

## Abordagens computacionais para *Big Data* aplicadas ao mercado financeiro: um mapeamento sistemático

### Big Data Computational Approaches Applied to the Financial Market: a systematic mapping

#### RESUMO

**Nathiely Laiane Moraes Macedo**  
[nathiely@alunos.utfpr.edu.br](mailto:nathiely@alunos.utfpr.edu.br)  
Universidade Tecnológica Federal do Paraná, Comélio Procópio, Paraná, Brasil

**Érica Ferreira de Souza**  
[ericasouza@utfpr.edu.br](mailto:ericasouza@utfpr.edu.br)  
Universidade Tecnológica Federal do Paraná, Comélio Procópio, Paraná, Brasil

**Giovani Volnei Meinerz**  
[giovaniminerz@utfpr.edu.br](mailto:giovaniminerz@utfpr.edu.br)  
Universidade Tecnológica Federal do Paraná, Comélio Procópio, Paraná, Brasil

A Gestão do Conhecimento (GC) emerge como um importante meio para gerir o conhecimento útil. No setor financeiro, o conhecimento gerado pode auxiliar investidores em sua tomada de decisão acerca das ações negociadas em uma bolsa de valores. Neste setor, o volume de dados, a velocidade de geração e processamento dos dados de diferentes fontes criam desafios, tais como o armazenamento, processamento, visualização e, principalmente análise dos dados. O grande volume de dados produzidos por diferentes fontes, distribuídas e descentralizadas que geram rapidamente dados com relações complexas e em evolução é chamado *Big Data*. O objetivo deste trabalho é sumarizar as principais abordagens computacionais para *Big Data* aplicadas no mercado financeiro a fim de identificar o estado da arte na área. O método utilizado é a condução de um mapeamento sistemático da literatura. Os resultados alcançados por esta pesquisa fornecem uma orientação para posicionar apropriadamente novas atividades de pesquisa no tópico investigado.

**PALAVRAS-CHAVE:** Gestão do conhecimento. Análise financeira. Ciência de dados.

**Recebido:** 19 ago. 2019.

**Aprovado:** 01 out. 2019.

**Direito autoral:** Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



#### ABSTRACT

Knowledge Management (KM) emerges as an important means for managing useful knowledge. In the financial sector, the knowledge generated can assist investors in their decision making about stocks traded on a stock exchange. In this domain, the data volume, generation speed and data processing from different sources create challenges such as storage, processing, visualization and especially data analysis. The large data volume produced by different distributed and decentralized sources that quickly generate data with complex and evolving relationships is called Big Data. The objective of this study is to summarize the main computational approaches to Big Data applied to the financial market in order to identify the state of art in the area. We conducted a systematic mapping. The results achieved by this study provide guidance for appropriately positioning new research activities on the topic investigated.

**KEYWORDS:** Knowledge management. Financial analysis. Data science.

## INTRODUÇÃO

Um dos maiores problemas no processo de tomada de decisão nas organizações é a baixa taxa de reutilização de conhecimento. Isso ocorre porque a maior parte do conhecimento nas organizações não é tratada, tornando difícil sua articulação. Diante deste contexto, a Gestão do Conhecimento (GC) emerge como um importante meio para gerir o conhecimento útil (DAVENPORT; PRUSAK, 2000). A GC tem como objetivo armazenar, organizar e transferir conhecimento essencial aos negócios de uma empresa, de maneira a melhorar o desempenho do funcionário e a competitividade corporativa (NONAKA; TAKEUCHI, 1995).

No setor financeiro, o conhecimento quando bem gerenciado pode auxiliar investidores em sua tomada de decisão acerca das ações de determinadas empresas, negociadas em uma bolsa de valores, que em um determinado momento estejam mais atrativas. Neste setor, o volume de dados, a velocidade de geração e processamento dos dados de diferentes fontes criam desafios, tais como o armazenamento, processamento, visualização e, principalmente análise dos dados. “O grande volume de dados produzidos por diferentes fontes autônomas, distribuídas e descentralizadas que geram rapidamente dados com relações complexas e em evolução é chamado *Big Data*” (MARQUESONE, 2016). *Big Data* refere-se a grandes conjuntos de dados impossíveis de serem gerenciados e processados usando apenas ferramentas tradicionais de gerenciamento de dados, exigindo plataformas computacionais mais complexas para serem analisados (AKOKA et al., 2017).

Na literatura, tanto pesquisas sobre *Big Data*, como mercado financeiro têm aumentado consideravelmente e diversas abordagens computacionais têm sido propostas para dar suporte às atividades de processamento de dados neste domínio. No mercado financeiro, *Big Data* passou a ser utilizado em várias empresas que buscam criar rotinas mais eficazes, avaliar com mais velocidade as tendências de mercado e rastrear fraudes com maior precisão. Dessa forma, a análise dos dados gerados passou a ser uma ferramenta estratégica, tornando a empresa mais preparada para lidar com mudanças no mercado e com os desafios do mundo digital.

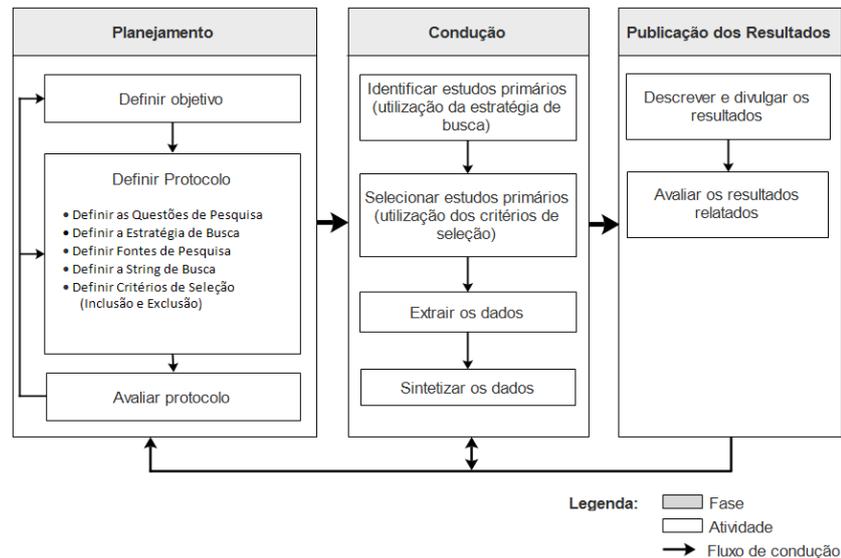
O objetivo deste trabalho é sumarizar as principais abordagens computacionais para *Big Data* aplicadas no mercado financeiro a fim de identificar o estado da arte na área. Para alcançar o objetivo proposto, um mapeamento sistemático da literatura foi conduzido. Um mapeamento sistemático fornece uma visão ampla de uma área de pesquisa para determinar se há evidências de pesquisa sobre um determinado tópico (Kitchenham, 2007).

## MATERIAIS E MÉTODOS

O objetivo da condução de estudos secundários é integrar e sintetizar evidências relacionadas a determinadas questões de pesquisa a partir de um conjunto de estudos chamados estudos primários (NAKAGAWA et al., 2017). As Revisões Sistemáticas (RS) e Mapeamentos Sistemáticos (MS) são tipos de estudos secundários que seguem um processo de pesquisa bem definido para identificar, analisar e interpretar tais evidências (KITCHENHAM; CHARTERS, 2007).

O processo para conduzir uma RS ou MS, envolve as mesmas atividades: planejamento, condução e divulgação dos resultados. Esse é um processo fortemente iterativo e que pode requerer muitas revisões. A Figura 1 apresenta quais são as principais atividades que compõem cada uma dessas etapas.

Figura 1 – Processo de condução de um estudo secundário.



Fonte: Adaptado de (NAKAGAWA et al., 2017).

Para alcançar o objetivo proposto neste trabalho, um mapeamento sistemático da literatura foi conduzido. Conforme descrito na Figura 1, na etapa de planejamento é definido um protocolo a ser seguido, o qual define todos os procedimentos a serem realizados para condução do mapeamento. A seguir é apresentado o protocolo utilizado para guiar o mapeamento sistemático conduzido neste trabalho.

- A) **Questões de pesquisa (QP):** (QP1) Onde os estudos foram publicados?; (QP2) Qual o propósito do estudo no mercado financeiro?; (QP3) Quais são as fontes de dados utilizadas das quais os dados são coletados?; e (QP4) Quais técnicas, algoritmos, modelos e ferramentas foram empregados?
- B) **String de Busca:** Na definição da *String* de busca foram consideradas duas áreas - “Big Data” e “Mercado Financeiro”: (“big data” OR “data science”) AND (“financial market” OR “stock market” OR “stock price” OR “stock data analysis” OR “stock movement” OR “financial big data”)
- C) **Base de Busca:** A base de busca utilizada para identificar os estudos primários a partir da *string* de busca é a Scopus<sup>1</sup>.
- D) **Critérios de Seleção:** Critério de Inclusão (CI): (CI1) O estudo deve apresentar abordagens computacionais para Big Data no mercado financeiro. Critérios de Exclusão (CE): (CE1) O artigo não possui um resumo (abstract); (CE2) A publicação é apenas um resumo (abstract), não possuindo um texto completo; (CE3) A publicação não é um estudo

<sup>1</sup> <https://www.scopus.com/>

primário; (CE4) A publicação não está escrita em inglês; (CE5) A publicação é uma cópia ou uma versão mais antiga de uma outra publicação já considerado; (CE6) Não foi possível ter acesso ao trabalho completo.

- E) **Avaliação:** Antes da condução do mapeamento, o protocolo foi testado. O teste foi realizado com o objetivo de verificar a viabilidade e adequação do protocolo, com base em um conjunto de estudos pré-selecionados e considerados relevantes para a investigação. Primeiramente, o processo de revisão foi conduzido pela primeira autora deste artigo e, posteriormente, os demais autores realizaram uma validação da revisão.

Seguindo o protocolo definido para identificar e selecionar os estudos relevantes, os dados contidos nos estudos primários foram extraídos, categorizados e sumarizado a fim de responder as questões de pesquisa.

## RESULTADOS E DISCUSSÕES

Nesta seção são apresentados resultados preliminares dos estudos selecionados após a condução do protocolo. A *string* de busca retornou 443 estudos da base da *Scopus*. Uma análise foi realizada nos primeiros 90 artigos retornados. Na primeira etapa do mapeamento, os critérios de seleção foram aplicados na leitura dos títulos e abstracts dos 90 artigos. Após a aplicação dos critérios de seleção, permaneceram 47 estudos. Na segunda etapa, os mesmos critérios de seleção foram aplicados na leitura do artigo inteiro. Permaneceram 25 estudos.

As questões de pesquisa foram respondidas com base nos 25 artigos selecionados. As classificações foram definidas com base nas repostas das questões de pesquisa, sendo assim, alguns artigos podem abranger mais de uma classificação.

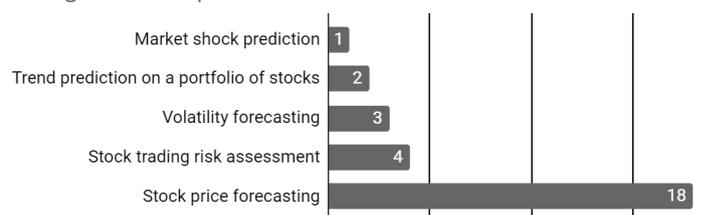
### QP1. Onde os estudos foram publicados?

Os estudos foram classificados em dois meios de publicação: revista e conferência. O meio de publicação na área é equilibrado, dos estudos selecionados; 13 estudos foram publicados em revistas e 12 em conferências.

### QP2. Qual o propósito do estudo no mercado financeiro?

O principal tópico investigado nos estudos selecionados é a predição dos preços das ações, seguido pela avaliação de risco em negociações de ações, predição de volatilidade do mercado e predição de tendências em um portfólio de ações. O gráfico da Figura 2 apresenta os principais propósitos dos estudos selecionados.

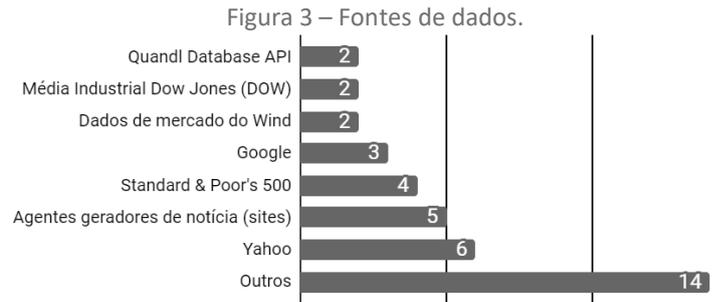
Figura 2 – Propósito do estudo no mercado financeiro.



Fonte: Autoria própria (2019).

**QP3. Quais são as fontes de dados utilizadas das quais os dados são coletados?**

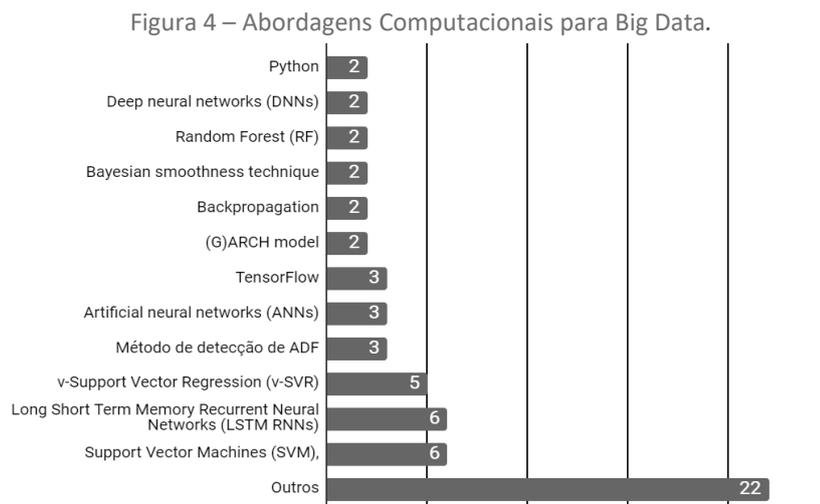
Entre os estudos selecionados, a fonte de dados mais utilizada é o *Yahoo*, citado por 6 artigos. O *Yahoo* fornece dados financeiros através do *Yahoo Finance*. As notícias publicadas em sites podem fornecer informações importantes sobre o mercado financeiro, por isso os agentes geradores de notícias também são muito utilizados nos estudos selecionados. O gráfico da Figura 3 apresenta as principais fontes de dados identificados neste mapeamento.



Fonte: Autoria própria (2019).

**QP4. Quais técnicas, algoritmos, modelos e ferramentas foram empregados?**

Para trabalhar com o alto volume de dados gerado no setor financeiro, a principal metodologia utilizada pelos autores é o aprendizado de máquina. Os modelos mais utilizados são baseados em redes neurais e regressão. Existem várias ferramentas que auxiliam no aprendizado de máquina, a biblioteca *TensorFlow* é a mais citada entre os estudos selecionados. Para validação, este mapeamento sistemático identificou o método de detecção de ADF como um dos mais utilizados nesta área. Demais técnicas, algoritmos, modelos e ferramentas são mostrados no gráfico da Figura 4.



Fonte: Autoria própria (2019).

**CONCLUSÃO**

O objetivo deste artigo foi sumarizar as principais abordagens computacionais para *Big Data* aplicadas ao mercado financeiro a fim de identificar o estado da arte na área por meio da condução de um mapeamento sistemático. Os resultados alcançados, mesmo que preliminares, indicam uma crescente evolução nos métodos utilizados para extrair, analisar e visualizar dados históricos de ações e informações implícitas que são influentes no mercado financeiro, combinando tecnologias de aprendizado de máquina, inteligência artificial, análise estatística e *Big Data*.

Os estudos propõem metodologias e soluções para problemas causados pela complexidade dos dados financeiros a fim de prever os preços das ações, avaliar riscos e a volatilidade financeira baseado em análises sentimentais e dados históricos. A principal contribuição deste mapeamento é destacar as principais ferramentas, técnicas, métodos e algoritmos utilizadas na literatura. O conteúdo deste trabalho é importante para pesquisadores envolvidos na área da computação com aplicações no mercado financeiro, podendo direcionar suas pesquisas futuras e fornecer um posicionamento apropriado para novos estudos.

Como trabalho futuro pretende-se realizar a seleção e síntese dos dados dos demais estudos deste mapeamento, de forma a cobrir os 443 estudos retornados pela *string* de busca.

## REFERÊNCIAS

AKOKA, J.; COMYN-WATTIAU, I.; LAOUFI, N. Research on big data – a systematic mapping study. *Computer Standards & Interfaces*, v. 54, p. 105–115, 2017.

DAVENPORT, T. H.; PRUSAK, L. **Working knowledge: how organizations manage what they know**. 2. ed. Boston, USA: Harward Business School Press, 2000.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. Durham, UK, 2007. v. 2, n. EBSE 2007-001, 1051 p.

MARQUESONE, R. **Big Data: Técnicas e tecnologias para extração de valor dos dados**. Brasil: Casa do Código, 2016. ISBN 9788555192326. Disponível em <https://books.google.com.br/books?id=cbWIDQAAQBAJ>. Acesso em: 03 ago. 2019

NAKAGAWA, E. et al. **Revisão Sistemática da Literatura em Engenharia de Software: Teoria e Prática**. Brazil: Elsevier Brasil, 2017. ISBN 9788535285970.

NONAKA, I.; TAKEUCHI, H. **The Knowledge Creating Company: How Japanese Companies Create The Dynamics Of Innovation**. 1995.