



# INTRODUÇÃO À MINERAÇÃO DE TEXTO

**Luciano Ignaczak**

Sócio-Fundador

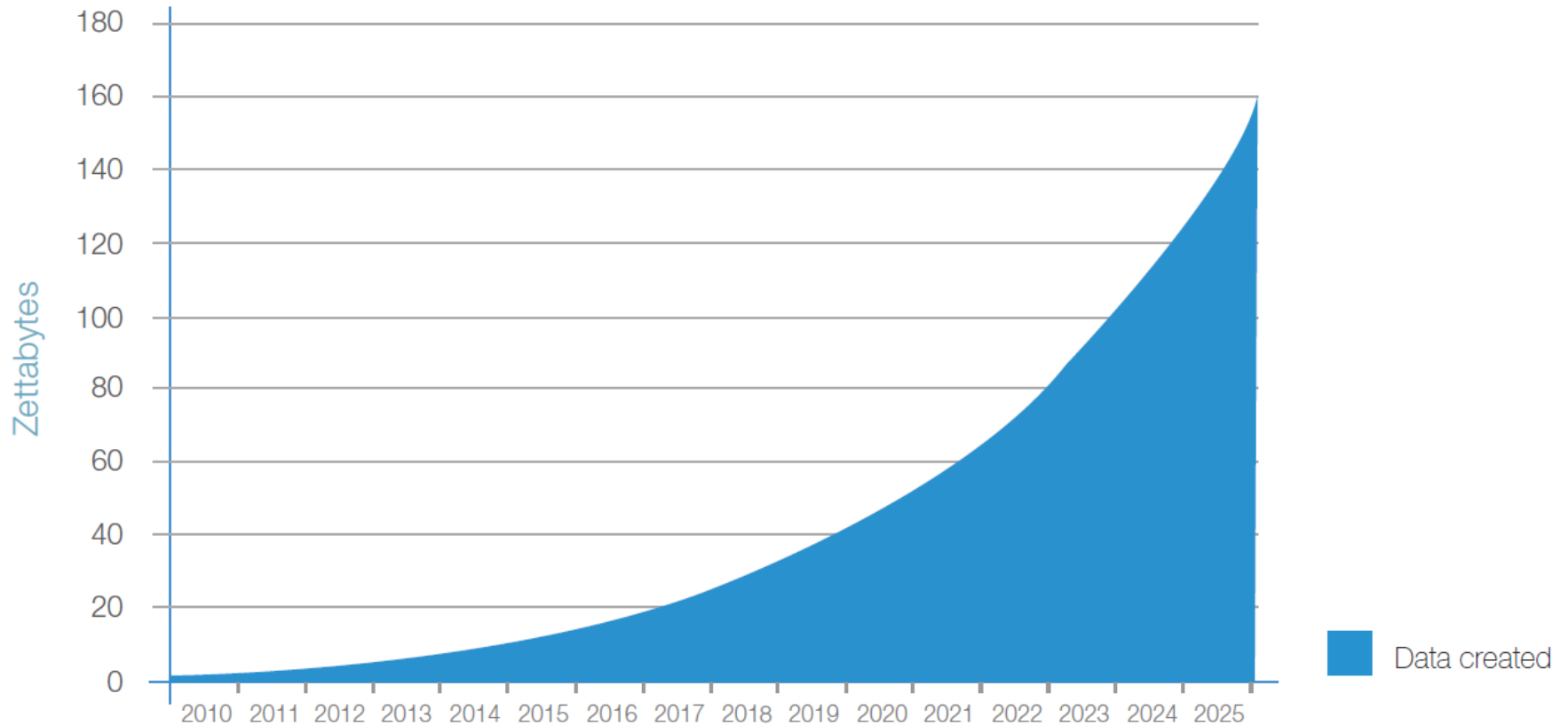
[ignaczak@icybersec.com](mailto:ignaczak@icybersec.com)

**Márcio Garcia Martins**

Sócio-Fundador

[marciogm@icybersec.com](mailto:marciogm@icybersec.com)

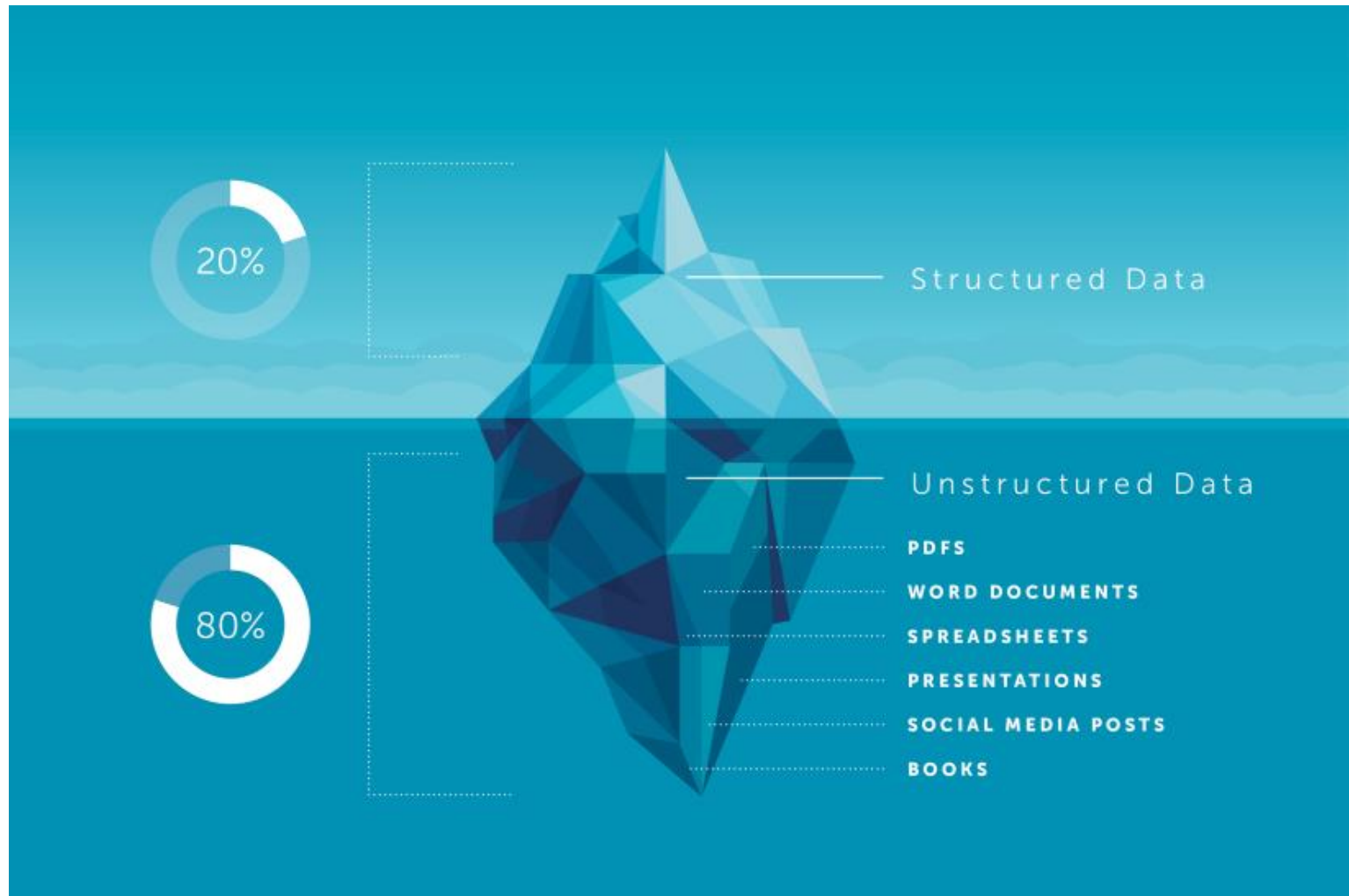
# Cenário Atual



Source: IDC's Data Age 2025 study, sponsored by Seagate, April 2017

Fonte: IDC. **Data Age 2025: The Evolution of Data to Life-Critical.** 2019.

# Cenário Atual

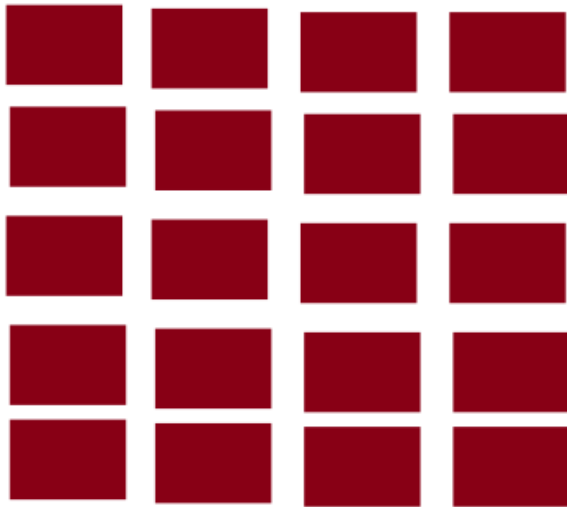


Fonte: EPAM. **Unstructured Data: Elevating Your Business With One of Its Most Valuable Hidden Resources.** 2018.

# Mineração de Texto

- Mineração de texto é o processo para extrair conhecimento implícito a partir de dados textuais;
- A mineração de texto utiliza um processo semelhante a mineração de dados. No entanto, a mineração de texto tem como alvo fontes de dados formadas por coleções de documentos eletrônicos;

# Dados não estruturados



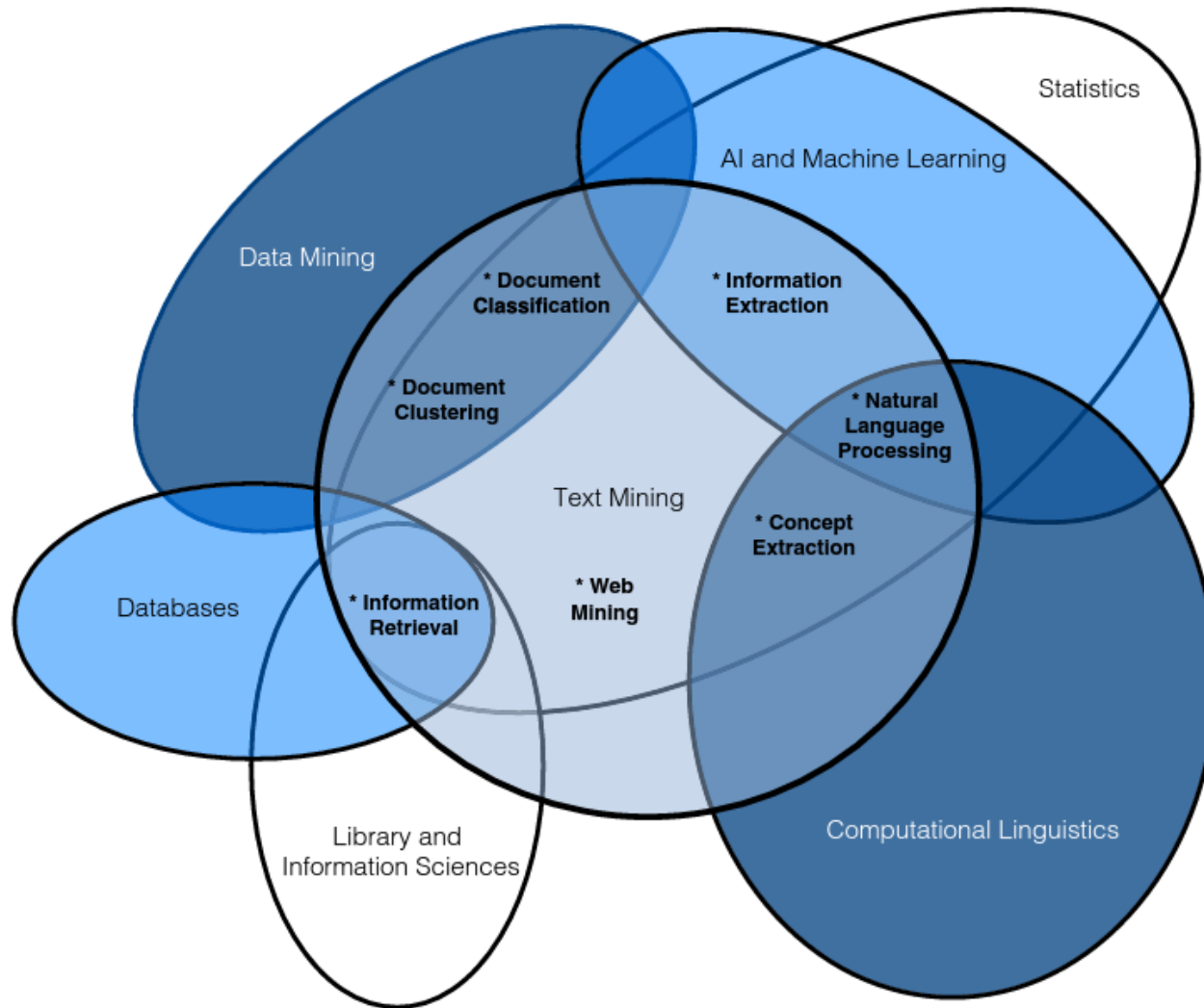
*Structured Data*



*Unstructured Data*

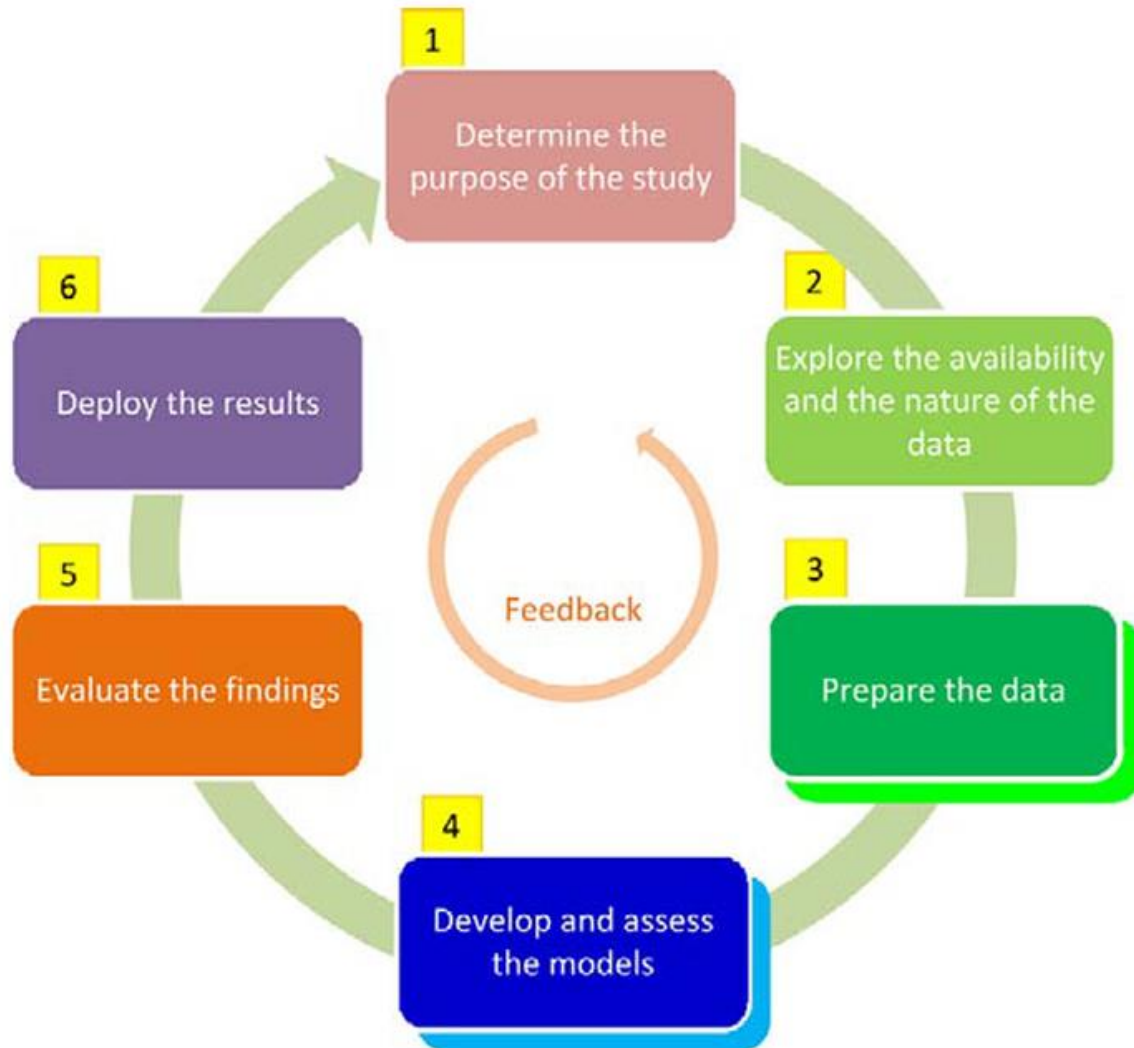
Fonte: EMC. **Unstructured Data: Growth and Challenges.** 2015.

# Áreas de Aplicação



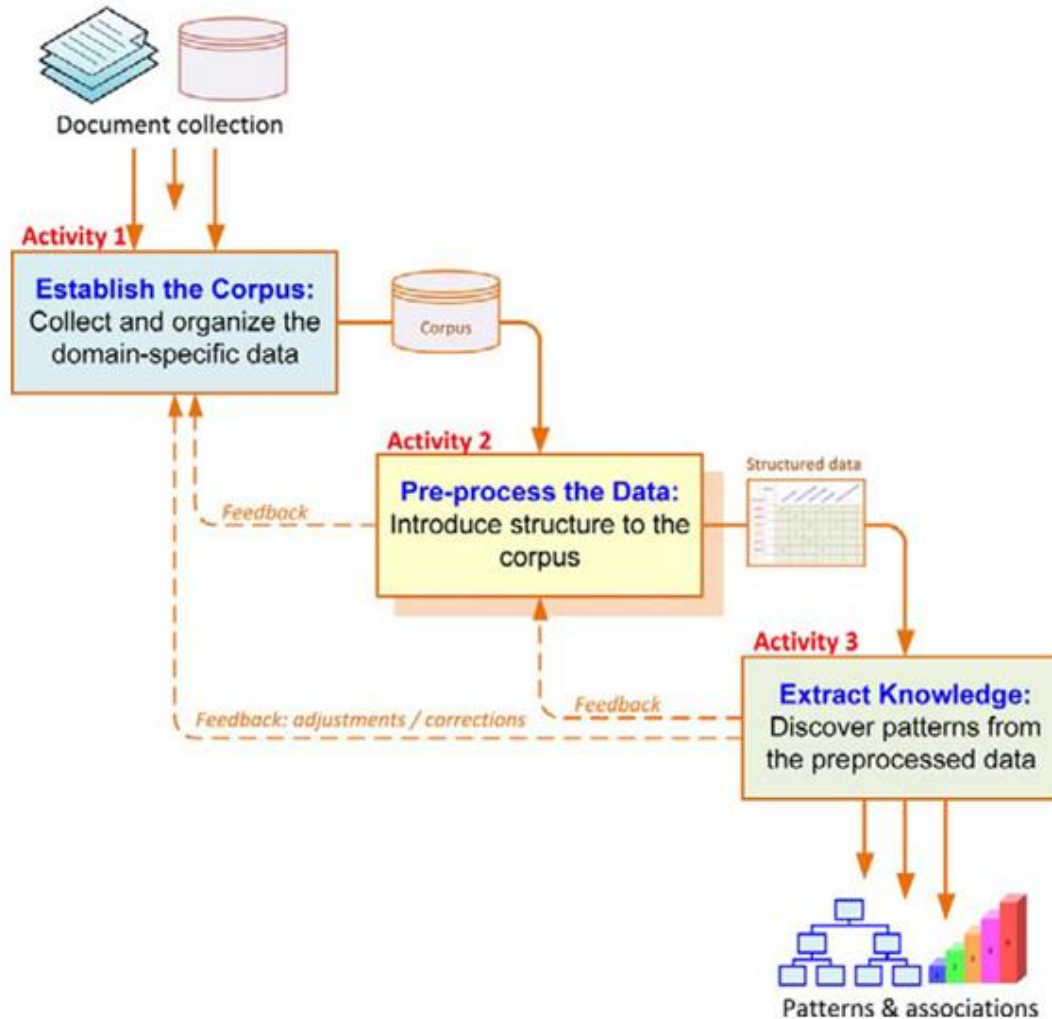
Fonte: MINER, GARY D. et al. **Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications**. 2012.

# Processo CRISP-DM



Fonte: MINER, GARY D. et al. **Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications**. 2012.

# Adaptação para Mineração de Texto



Fonte: MINER, GARY D. et al. **Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications**. 2012.

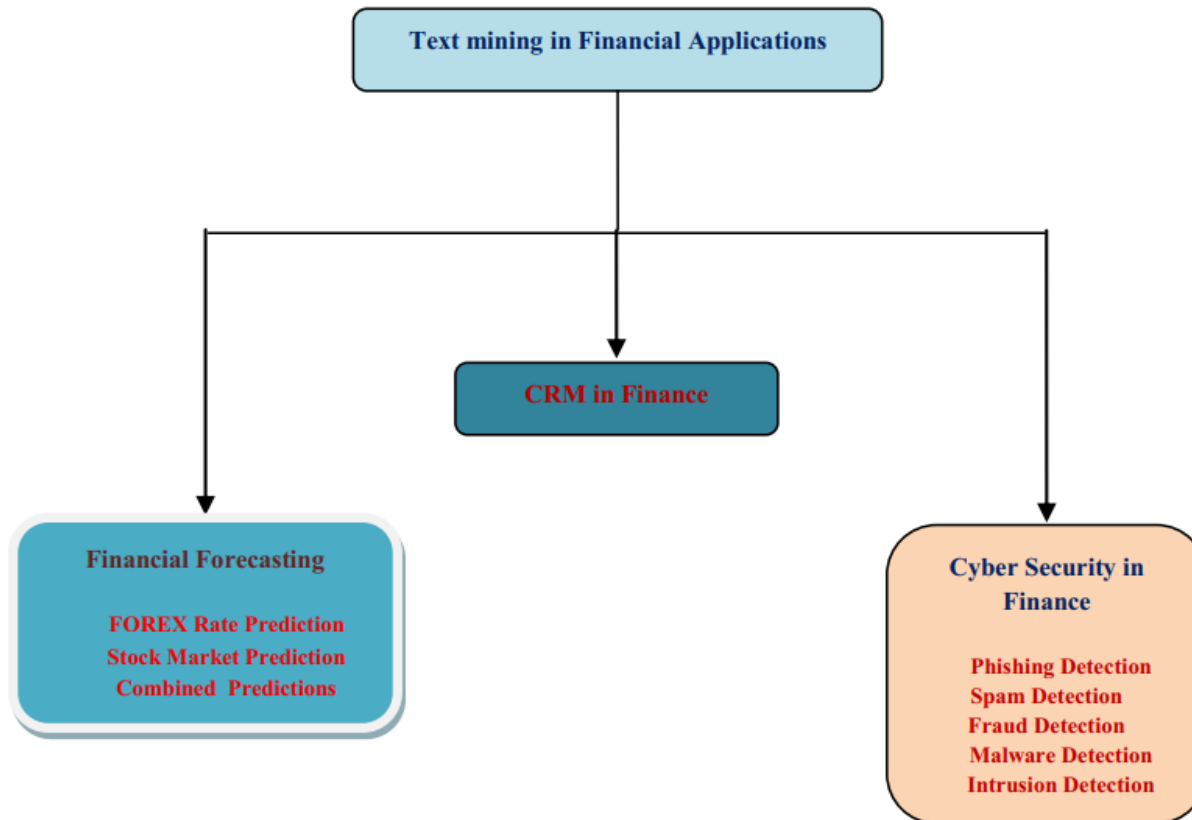


# Exemplos de Aplicação

- Serviços para atendimento ao cliente;
- Gestão de riscos;
- Manutenção;
- Saúde;
- Identificação de SPAM.

Fonte: IBM. **Text Mining**. 2020.

# Exemplos de Aplicação em Finanças



Fonte: KUMAR, SHRAVAN B. and RAVI, VADLAMANI. A survey of the applications of text mining in financial domain. 2016.

# Exemplos de Aplicação em Segurança Cibernética

- Threat Intelligence;
- Cyberbullying;
- Fake news;
- Classificação da Informação;
- Forense digital;
- Predição de incidentes;

Fonte: IGNACZAK, LUCIANO, et al. **Text Mining in Cybersecurity**: A Systematic Literature Review. 2021.