



ESCOLA DE APERFEIÇOAMENTO DE OFICIAIS DA AERONÁUTICA
CURSO DE APERFEIÇOAMENTO DE OFICIAIS 3/2023

ANDRÉ MUNIZ MARINHO DA **ROCHA**, Cap QOEMET

Text Mining: ferramenta de análise automática e inteligente para otimização do assessoramento aos decisores do CIAER

Rio de Janeiro

2023

ESCOLA DE APERFEIÇOAMENTO DE OFICIAIS DA AERONÁUTICA
CURSO DE APERFEIÇOAMENTO DE OFICIAIS 3/2023

ANDRÉ MUNIZ MARINHO DA **ROCHA**, Cap QOEMET

Text Mining: ferramenta de análise automática e inteligente para otimização do assessoramento aos decisores do CIAER

Trabalho de conclusão de curso apresentado no Curso de Aperfeiçoamento de Oficiais da Aeronáutica como requisito parcial para aprovação no Curso de Pós-graduação em Lato Sensu em Liderança com Ênfase em Gestão no COMAER.

Linha de Pesquisa: Doutrina de Inteligência
Orientador: Danilo Bichir, Cap Inf

Rio de Janeiro

2023

ANDRÉ MUNIZ MARINHO DA **ROCHA**, Cap QOEMET

Text Mining: ferramenta de análise automática e inteligente para otimização do assessoramento aos decisores do CIAER

Trabalho de conclusão de curso apresentado no Curso de Aperfeiçoamento de Oficiais da Aeronáutica.

Aprovado por:

Danilo Bichir, Cap Int
EAOAR

Prof. Dr. **André** da Costa Gonçalves
EAOAR

Rio de Janeiro

2023

RESUMO

O Analista de Inteligência transforma os dados em cenário inteligível para a compreensão do que ocorreu, do que ocorre hoje e do que possivelmente ocorrerá, com objetivo específico de auxiliar o decisor. A maioria das pistas que o Analista de Inteligência tem para elaborar os relatórios é de textos, em que é preciso saber analisar e descobrir o conteúdo ocultos. Porém, o rápido crescimento tecnológico acarreta na criação de um grande volume de dados, *Big Data*. Esse novo paradigma impacta na atuação da área de Inteligência, que têm a razão de sua existência fundada na necessidade de informação para auxílio no processo de tomada de decisão. Neste cenário, torna-se importante para a atividade de Inteligência a busca pela melhoria contínua de seus processos e métodos. Como ferramenta de auxílio ao Analista de Inteligência tem-se o *Text Mining*, inspirado no *Data Mining*, para obtenção de informação a partir de textos. Neste cenário, o presente ensaio sustenta a tese da implementação de *Text Mining* como ferramenta de apoio na área de inteligência e sua utilização para aumento da capacidade de interpretar e examinar elementos para a otimização de assessorias no CIAER. Destarte, a partir de diferentes referenciais teóricos pretende-se evidenciar a relevante contribuição da utilização de *Text Mining* como uma ferramenta para extração de informação de dados estruturados ou semi estruturados, e o seu auxílio na análise automática e inteligente de grandes bases de dados, a fim de obter informações que possam ser utilizadas pelos Analistas de Inteligência no assessoramento aos decisores.

Palavras-chave: *Text Mining*. *Data Mining*. Banco de Dados. Processo decisório. Assessoramento de inteligência.

1 INTRODUÇÃO

Atualmente vivencia-se na sociedade em geral a denominada Era da Informação, em que existe um crescimento da quantidade de dados disponível, de modo diversificado e com a constante expansão do armazenamento de dados e informações, muitas vezes desordenados.

A esse fenômeno de acúmulo de grandes massas de dados se atribui o nome de *Big Data*, cujas características desafiam as capacidades globais atuais de transformar dados em informações úteis.

Dentro da Concepção Estratégica da Força Aérea Brasileira (FAB), o uso de ferramentas de auxílio para extração de informações dessas grandes massas de dados é percebida como um dos desafios cruciais da atualidade.

Neste contexto, especificamente na área de inteligência, a produção de documentos textuais tem impulsionado um crescimento exponencial em seus repositórios de dados, ultrapassando a capacidade habitual de interpretar e examinar seus elementos, o que tem levado à necessidade imperativa de adotar novas ferramentas e técnicas para uma análise automática e inteligente dessas bases de dados. Na área de inteligência da FAB, cabe ao Centro de Inteligência da Aeronáutica (CIAER), por meio da análise de seus analistas, o emprego dos dados armazenados nestes grandes repositórios de dados, *Big Data*. Assim, torna-se desafiador e apresenta-se como uma fonte de oportunidade, a análise de tais textos de forma integrada, uma vez que a superioridade da utilização das informações está firmemente baseada na produção de dados oportunos, precisos e de alta fidelidade, para assessoramento das autoridades decisoras.

Considerando esse ponto de vista, o presente ensaio sustenta a tese da implantação de *Text Mining* como ferramenta de análise automática e inteligente para otimização do assessoramento no CIAER.

Partindo dessa proposição, argumenta-se que o *Text Mining* é uma ferramenta que pode ser usada na extração de informação de dados estruturados ou semi estruturados, a fim de obter informações que possam assessorar os decisores. Com o mesmo propósito, o outro argumento refere-se à utilização de *Text Mining*, para auxílio na análise automática e inteligente de grandes bases de dados, a fim de obter informações valiosas, tendências e insights ocultos nos dados de texto.

2 DESENVOLVIMENTO

O Sistema de Inteligência da Aeronáutica (SINTAER) tem por finalidade produzir conhecimentos necessários ao preparo e ao emprego da FAB e promover e manter a salvaguarda das fontes cujo sigilo seja de interesse da Aeronáutica, tendo como um de seus objetivos a produção de conhecimentos que permitam o assessoramento aos Comandantes, Chefes e Diretores das Organizações Militares (OM) do Comando da Aeronáutica (COMAER). Em conformidade com a Doutrina de Inteligência de Defesa do Sistema de Inteligência de Defesa (SINDE) o CIAER, elo central do SINTAER, estabelece um conjunto de conceitos, princípios, normas, métodos e processos que orientam e disciplinam a Atividade de Inteligência no âmbito do COMAER.

Para consecução de seus objetivos a partir do ano de 2021 os documentos textuais produzidos pelos Órgãos de Inteligência (OI), que compõem o SINTAER, passaram a ser enviados ao CIAER no formato *Portable Document Format* (PDF), possibilitando a posterior leitura e identificação de palavras em seu conteúdo.

Para auxílio, na análise desta grande massa textual produzida, tem-se a possibilidade da utilização de *Text Mining*, inspirado no *Data Mining*, para uma melhor investigação e compreensão de *Big Data*. Tal ferramenta pode ser aplicada no CIAER a uma série de problemas de informação, como a compreensão de temas relacionados em documentos de inteligência ou a facilitação da recuperação de informações em dados estruturados, altamente organizados com uma estrutura predefinida na doutrina de produção de documentos de inteligência, ou semi estruturados, com uma estrutura mais flexível, encontrados nas mensagens informais de inteligência. Pode ser de grande valia no início da exploração de pesquisas, permitindo que os dados textuais sugiram temas e conceitos ao Analista de Inteligência durante o processo de análise, fornecendo um ponto de partida útil para avaliar itens a serem investigados.

Como desafio na exploração de pesquisas em dados textuais, Cooper (2005, p. 1) afirma que o maior problema na área de inteligência americana não está na estrutura e nas autoridades, mas, sim no método de análise das informações disponíveis.

2.1 O *Text Mining* como uma ferramenta para extração de informação

As dificuldades do fenômeno de extração de informações de grandes bases textuais são percebidas na prática, pelas Agências de Inteligências dos mais diversos países. Como os dados de texto codificam grande parte de nosso conhecimento acumulado, eles geralmente não podem ser descartados. Como consequência, gera o acúmulo de uma grande quantidade de dados, *Big Data*, que agora estão além da capacidade de qualquer indivíduo examinar, Zhai e Massung (2016, p. 4). Neste cenário, a recepção de documentos textuais provenientes de todos os OI do SINTAER representa uma fonte considerável de dados. A partir dessa vasta quantidade de dados e com a utilização da ferramenta de *Text Mining*, é possível ao Analista de Inteligência do CIAER a otimização da extração de informações que influenciam na tomada de decisões estratégicas.

Desta forma, o *Text Mining* apresenta-se como uma ferramenta que pode ser usada na extração de informação de dados estruturados ou semi estruturados, a fim de obter informações que possam assessorar os decisores, Morais e Ambrósio (2007, p. 1). Portanto, apresenta-se como uma ferramenta de potencial valor, como suporte na tomada de decisões, na otimização das rotinas de análise do CIAER e facilitadora da extração de informações de dados estruturados, ou semi estruturados, pelo Analista de Inteligência.

Dentro do processo de descoberta de conhecimento têm sido empregadas técnicas de análise e extração de informação para encontrar informações ou padrões de informações. A descoberta de conhecimento em documentos textuais está inserida no contexto da descoberta de conhecimento em dados não estruturados, Silva (2021, p. 33). Assim, a descoberta de conhecimento é um processo que está inserido na rotina do Analista de Inteligência, no qual a análise de dados não estruturados, como textos, permite a identificação de relações e correlações que podem não ser evidentes à primeira vista. A aplicação de técnicas apropriadas, como *Text Mining*, facilita a identificação dessas relações, proporcionando uma visão mais profunda e abrangente do conhecimento presente nos documentos de inteligência, otimizando o método de leituras convencionais, bem como a capacidade de concatenação do Analista de Inteligência do CIAER.

Neste campo surge a imprescindibilidade do domínio de técnicas e ferramentas de *Big Data* para a eficácia da atividade de Inteligência no mundo atual, Alves (2018,

p. 40). Segundo, Loh, *et al.*, (2003, p. 357), a tecnologia de *Text Mining* serve para “identificar” os conceitos presentes nos textos, que representam “classes” do mundo real, tais como objetos, sentimentos, eventos e entidades, que permitem compreender quais assuntos estão presentes nos textos ou do que se tratam.

O levantamento de trabalhos científicos que abordam o tema de *Text Mining* auxilia no propósito de buscar pesquisas que possam contribuir, também, para extração de informação em investigações criminais. De acordo com Sampaio e Mancini (2007, p. 84), as revisões sistemáticas propiciam a incorporação de um espectro maior de resultados mais significativo. Desta forma, a ferramenta de *Text Mining*, também é utilizada no campo das investigações criminais na resolução de casos, na extração de informações de relatórios policiais, mensagens de texto, revelando evidências importantes na identificação de padrões, relações e tendências ocultas em grandes volumes de dados textuais. Portanto, da mesma forma, tal ferramenta pode aprimorar a qualidade e a objetividade das informações disponíveis para os Analistas de Inteligência, facilitando o assessoramento disponibilizado pelo CIAER.

Destarte, o argumento da utilização de *Text Mining*, pelo Analista de Inteligência, como ferramenta para otimização da extração de informação de dados estruturados ou semi estruturados, dos documentos de inteligência armazenados no CIAER, suporta a tese da obtenção de informações que possam assessorar os tomadores de decisão.

2.2 Auxílio na análise automática e inteligente da base de dados

Atualmente as áreas governamentais corporativas e científicas tem promovido um crescimento explosivo em seus bancos de dados, superando em muito a usual capacidade de interpretar e examinar estes dados. Esses fatores geram a necessidade de novas ferramentas e técnicas para análise automática e inteligente de bancos de dados, Fayyad, *et al.*, (1996, p. 37). Assim, com o avanço tecnológico houve também um grande crescimento dos volumes de dados coletados e textos armazenados no CIAER, bem como o aumento da exigência da capacidade convencional do Analista de Inteligência de interpretar e examinar tais informações de maneira eficiente e precisa. Na esteira de entendimento, tornou-se imperioso o desafio de lidar com muitos dados, bem como a busca de ferramentas que possam

converter *Big Data* em informações valiosas, acessíveis e mais eficazes para as Assessorias de Inteligência.

Desta forma, a análise automática e inteligente de grandes volumes de dados possui o facilitador da utilização de tecnologias com algoritmos avançados de análise de dados, especialmente no contexto de *Big Data*, que podem proporcionar a automatização e inteligência na interpretação de vasta quantidade de informações armazenadas. Portanto, a automação desse processo maximiza a precisão e a profundidade da análise, porém não dispensa a presença e interferência do Analista de Inteligência para obtenção do conhecimento final. Desta forma, o método de extração de informações identifica as principais palavras e relações dentro do texto. Tal fato ocorre com a procura de sequências predefinidas no texto, em um processo denominado de busca por correspondências de padrões textuais, Vijayarani, *et al.*, (2015, p.1). Assim, o *Text Mining* apresenta-se como uma ferramenta valiosa na descoberta de padrões, associações e regras de classificação, permitindo análise automática e inteligente que pode facilitar o trabalho do Analista de Inteligência do CIAER no assessoramento de autoridades.

Na contínua busca pelo aprimoramento o *Text Mining* apresenta-se no campo interdisciplinar baseado na recuperação de informações, *Data Mining*, aprendizado de máquina, estatística e linguística computacional. São considerados processos necessários no projeto e implementação de uma ferramenta de *Text Mining* o pré-processamento de coleta de documentos (categorização de texto, informações extração, extração de termo), armazenando e analisando representações intermediárias, bem como usando uma técnica selecionada, como análise de distribuição, agrupamento, análise de tendências e regras de associação, e visualizar os resultados, Jusoh (2012, p. 1). Assim, com o crescimento da produção de documentos e textos provenientes da área de inteligência enviados ao CIAER, é plausível a necessidade premente da adoção de novas ferramentas e técnicas, como o *Text Mining*.

Dessarte, o argumento da utilização de *Text Mining*, pelo Analista de Inteligência, para auxílio na análise automática e inteligente de grandes bases de dados, a fim de obter informações valiosas, tendências e insights ocultos nos dados de texto, suporta a tese da implantação de tal ferramenta de análise automática e inteligente para otimização do assessoramento no CIAER.

3 CONCLUSÃO

Diante do exposto, a implantação de *Text Mining* como ferramenta de análise automática e inteligente oferece uma gama de benefícios significativos para a otimização do assessoramento no CIAER, devido ao grande volume de documentos de inteligência recebidos atualmente.

Destarte, a ferramenta de *Text Mining* demonstra ser de grande valia na extração de informações valiosas a partir de dados, sejam eles estruturados ou semi estruturados, possuindo forte potencial para ser utilizada no auxílio na tomada de decisão, a partir de análise da base de dados textuais produzida no CIAER. A utilização da ferramenta de *Text Mining* oferece um suporte valioso, auxiliando na tomada de decisões embasadas em informações mais precisas e relevantes, contudo, destaca-se que a presença e intervenção do Analista de Inteligência continuam sendo indispensáveis para alcançar o conhecimento final.

Neste mesmo diapasão, diante do aumento do recebimento de documentos, textuais da área de inteligência no CIAER, tem-se como favorável a utilização da ferramenta de *Text Mining*, pois auxilia a análise automática e inteligente de grandes bases de dados. Tal ferramenta facilita a atividade do Analista de Inteligência, visto que possibilita a obtenção de informações valiosas, tendências e insights ocultos nos dados de texto, que seriam difíceis ou impossíveis de serem percebidos em leituras convencionais, uma vez que superam a capacidade de concatenação do Analista de Inteligência.

Desta forma, a adequabilidade e as afirmativas apresentadas levam a defesa da tese da implantação de *Text Mining* como ferramenta de análise automática e inteligente para otimização do assessoramento no CIAER. Sobremaneira, apresenta-se ainda a possibilidade da extrapolação do uso da ferramenta de *Text Mining* para outros Órgãos de Direção Geral e Setorial (ODGS) da FAB, que compõem o SINTAER, tais como o Estado-Maior da Aeronáutica (EMAER), Comando-Geral de Apoio (COMGAP), Comando-Geral de Pessoal (COMGEP), Comando de Operações Aeroespaciais (COMAE), e Comando de Preparo (COMPREP), uma vez que possuem semelhantes desafios e missões a serem cumpridas, na atividade de inteligência.

REFERÊNCIAS

- ALVES, P. M. M. R.. O impacto de Big Data na Atividade de Inteligência. **Revista Brasileira de Inteligência**, Brasília, v. 13, n. 1, p. 40, dez. 2018. Disponível em: <https://rbi.enap.gov.br/index.php/RBI/article/view/148/122>. Acesso em: 03 out. 2023.
- COOPER, J. R. **Curing analytic pathologies: Pathways to improved intelligence analysis**. Washington, DC: Center for the Study of Intelligence, p. 1, 2005.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P.. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37-37, 1996.
- JUSOH, S.; ALFAWAREH, H. M.. Techniques, applications and challenging issue in text mining. **International Journal of Computer Science Issues (IJCSI)**, v. 9, n. 6, p. 431, 2012.
- LOH, S.; DE OLIVEIRA, J. P. M.; GAMEIRO, M. A.. Knowledge discovery in texts for constructing decision support systems. **Applied Intelligence**, v. 18, p. 357-366, 2003.
- MORAIS, E. A. M.; AMBRÓSIO, A. P. L.. **Mineração de Textos**. Relatório Técnico – Instituto de Informática. Universidade Federal de Goiás, 2007, p. 1.
- SAMPAIO, R. F.; MANCINI, M. C.. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. **Brazilian Journal of Physical Therapy**, v. 11, 2007, p. 84.
- SILVA, M. P. da. **Descoberta de Conhecimento em Peças de Inquérito Policial Relatado**. 2021. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Santa Catarina, Santa Catarina, 2021, p. 33.
- VIJAYARANI, S. et al. Preprocessing techniques for text mining-an overview. **International Journal of Computer Science & Communication Networks**, v. 5, n. 1, p. 1, 2015.
- ZHAI, C. X.; MASSUNG, S.. **Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining**. USA: Association for Computing Machinery and Morgan & Claypool, 2016, p. 4.