autores pertencem a países diferentes (SCP: *Single Country Publications* e MCP: *Multiple Country Publications*). Os dois países com maior produtividade são a China e os Estados Unidos e as pesquisas são desenvolvidas por autores da mesma nacionalidade. Não foram encontradas pesquisas desenvolvidas por autores de múltiplas nacionalidades, o que indica a inexistência de contribuição de grupos de pesquisas entre os países, para os artigos coletados sobre esta temática. O Brasil foi representado por um artigo publicado.

A organização das palavras-chave Plus mais frequentes em grupos por correspondência múltipla está disposta na Figura 4, na qual é possível identificar a estrutura conceitual dos termos.

Figura 4 – Dendrograma de palavras-chave Plus dos artigos sobre a temática da pesquisa de 1985 a 2020



Fonte: Autores (2022).

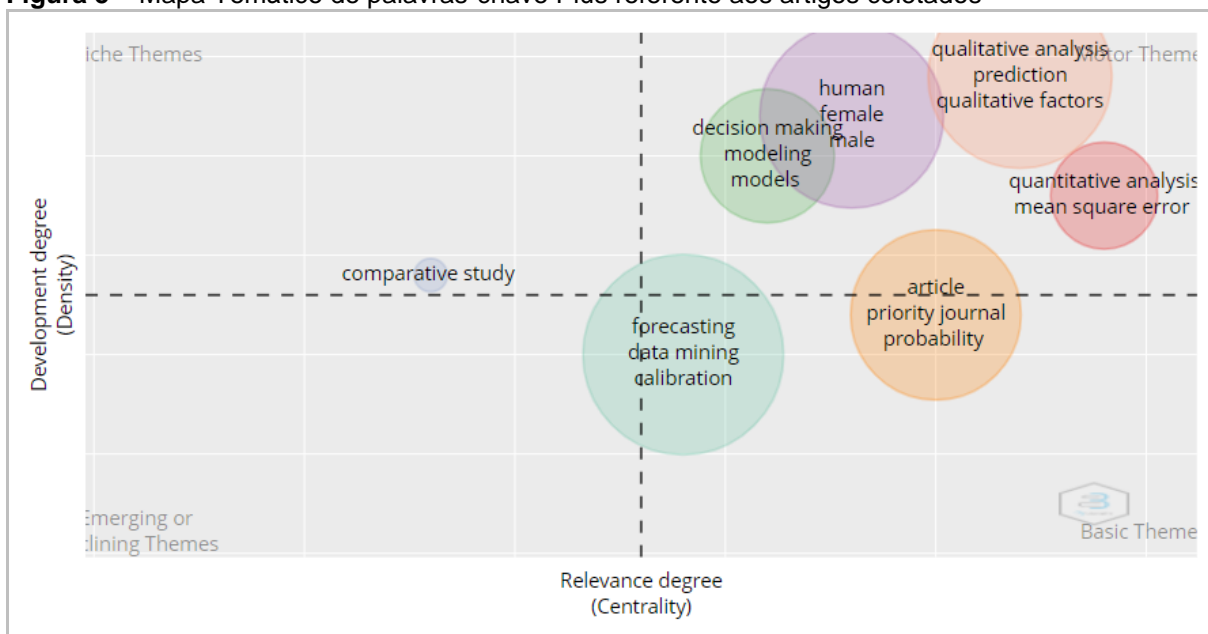
Na Figura 4 o dendrograma divide as palavras-chave *Plus* em dois agrupamentos. O agrupamento azul é denominado “*mortality*”, os termos “*female*”, “*male*” e “*mortality*” são os que possuem ligação mais forte. Em vermelho o agrupamento “*models*” é composto por termos relacionado a modelos, as duas palavras com ligação mais forte são “*prediction*” e “*quantitative variables*”. As

Revista Produção Online. Florianópolis, SC, v. 22, n. 1, p. 2349-2372, 2022.

relações mais fortes indicam pesquisas focadas em modelos de previsão, com precisão e a inserção de variáveis qualitativas, pois os termos que estão relacionados com ligações mais próximas são, na maioria sinônimos.

A densidade e centralidade das palavras são apresentadas na Figura 5.

Figura 5 – Mapa Temático de palavras-chave Plus referente aos artigos coletados



Fonte: Autores (2022).

Na Figura 5 os temas motores estão posicionados no quadrante direito superior, possuem alta densidade e forte centralidade. Destacam-se os termos “*qualitative analysis*”, “*quantitative factors*” e “*prediction*” que concordam com o *String* da pesquisa. Esse quadrante é indicativo de temáticas com alto potencial de desenvolvimento, objetos de pesquisas atuais. Os temas de importância marginal estão no quadrante superior esquerdo, são muito especializados, mas com caráter periférico, as palavras nesse quadrante são “*comparative study*”. Esse quadrante indica temas importantes, mas com enfoque secundário se comparados com o primeiro quadrante.

Os temas emergentes ou desaparecidos são localizados no quadrante inferior esquerdo, compostos por temas de baixa densidade e baixa centralidade. Neste quadrante não constam os termos para os resultados desse estudo. Os temas incluídos nesse quadrante, em geral, são os que já foram explorados e estão em queda com relação ao conjunto de artigos selecionados. Os temas transversais e

gerais, importantes, mas não desenvolvidos estão no quadrante inferior direito. Neste constam os temas como “*priority journal*” e “*probability*”. Este pode ser um indicativo da necessidade de desenvolvimento e publicação de mais estudos de modelos probabilísticos que unam as análises qualitativas e quantitativas, ou seja, os temas motores.

Posicionados sobre o eixo que divide os quadrantes inferiores estão os termos “*forecasting*”, “*data mining*” e “*calibration*”. Os temas identificados apontam para o uso de modelos quantitativos de previsão e análises qualitativas que auxiliem na tomada de decisão. Esse eixo é considerado indicador de centralidade, ou seja, isso indica que esses termos são necessários para a compreensão do cenário pesquisado. O posicionamento dos temas nos quadrantes primeiro e quarto indicam a potencialidade do desenvolvimento de pesquisas na área. Apesar de modelos apenas quantitativos bastante precisos sempre há capacidade de melhoria se observadas as mudanças que variáveis categóricas possuem de interferência nos sistemas.

A *String* da pesquisa tem como foco da pesquisa os artigos que unem variáveis qualitativas e modelos de previsão quantitativos. Os resultados indicaram a aplicação dessas palavras como ferramentas para a geração de resultados em outras áreas de conhecimento. Pela análise conjunta das Figuras 4 e 5 é possível inferir que há potencial para o desenvolvimento de modelos de previsão com a inclusão de variáveis/fatores qualitativos que auxiliem na geração de previsões mais precisas que sirvam de apoio à tomada de decisão.

Para ampliar a compreensão sobre os estudos foi efetuada a verificação dos estudos por ordem de relevância. A quantidade de citações em cada artigo pesquisado está especificada na Tabela 1, que apresenta os periódicos, o ano e DOI.

Tabela 1 – Os artigos e periódicos por relevância (continua)

Primeiro Autor	Ano	Periódico	DOI	Total de citações
Alfaro E	2008	Decis Support Syst	10.1016/j.dss.2007.12.002	180
Harvey AC	1989	J Bus Econ Stat	10.1080/07350015.1989.10509750	149
Brauers WKM	2011	Technol Econ Develop Econ	10.3846/20294913.2011.580566	86
Dueker M	2005	J Bus Econ Stat	10.1198/073500104000000613	69
Hassler U	2003	J R Stat Soc Ser D Stat	10.1111/1467-9884.00365	34

Tabela 1 – Os artigos e periódicos por relevância

(conclusão)

Primeiro Autor	Ano	Periódico	DOI	Total de citações
Figini S	2011	J Oper Res Soc	10.1057/jors.2010.41	32
Kuo RJ	2016	J Intell Manuf	10.1007/s10845-014-0944-1	31
Cao S	2017	Sci Total Environ	10.1016/j.scitotenv.2016.10.088	25
Helbok R	2005	Am J Trop Med Hyg	10.4269/ajtmh.2005.72.150	23
Christensen R	1986	Int J Gen Syst	10.1080/03081078608934938	18
Billio M	1998	J Stat Plann Inference	10.1016/S0378-3758(97)00136-5	15
Niedbala G	2019	Appl Sci	10.3390/app9142773	15
Séverin E	2010	Neurocomputing	10.1016/j.neucom.2009.12.024	13
Spooner J	1990	Acad Med	10.1097/00001888-199003000-00011	12
Pecican ES	2010	Rom J Econ Forecast	NA	10
Anagnostis A	2020	Sustainability	10.3390/SU12166409	8
Ramírez MD	1998	North Am J Econ Financ	10.1016/S1062-9408(99)80080-1	7
Jurasz J	2016	Prz Elektrotech	10.15199/48.2016.04.32	7
Annarelli A	2018	J Clean Prod	10.1016/j.jclepro.2018.07.220	6
Costa NR	2016	Cienc Saude Coletiva	10.1590/1413-812320152110.18292016	6
Ramirez MD	1996	Atl Econ J	10.1007/BF02298433	6
Sattari MT	2020	Environ Monit Assess	10.1007/s10661-020-08577-8	6
Qu S	2017	Int J Environ Res Public Health	10.3390/ijerph14111276	5
Elwakil E	2014	Can J Civ Eng	10.1139/cjce-2014-0153	5
Kokodey TA	2012	J Int Food Agribus Mark	10.1080/08974438.2012.691815	3
M'hallah R	2019	Appl Soft Comput J	10.1016/j.asoc.2018.10.009	3
Jakubowski J	2014	Arch Min Sci	10.2478/amsc-2014-0049	2
Wang XZ	2012	Transp Porous Media	10.1007/s11242-012-0037-6	2
Farhadi R	2020	Int J Food Eng	10.1515/ijfe-2019-0110	2
Victory W	2016	Intern J Eng Technol	10.21817/ijet/2016/v8i5/160805233	1
Dohnal M	2015	Int J Bifurcation Chaos	10.1142/S0218127415501734	1
Pérez-Rodríguez JV	2011	Quant Financ	10.1080/14697681003685555	1
Varga T	2008	Comput Aided Chem Eng	10.1016/S1570-7946(08)80179-4	1
Herrmann R	1989	J Agric Econ	10.1111/j.1477-9552.1989.tb01095.x	1
Alberti TS	2020	Pesqui Vet Bras	10.1590/1678-5150-PVB-6406	0
Boada A	2016	Opcion	NA	0
Borgegard LE	1985	Espace-Populations-Societes	10.3406/espos.1985.996	0
Total				785

Fonte: Autores (2022).

Os artigos apresentados na Tabela 1 estão identificados pelo sobrenome do primeiro autor. Destacam-se os dois primeiros artigos que, pelo número de citações são considerados os mais relevantes pelas bases científicas. O artigo de Alfaro et al.; (2008) foi citado em outros 180 estudos. Os autores criaram o AdaBoost, método alternativo para previsão de falências corporativas que consiste em um algoritmo de

aprendizagem de conjunto que constrói classificadores de base usando diferentes conjuntos de dados. O método foi comparado a redes neurais artificiais e aplicado em um conjunto de empresas europeias, em variáveis de previsão, como índices financeiros, e em variáveis qualitativas, como tamanho da empresa, atividade e estrutura jurídica. Os resultados indicaram que a estratégia AdaBoost, que combina árvores únicas atingiu uma redução de 30% no erro de teste em comparação com a rede neural individual, com isso as empresas falidas são diferenciadas das empresas saudáveis.

O segundo artigo mais citado, com 149 citações, foi escrito por Harvey e Fernandes (1989) que desenvolveram um modelo que, com a série em nível, gera observações que mudam ao longo do tempo. O modelo serve para dados de contagem e pode incluir variáveis planificatórias, as inclinações não estocásticas, os componentes sazonais, as análises de intervenção, as variáveis explicativas, as variáveis qualitativas, baseadas em distribuição binomial ou multinomial. A seleção do modelo segue os moldes de modelos utilizados para séries temporais estruturais gaussianas. Os testes foram feitos nos resultados dos jogos internacionais de futebol disputados entre a Inglaterra e a Escócia e uma avaliação do efeito da lei britânica do cinto de segurança nos condutores de veículos de mercadorias.

O artigo mais antigo é datado em 1985 (Borgegard Haggstrom, 1985) que descreve um modelo de micro simulação por Holm e Oberg de migração relacionada à ligação entre os níveis micro e macro. A ênfase foi descrever o modelo de migração entre a Finlândia e a Suécia desde a Segunda Guerra Mundial. O objetivo do modelo é tentar integrar variáveis quantitativas e qualitativas. Um ponto forte é que inclui mudanças sociais e não se limita a uma perspectiva de curto prazo. A desvantagem é a natureza geral do modelo e a impossibilidade de usá-lo para fins de previsão, esta pode ser a razão pela baixa relevância do estudo, em comparação com os demais.

Para identificar os métodos aplicados nos estudos foi desenvolvida a Tabela 2.

Tabela 2 – Principais resultados obtidos pela aplicação de variáveis qualitativas em modelos quantitativos de séries temporais

Modelo utilizado	Nº artigos que usaram o modelo		Autores
Criação de modelo	16	43,24%	(Dueker, 2005) (Figini Giudici, 2011) (Helbok et al.; 2005) (Billio Monfort, 1998) (Pecican, 2010) (Ramírez, 1998) (Costa et al.; 2016) (Qu Zhou, 2017) (Ramirez, 1996) (Hallah Aboukhamseen, 2019) (Sattari et al.; 2020) (Farhadi et al.; 2020) (Victory Ali Ahmed, 2016) (Varga et al.; 2008) (Herrmann, 1989) (Boada, 2016)
Rede neural	8	21,62%	(Alfaro et al.; 2008) (Kuo et al.; 2016) (Séverin, 2010) (Niedbala et al.; 2019) (Jurasz Mikulik, 2016) (Anagnostis et al.; 2020) (Jakubowski Tajduś, 2014) (Pérez-Rodríguez, 2011)
Modelo Fuzzy	4	10,81%	(Brauers et al.; 2011) (Christensen, 1986) (Elwakil Zayed, 2015) (Dohnal Doubravsky, 2015)
Análise de regressão	3	8,11%	(Cao et al.; 2017) (Kokodey, 2012) (Wang et al.; 2012)
Análise de correlação	2	5,41%	(Annarelli et al.; 2018) (Hassler Thadewald, 2003)
Revisão de literatura	2	5,41%	(Spooner, 1990) (Borgegard Haggstrom, 1985)
Análises qualitativas e quantitativas separadas	1	2,70%	(Alberti et al.; 2020)
Variável qualitativa como série exógena	1	2,70%	(Harvey Fernandes, 1989)
Total Geral	37	100,00%	

Fonte: Autores (2022).

De acordo com os resultados indicados pela Tabela 2, em 43,24%, ou seja, 16 dos estudos os autores desenvolveram novos modelos para incluir variáveis qualitativas em quantitativos. Dois estudos utilizaram as qualitativas de forma binária, para estimar um modelo de crescimento linear derivado de uma função de produção neoclássica modificada que incorpora um número de variáveis *dummy* quantitativas e qualitativas relevantes. As variáveis qualitativas foram aumentos nos gastos de investimento público e privado e movimentos positivos nos termos de troca (Ramirez, 1996), e aumentos nos gastos com investimentos nos setores público e privado (Ramírez, 1998).

Nos 14 artigos as variáveis qualitativas foram representadas por números definidos pelos autores tomando por base a expertise obtida pelo estudo das temáticas. Os números têm a função de indicar os pesos que cada variável possui em função da capacidade de influenciar no sistema de variáveis qualitativas definidas nos estudos. O destaque é análise subjetiva aplicada que pode ser contestada ou receber avaliação distinta daquela dada pelos autores se for submetida a outro especialista.

Destaca-se o modelo Qual VAR (Dueker, 2005) permite incorporar informações qualitativas ou discretas em autorregressões vetoriais, foi usado para prever recessão de negócios em 2001. O modelo permite a inclusão de variável qualitativa que pode melhorar a qualidade de previsões de densidade das demais variáveis do sistema.

Outra metodologia relevante é a fusão de dados em modelos longitudinais e de duração de sobrevivência usando variáveis quantitativas e qualitativas separadamente na função de verossimilhança e combinaram as pontuações linearmente por um peso, para obter a probabilidade de inadimplência correspondente para Pequenas e Médias Empresas (Figini Giudici, 2011).

Para quantificar a relação entre variáveis foi utilizada a correlação de Spearman, o teste U de Mann-Whitney foi usado para comparar o *multi-organ dysfunction score* – MODS, score de disfunção de múltiplos órgãos de pacientes com malária *P. falciparum* grave da população do estudo (Helbok et al.; 2005). A pontuação encontrada pode fornecer um valor preditivo para a morbidade na malária *P. falciparum*.

O modelo geral de comutação de espaço de estados, uma combinação do filtro de Kalman parcial e de técnicas de amostragem para calcular a função de verossimilhança e métodos de redução de variância com base em abordagens sequencialmente ideais (Billio Monfort, 1998). As abordagens são computacionalmente mais simples se a série for uma variável qualitativa, leva apenas um número finito de valores.

Para prever a evolução dos principais preços (preço ao consumidor, taxa de câmbio, salários, taxa de juros), foram elaboradas variantes do modelo VAR e foi mantida uma, modelo VAR aberto, que inclui um número limitado de variáveis exógenas que expressam abordagens como teoria econômica e o papel da política econômica (Pecican, 2010). As previsões de curto prazo são mais adequadas e recomenda-se a atualização constante.

Foram utilizadas redes neurais em 21,62% dos estudos, dentre os mais relevantes tem-se a rede neural difusa, utilizada para capturar o conhecimento sobre fatores qualitativos, foi combinada a série temporal por uma rede neural de retropropagação baseada no sistema imunológico artificial (Kuo et al.; 2016). A

aplicação de redes neurais com topologia MLP (*perceptron* multicamadas) originou três modelos multicritério para a previsão e simulação da produtividade do trigo de inverno, com base em variáveis quantitativas e qualitativas que podem ser usados na agricultura de grandes áreas e servir de sistema de apoio à tomada de decisão (Niedbala et al.; 2019).

O algoritmo de Kohonen aplicado em redes neurais com variáveis qualitativas que permitem questionar os modelos de pontuação permitiu uma melhor percepção dos fatores capazes de explicar o financiamento de leasing (Séverin, 2010). Anagnostis et al. (2020) exploram três tipos de abordagens de rede neural, a artificial *perceptron* simples, a de memória de longo prazo e a rede neural profunda proposta com a inclusão das variáveis qualitativas que regem o comportamento humano. Os resultados indicaram que terceira rede supera as demais.

Modelos Fuzzy compuseram a base de 10,81% do resultado. A teoria fuzzy foi usada para o método MULTIMOORA (*Multiplicative and Multi-Objective Ratio Analysis*), que consiste em três partes, Sistema de Razões, Ponto de Referência e Forma Multiplicativa Completa. As partes são sintetizadas em três conjuntos de classificação, mas permaneceu o problema de resumir os resultados com matrizes de resposta grandes (Brauers et al.; 2011). Christensen (1986) verificou as aplicações da entropia minimax abrangem padrões discretos em espaços multidimensionais de variáveis quantitativas e qualitativas mistas e padrões contínuos que empregam conceitos de funções potenciais e entropias fuzzy.

Entre os estudos que utilizaram análise de regressão destaca-se a aplicação de modelos de interpolação espacial, incluindo krigagem ordinária e krigagem de regressão com variável auxiliar para medir a melhoria na precisão do mapeamento da distribuição espacial do cádmio em solos usando métodos geoestatísticos combinados com fatores auxiliares, especialmente variáveis qualitativas (Cao et al.; 2017). A comparação dos modelos indicou que a introdução de variáveis qualitativas melhorou a precisão da previsão, e enfraqueceu o efeito de fatores quantitativos.

A análise de correlação foi empregada em 5,41% dos estudos. Para examinar se o comportamento de dois grupos heterogêneos é distinto, mesmo que o coeficiente de correlação seja aproximadamente o mesmo. A heterogeneidade dos grupos pode ser interpretada como uma variável qualitativa oculta e os resultados

podem ser quantificados a partir de uma fórmula assintótica para o coeficiente de correlação da amostra combinada (Hassler Thadewald, 2003).

Dos 34 artigos que produziram modelos três usaram as variáveis qualitativas como entradas binárias e trinta e um com valores numéricos atribuídos pelos autores. Em comparação cronológica os modelos fuzzy foram utilizados até 2015, modelos de correlação e regressão até 2018, modelos de redes neurais foram utilizados de 2008 a 2020 e, em destaque, os modelos desenvolvidos pelos autores, baseados em modelos de séries temporais que perpassam pela maior parte do período pesquisado de 1989 a 2020.

Ficam evidenciados na Tabela 2 a aplicação de outros métodos quantitativos que foram combinados com variáveis qualitativas para melhorar a precisão da previsão. Além dos métodos foram verificadas as lacunas de conhecimento geradas após a conclusão dos estudos. Na Tabela 3 estão descritas as lacunas identificadas nos artigos coletados.

Tabela 3 – Principais lacunas identificadas pelos autores dos artigos coletados (continua)

Lacunas identificadas	Nº de artigos		Autores
A metodologia foi proposta para integração de risco que pode ser generalizada para prestadores de serviços financeiros.	2	5,41%	(Figini Giudici, 2011) (Boada, 2016)
O modelo aumenta a precisão da e pode ser aplicado para estudar o desenvolvimento de transportadoras intermunicipais	1	2,70%	(Victory Ali Ahmed, 2016)
O modelo deve ser aplicado em amostras maiores para sua confirmação.	1	2,70%	(Helbok et al.; 2005)
O modelo pode melhorar a qualidade das previsões de outras variáveis do sistema	7	18,92%	(Dueker, 2005) (Billio Monfort, 1998) (Sattari et al.; 2020) (Wang et al.; 2012) (Jakubowski Tajduś, 2014) (Farhadi et al.; 2020) (Herrmann, 1989)
O modelo pode ser ampliado e/ou utilizado com outros objetivos	21	56,75%	(Alfaro et al.; 2008) (Harvey Fernandes, 1989) (Brauers et al.; 2011) (Hassler Thadewald, 2003) (Kuo et al.; 2016) (Christensen, 1986) (Cao et al.; 2017) (Niedbała et al.; 2019) (Séverin, 2010) (Pecican, 2010) (Anagnostis et al.; 2020) (Jurasz Mikulik, 2016) (Costa et al.; 2016) (Annarelli et al.; 2018) (Elwakil Zayed, 2015) (Qu Zhou, 2017) (Kokodey, 2012) (Hallah Aboukhamseen, 2019) (Varga et al.; 2008) (Pérez-Rodríguez, 2011) (Dohnal Doubravsky, 2015)

Tabela 3 – Principais lacunas identificadas pelos autores dos artigos coletados (conclusão)

Lacunas identificadas	Nº de artigos		Autores
Os métodos qualitativo e quantitativo foram analisados de forma separada.	1	2,70%	(Alberti et al.; 2020)
Os resultados questionam a tendência entre os países latino-americanos de reduzir os gastos do governo e a futura fonte de renda e crescimento do emprego	2	5,41%	(Ramírez, 1998) (Ramirez, 1996)
Uma revisão de literatura para analisar as descobertas	2	5,41%	(Spooner, 1990) (Borgegard Haggstrom, 1985)
Total Geral	37	100%	

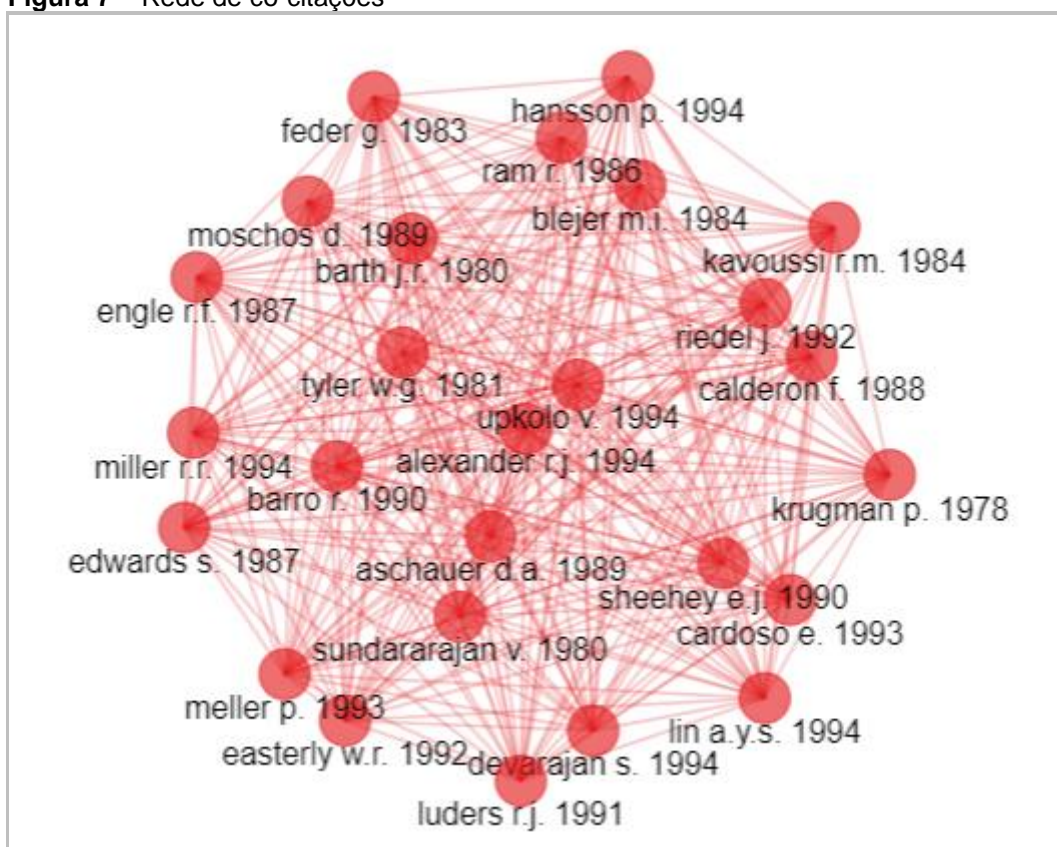
Fonte: Autores (2022).

O resultado que mais se destaca na Tabela 3 é a possibilidade identificada de ampliação do modelo que foi criado pelos autores, bem como a utilização com outros tipos de dados, com 56,75%. Os modelos de séries temporais já consolidados possuem como característica a possibilidade de aplicação em séries de sistemas distintos, baseados em um ciclo iterativo, parcimônia e que apresentem ruído branco nos resíduos (Box et al.; 1994). No entanto, a diferença entre os modelos indicados é a inclusão de variáveis qualitativas, e a ampla possibilidade de aprimoramento para a melhoria da qualidade das previsões geradas.

Outro destaque é a indicação de que o modelo criado pode melhorar a qualidade das previsões de outras variáveis do sistema, com 18,92%. Neste caso, as descrições dos autores variaram entre a alteração de variáveis qualitativas para aprimorar as previsões do sistema e a geração de previsões de outras séries vinculadas ao mesmo sistema estudado. A aplicação do modelo para outros tipos de sistema não foi mencionada, mas sua aplicação não é deve ser descartada sem prévio teste.

Os demais autores indicaram a possibilidade de uso dos modelos criados nas mesmas áreas de conhecimento aos quais foram aplicados com a diferença da adoção de séries de outras localidades geográficas. As referências utilizadas para desenvolver esses artigos estão brevemente apresentados na Figura 7, que contém a rede de co-citações, ou seja, o embasamento teórico utilizado para a geração dos estudos.

Figura 7 – Rede de co-citações



Fonte: Autores (2021).

A Figura 7 representa a relação de citações das 50 referências citadas com maior frequência entre os artigos da pesquisa. A rede indica as ocorrências relacionadas e o tamanho do nó representa a quantidade de citações de cada autor. Neste caso, as 50 referencias indicaram uma rede interligada com quantidades semelhantes de citações para cada autor, visto que os círculos possuem tamanhos similares. Essa característica indica a proximidade conceitual e metodológica das pesquisas coletadas pelo estudo devido a identificação dos artigos basilares que dão sustentação aos artigos.

A análise da rede de cocitações possibilita a compreensão dos trabalhos mais influentes e o que já foi estudado sobre a relação entre modelos quantitativos e variáveis qualitativas. De acordo com os resultados extraídos, essas fontes de referência não geraram uma rede de citações direta históricas. Essa rede consiste em um mapa cronológico das citações relevantes resultantes do conjunto bibliográfico em estudo. Os estudos coletados são independentes entre si a nível de

referências, um estudo não utiliza o outro como fonte, tampouco para a evolução da temática.

O conhecimento gerado pelos estudos é a união de modelos quantitativos e a inserção de variáveis qualitativas como entrada em formato quantitativo. O valor atribuído para as variáveis qualitativas foi definido de várias formas, como pelo uso da probabilidade e das distribuições, aplicação de ponderações devido a conhecimento prévio dos especialistas ou frequências absoluta ou relativa. Os artigos analisados adaptam e integram variáveis qualitativas aos modelos quantitativos para gerar previsões ou apoiar tomadas de decisões. Na maior parte deles os autores criaram um algoritmo, modelaram equações, nas quais as principais características identificadas nos cenários foram inseridas como uma entrada de informação e influenciaram no resultado, na previsão. Também foram utilizados modelos fuzzy, redes neurais, regressão simples e múltipla, correlação, árvore de decisão entre outros, para compor novos modelos de previsões. Não foram encontradas avaliações do impacto das variáveis nas séries ao longo do tempo, o que foi considerada uma lacuna no conhecimento possível de ser explorada.

3 CONSIDERAÇÕES

Através da revisão sistemática de literatura pode-se verificar o conhecimento que já foi gerado sobre determinada temática e, a partir disso, pesquisar para sua evolução. O resultado obtido da inserção da *String* nas bases científicas foi de 37 artigos. Da análise desses artigos identificou-se a união de variáveis qualitativas e quantitativas para a criação de modelos gerados a partir de modelos quantitativos adaptados em 33 deles. Nesses casos as variáveis qualitativas receberam ponderações ou alguma forma de transformação para torná-las quantitativas. Foram identificadas as inserções de variáveis qualitativas como variáveis quantificadas por ponderações ou por frequência juntamente com modelos Fuzzy, VAR, regressão, correlação entre outros. Em outros dois casos as variáveis qualitativas foram tratadas como binárias. Os modelos foram aplicados em diversas áreas como economia, saúde, ecologia, energia, comportamento humano, entre outros.

Os modelos quantitativos servem de apoiadores a tomada de decisões e para geração de conhecimento com relação ao comportamento das variáveis em estudo. A integração de variáveis qualitativas e quantitativas auxilia na exploração ampla da complexidade do campo de estudo e incentiva o desenvolvimento da modelagem e do conhecimento como um todo. A importância de estudos como este se dá em função da demonstração da relevância da integração de conhecimentos qualitativos e quantitativos que já foram efetuados e de incentivo ao aprofundamento das pesquisas e engrandecimento das conclusões, pois características não mensuráveis podem ter tanto impacto quanto às mensuráveis.

A aplicação das variáveis qualitativas foi efetuada de forma quantificada ou como entrada de informação. Não foram encontradas análises da interferência ou influência dessas variáveis para as séries temporais. Nas comparações de aplicação dos modelos com ou sem entradas qualitativas os autores relataram melhoria da precisão das previsões geradas em modelos com qualitativas inclusas.

A lacuna identificada em 56,75% dos estudos foi a capacidade de melhoria dos modelos criados, assim como aplicação de outros sistemas de dados. E em aproximadamente 19% dos estudos os autores identificaram a possibilidade de melhoria da qualidade das previsões de outras variáveis do mesmo sistema. Esses resultados indicam que, mesmo com a alta precisão dos modelos quantitativos previsão, é possível desenvolver modelos de previsão mais precisos através da inclusão de variáveis qualitativas.

O mapa temático resultante das análises indicou como temáticas potenciais as análises qualitativas e quantitativas e erro quadrático médio, que indica a união dos modelos com o foco na redução de erros. Esse resultado indica que há potencial para a criação de modelo que unam essas variáveis em modelos mais precisos para serem usados como apoio a tomada de decisão. O baixo nível de contribuição entre grupos de pesquisa de países diferentes também pode ser destacado, e serve de indicativo de desafios a superar para o avanço colaborativo da temática.

Os resultados encontrados não têm a pretensão de esgotar a temática. A limitação encontrada foi a publicação de artigos em que os modelos obtiveram sucesso nas análises propostas. A não publicação de estudos em que os modelos não sugerem melhorias efetivas anula a possibilidade de verificação de reversão

desses resultados, bem como pode gerar viés no estudo e a perda de eficiência na geração do conhecimento científico, pois se resultados negativos forem expostos caminhos dúbios permanecem ocultos e sem desenvolvimento de discussões e melhorias. Sugere-se para outras pesquisas que o estudo seja aplicado para outras bases científicas, que podem acrescentar informações relevantes aos resultados encontrados.

REFERÊNCIAS

ALBERTI, T. S.; BRUHN, F. R. P.; ZAMBONI, R.; VENANCIO, F. R.; SCHEID, H. V.; RAFFI, M. B.; SCHILD, A. L.; SALLIS, E. S. V. (2020). Epidemiological analysis of bovine tuberculosis in the southern region of Rio Grande do Sul from 2000 to 2015. **Pesquisa Veterinária Brasileira**, v. 40, n. 2, p. 77–81. DOI

<https://doi.org/10.1590/1678-5150-PVB-6406>

ALFARO, E.; GARCÍA, N.; GÁMEZ, M.; ELIZONDO, D. Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. **Decision Support Systems**, v. 45, n. 1, p. 110–122, 2008. DOI DOI

<https://doi.org/10.1016/j.dss.2007.12.002>

ANAGNOSTIS, A.; PAPAGEORGIOU, E.; BOCHTIS, D. Application of artificial neural networks for natural gas consumption forecasting, **Sustainability** (Switzerland), v. 12, n. 16, p. 6409, 2020. DOI <https://doi.org/10.3390/SU12166409>

ANNARELLI, A.; BATTISTELLA, C.; BORGIANNI, Y.; NONINO, F. Estimating the value of servitization: A non-monetary method based on forecasted competitive advantage. **Journal of Cleaner Production**, 200, p. 74–85, 2018. DOI

<https://doi.org/10.1016/j.jclepro.2018.07.220>

BILLIO, M.; MONFORT, A. Switching state-space models Likelihood function, filtering and smoothing. **Journal of Statistical Planning and Inference**, v. 68, n. 1, p. 65–103, 1998. DOI [https://doi.org/10.1016/S0378-3758\(97\)00136-5](https://doi.org/10.1016/S0378-3758(97)00136-5)

BOADA, A. “Sistema Forecast”. **Predicción automatizada en empresas de venta directa**, v. 32, n. 11, p. 121–142, 2016.

BORGEARD, L. E.; HAGGSTROM, N. Migration and social development in a household perspective: an attempt to develop an integrated model of migration. **Espace-Populations-Societes**, v. 1985, n. 1, p. 26–32, 1985. DOI

<https://doi.org/10.3406/espos.1985.996>

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis: Forecasting and Control**. 3 ed. 1994. Disponível em:
<https://books.google.com.br/books?id=sRzvAAAAMAAJ>

BRAUERS, W. K. M.; BALEŽENTIS, A.; BALEŽENTIS, T. Multimoora for The Eu Member States Updated With Fuzzy Number Theory. **Technological and Economic Development of Economy**, v. 17, n. 2, p. 259–290, 2011. DOI
<https://doi.org/10.3846/20294913.2011.580566>

CAO, S.; LU, A.; WANG, J.; HUO, L.; BARCELO, D. Modeling and mapping of cadmium in soils based on qualitative and quantitative auxiliary variables in a cadmium contaminated area. **Science of the Total Environment**, v. 580, p. 430–439, 2017. DOI <https://doi.org/10.1016/j.scitotenv.2016.10.088>

CHRISTENSEN, R. Entropy minimax multivariate statistical modeling—II: Applications. **International Journal of General Systems**, v. 12, n. 3, p. 227–305, 1986. DOI <https://doi.org/10.1080/03081078608934938>

COSTA, N. DO R.; MARCELINO, M. A.; DUARTE, C. M. R.; UHR, D. Proteção social e pessoa com deficiência no Brasil. **Ciência Saúde Coletiva**, v. 21, n. 10, p. 3037–3047, 2016. DOI <https://doi.org/10.1590/1413-812320152110.18292016>

DOHNAL, M.; DOUBRAVSKY, K. Qualitative upper and lower approximations of complex nonlinear chaotic and nonchaotic models. **International Journal of Bifurcation and Chaos**, v. 25, n. 13, 1550173, 2015. DOI
<https://doi.org/10.1142/S0218127415501734>

DUEKER, M. Dynamic Forecasts of Qualitative Variables. **Journal of Business Economic Statistics**, v. 23, n. 1, p. 96–104, 2005. DOI
<https://doi.org/10.1198/073500104000000613>

ELWAKIL, E.; ZAYED, T. Construction knowledge discovery system using fuzzy approach. **Canadian Journal of Civil Engineering**, v. 42, n. 1, p. 22–32, 2015.
<https://doi.org/10.1139/cjce-2014-0153>

FARHADI, R.; AFKARI-SAYYAH, A. H.; JAMSHIDI, B.; MOUSAPOUR GORJI, A. Prediction of internal compositions change in potato during storage using visible/near-infrared (Vis/NIR) spectroscopy. **International Journal of Food Engineering**, v. 16, n. 4, 2020. DOI <https://doi.org/10.1515/ijfe-2019-0110>

FIGINI, S.; GIUDICI, P. Statistical merging of rating models. **Journal of the Operational Research Society**, v. 62, n. 6, p. 1067–1074, 2011. DOI
<https://doi.org/10.1057/jors.2010.41>

HALLAH, R. M.; ABOUKHAMSEEN, S. Cross-calibration of categorical variables: An evaluation of the genetic algorithm approach. **Applied Soft Computing Journal**, 74, p. 154–166, 2019. DOI <https://doi.org/10.1016/j.asoc.2018.10.009>

HARVEY, A. C.; FERNANDES, C. Time Series Models for Count or Qualitative Observations. **Journal of Business Economic Statistics**, v. 7, n. 4, p. 407–417, 1989. DOI <https://doi.org/10.1080/07350015.1989.10509750>

HASSLER, U.; THADEWALD, T. Nonsensical and Biased Correlation Due to Pooling Heterogeneous Samples. **Source: Journal of the Royal Statistical Society. Series D (The Statistician)**, v. 52, n. 3, p. 367–379, 2003. Disponível em: <https://about.jstor.org/terms>

HELBOK, R.; KRUDSOOD, S.; WILAIRATANA, P.; LACKNER, P.; TREEPRASERTSUK, S.; DENT, W.; NACHER, M.; SILACHAMROON, U.; SCHMUTZHARD, E.; LOOAREESUWAN, S. The use of the Multi-organ-Dysfunction Score to discriminate different levels of severity in severe and complicated Plasmodium falciparum malaria. **The American Journal of Tropical Medicine and Hygiene**, v. 72, n. 2, p. 150–154, 2005. DOI <https://doi.org/10.4269/ajtmh.2005.72.150>

HERRMANN, R. Agricultural Price Protection, Import Dependence and Economic Development: The Case of Wheat. **Journal of Agricultural Economics**, v. 40, n. 2, p. 152–167, 1989. DOI <https://doi.org/10.1111/j.1477-9552.1989.tb01095.x>

JAKUBOWSKI, J.; TAJDUŚ, A. Predictive Regression Models of Monthly Seismic Energy Emissions Induced by Longwall Mining. **Archives of Mining Sciences**, v. 59, n. 3, p. 705–720, 2014. DOI <https://doi.org/10.2478/amsc-2014-0049>

JURASZ, J.; MIKULIK, J. **Day ahead electric power load forecasting** by WT-ANN. **PRZEGLĄD ELEKTROTECHNICZNY**, v. 1, n. 4, p. 154–156, 2016. DOI <https://doi.org/10.15199/48.2016.04.32>

KOKODEY, T. A. A Composite Technique for Modeling and Projecting Food Consumer Behavior. **Journal of International Food Agribusiness Marketing**, v. 4, n. 3, p. 231–249, 2012. DOI <https://doi.org/10.1080/08974438.2012.691815>

KUO, R. J.; TSENG, Y. S.; CHEN, Z.-Y. Integration of fuzzy neural network and artificial immune system-based back-propagation neural network for sales forecasting using qualitative and quantitative data. **Journal of Intelligent Manufacturing**, v. 27, n. 6, p. 1191–1207, 2016. DOI <https://doi.org/10.1007/s10845-014-0944-1>

NIEDBAŁA, G.; NOWAKOWSKI, K.; RUDOWICZ-NAWROCKA, J.; PIEKUTOWSKA, M.; WERES, J.; TOMCZAK, R. J.; TYKSIŃSKI, T.; ÁLVAREZ PINTO, A. Multicriteria Prediction and Simulation of Winter Wheat Yield Using Extended Qualitative and

Quantitative Data Based on Artificial Neural Networks. **Applied Sciences**, v. 9, n. 14, 2773, 2019. DOI <https://doi.org/10.3390/app9142773>

PECICAN, E. S. Forecasting Based on Open VAR Model. **Romanian Journal of Economic Forecasting-1**, 1, p. 59–69, 2010.

PÉREZ-RODRÍGUEZ, J. V. Probability of an incoming order signal. *Quantitative Finance*, v. 11, n. 6, p. 901–916, 2011. DOI <https://doi.org/10.1080/14697681003685555>

QU, S.; ZHOU, Y. A Study of The Effect of Demand Uncertainty for Low-Carbon Products Using a Newsvendor Model. **International Journal of Environmental Research and Public Health**, v. 14, n. 11, 1276, 2017. DOI <https://doi.org/10.3390/ijerph14111276>

RAMIREZ, M. D. Public and private Investment in Mexico and Chile: An empirical test of the complementarity hypothesis. **Atlantic Economic Journal**, v. 24, n. 4, p. 301–320, 1996. DOI <https://doi.org/10.1007/BF02298433>

RAMÍREZ, M. D. **Does Public Investment Enhance Labor Productivity Growth in Chile? A Cointegration Analysis**. 1998.

SATTARI, M. T.; MIRABBASI, R.; JARHAN, S.; SHAKER SUREH, F.; AHMAD, S. Trend and abrupt change analysis in water quality of Urmia Lake in comparison with changes in lake water level. **Environmental Monitoring and Assessment**, v. 192, n. 10, p. 1–16, 2020. DOI <https://doi.org/10.1007/s10661-020-08577-8>

SÉVERIN, E. Self organizing maps in corporate finance: Quantitative and qualitative analysis of debt and leasing. **Neurocomputing**, v. 73, n. 10–12, p. 2061–2067, 2010. DOI <https://doi.org/10.1016/j.neucom.2009.12.024>



Artigo recebido em: 10/02/2022 e aceito para publicação em: 30/08/2022
DOI: <https://doi.org/10.14488/1676-1901.v22i1.4575>