

## AVALIAÇÃO DA QUALIDADE DE DADOS PÚBLICOS DE EXTRAÇÃO E TRATAMENTO DE PRODUTOS MINERAIS

**Renata Moreira Limiro, Universidade Federal de Catalão, renatamlimiro@gmail.com**  
**Douglas Farias Cordeiro, Universidade Federal de Goiás, cordeiro@ufg.br**  
**Núbia Rosa Da Silva, Universidade Federal de Catalão, nubia@ufg.br**

***Resumo:** A qualidade de dados é algo fundamental e necessário para a realização de procedimentos analíticos, principalmente no que tange à garantia de assertividade na geração de informação e descoberta de conhecimento. Diversos conjuntos de dados são disponibilizados para acesso público através das plataformas de dados abertos existentes. Um destes conjuntos se refere aos dados de extração e tratamento de produtos minerais, a qual pode ser utilizada em estudos de gestão de cadeia de suprimentos relacionados à área da exploração de minérios. Diante disso, o objetivo deste artigo é apresentar um estudo avaliativo da qualidade dos dados disponibilizados face à possíveis análises. Os resultados alcançados revelam a necessidade de emprego de técnicas de tratamento e limpeza de dados, assim como técnicas mais robustas que garantam a homogeneidade de dados necessária para comparações a partir de atributos categóricos.*

***Palavras-chave:** qualidade de dados; análise de dados; dados abertos; produtos minerais.*

### 1. INTRODUÇÃO

Gestão de Cadeia de Suprimentos (do inglês, *Supply Chain Management*) pressupõe a integração contínua e efetiva dos agentes de sua cadeia de suprimento e a análise dos impactos dessa integração nos processos internos e externos de cada agentes em busca de melhoria contínua dos resultados de cada participante e da cadeia como um todo (HUGOS, 2011). Diante disso, é de fundamental importância que os produtores de matéria-prima, as empresas de logística, e a indústria e os distribuidores da manufatura busquem se integrar de forma holística com seus pares e clientes, buscando estratégias que promovam o levantamento, compreensão e análise dos impactos de cada uma das etapas dos processos, tanto na etapa atual quanto nas etapas sucessivas.

De acordo com Simchi-Levi, Kaminsky e Simchi-Levi (2010), a cadeia de suprimentos pode ser descrita como o processo no qual matérias primas são manufaturadas, transformadas em produtos, transportadas para depósitos e armazéns, e por fim entregues para varejistas e clientes finais. Desta maneira, é possível entender que existem diferentes agentes envolvidos e diferentes áreas com entregas específicas em cada uma delas. Ainda de acordo com os autores, A Gestão da Cadeia de Suprimentos é responsável pela aplicação de esforços que busquem uma garantia e aprimoramento da integração entre os agentes envolvidos, de modo a potencializar a visibilidade, em termos de satisfação de demanda, face à minimização dos custos relacionados ao procedimentos operacionais de logística e produção.

Considerando os desenvolvimentos relacionados às soluções tecnológicas voltadas aos processos de vendas, manufatura, logística, armazenagem e distribuição, característicos da Indústria 4.0, a Tecnologia da Informação e Comunicação tem potencializado sua importância enquanto ferramenta de auxílio à gestão da cadeia de suprimentos (GOMES; RIBEIRO, 2013). Diante disso, um volume considerável de dados são gerados a partir da execução de processos e de suas atividades no âmbito da cadeia de suprimentos, gerando demandas em termos de gestão e integração, de modo que

informações úteis e de qualidade possam ser geradas e servir de apoio à tomada de decisão estratégica, se levando os variados níveis da cadeia de suprimentos.

No contexto da indústria de minérios, a gestão da cadeia de suprimentos perpassa por vários aspectos, dentre os quais estão incluídas questões que remetem a rotinas legais, tais como a autorização de extração por parte dos órgãos reguladores, o que possui um impacto direto em planejamento, volume e definição de estratégias de manejo e transporte. Para além disso, conjuntos de dados provenientes de atividades de extração e tratamento de produtos minerais podem ser úteis no sentido de identificação de possíveis irregularidades quanto à realização de atividades não autorizadas. No Brasil, o Instituto Brasileiro de Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA) disponibiliza em formato aberto uma base de dados contendo registros de relatórios de extração e tratamento de produtos minerais, a qual possui dados que são de considerável relevância para estudos exploratórios relacionados à Gestão da Cadeia de Suprimentos de minérios no país. A partir disso, este artigo propõe-se a realizar um estudo avaliativo sobre a qualidade de tal conjunto de dados, com o objetivo de identificar as possíveis necessidades de tratamento para propósitos de geração de informação e descoberta do conhecimento.

## 2. MÉTODOS

O objetivo deste artigo se refere à avaliação da qualidade do conjunto de dados disponibilizados acerca da exploração de minérios no Brasil. Os dados foram obtidos através do Portal Brasileiro de Dados Abertos, sendo disponibilizados pelo Instituto Brasileiro de Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA)<sup>1</sup>. A base de dados contém informações sobre os relatórios de Extração e Tratamento de Produtos Minerais (RAPP). O intuito é realizar uma avaliação sobre este conjunto de dados em face de sua qualidade para propósitos de geração de informações e descoberta do conhecimento no apoio à tomada de decisão em processos produtivos.

Neste sentido, com o propósito de avaliar a qualidade dos dados disponibilizados, serão realizadas verificações exploratórias sobre esse conjunto de dados, focada nos atributos constantes da base de dados. Para tanto, inicialmente serão verificados para todos os campos possíveis ocorrências de dados ausentes ou inconsistentes, de acordo com o formato específico. Após isso, serão avaliados os campos numéricos de dados, a fim de verificar a possível existência de *outliers*, ou seja, dados que possuam valores desproporcionais ao conjunto de dados como um todo, podendo provocar anomalias em possíveis aplicações de métodos de mineração de dados (DEAN, 2014). Para os atributos descritivos, nos quais estão enquadrados os atributos referentes ao nome do minério e a unidade de medida, serão verificadas questões de integridade e homogeneidade, ou seja, se uma dada instância apresenta ocorrências que se variam em termos de escrita (exemplo: balsato; BASALTO; bas.). A Figura 1 apresenta o conjunto de atributos constantes da base de dados obtida.

As análises de avaliação da qualidade são realizadas através da construção de rotinas computacionais utilizando a linguagem de programação Python, com as bibliotecas Pandas, Numpy, Matplotlib e Seaborn.

---

<sup>1</sup> <https://dados.gov.br/dataset/extracao-e-tratamento-de-produtos-minerais>

<i>Proponente</i>
<i>CNPJ</i>
<i>Razão Social</i>
<i>Estado</i>
<i>Município</i>
<i>Ano</i>
<i>Nome do Produto</i>
<i>Quantidade Explorada</i>
<i>Unidade de Medida</i>
<i>Tamanho da Área</i>
<i>Nº Decreto da Lavra</i>
<i>Data do Decreto</i>
<i>Ano de início da exploração</i>
<i>Ano de término da exploração</i>
<i>Projeto de Recuperação Ambiental (PRA) aprovado por</i>
<i>Data de Aprovação do PRA</i>
<i>Situação Cadastral</i>

Figura 1. Conjunto de atributos da base de dados explorada.

### 3. DISCUSSÃO

A base de dados obtida possui um total de 37.519 registros, os quais contemplam dados de registros de extração e tratamento de minérios no período compreendido entre os anos de 2006 e 2013, conforme informações apresentadas na documentação dos dados (PORTAL BRASILEIRO DE DADOS ABERTOS, 2021). Diante disso, inicialmente, conforme delimitado nos aspectos metodológicos deste estudo, foram avaliadas possíveis ocorrências de valores ausentes dentro da base de dados. A Tabela 1 apresenta as ocorrências de atributos que apresentam dados ausentes, com o quantitativo de dados ausentes detectados. Para o atributo *Data de Aprovação do PRA*, o qual apresentou um valor expressivo de dados ausentes, aproximadamente 23% do total de dados presente na base encontram-se vazios.

Tabela 1. Ocorrências de dados ausentes.

<b>Atributo</b>	<b>Quantidade dados ausentes</b>
Nome do Produto	1
Nº Decreto da Lavra	1
Data de Aprovação do PRA	8.915

A partir da representativa constatação de dados vazios no atributo *Data de Aprovação do PRA*, foi verificado, especificamente para tais registros vazios, o atributo *Situação Cadastral*, que especifica qual é a situação que um determinado registro se encontra. A Tabela 2 apresenta uma descrição quantitativa das instâncias, onde aproximadamente 83% dos registros encontram-se com situação cadastral ativa, embora não possuam dados referentes à data de aprovação do PRA, o que possivelmente representa uma inconsistência no referido atributo.

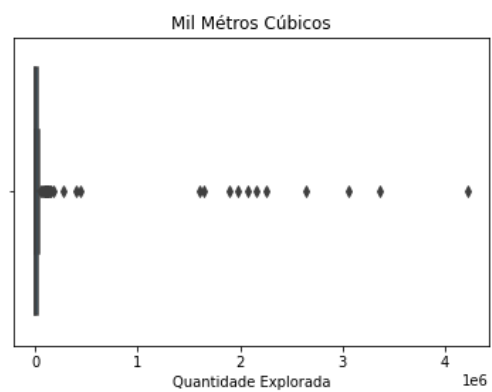
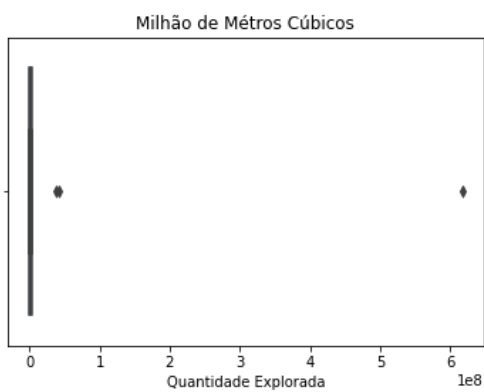
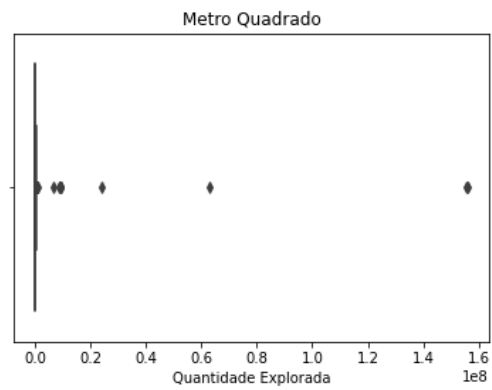
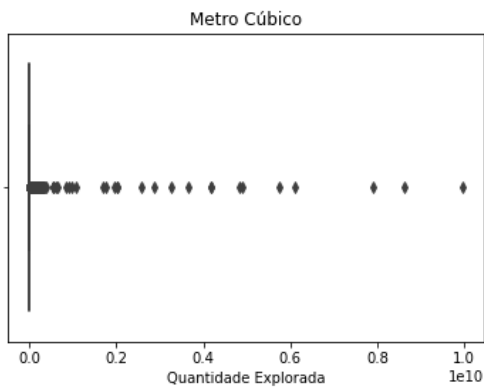
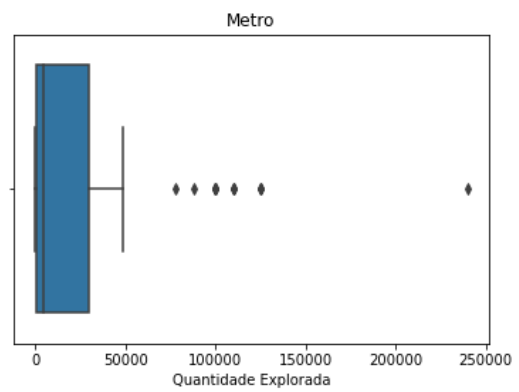
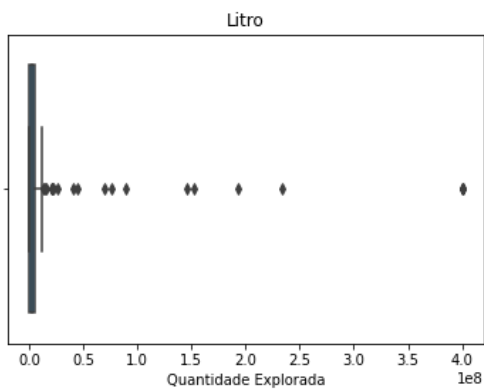
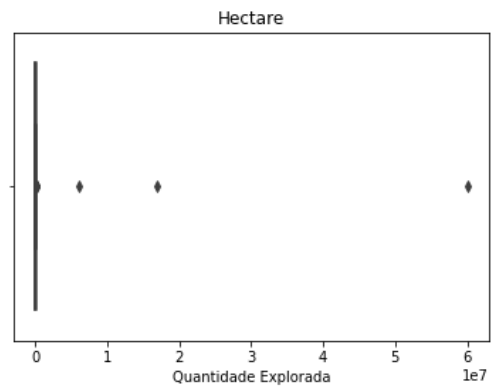
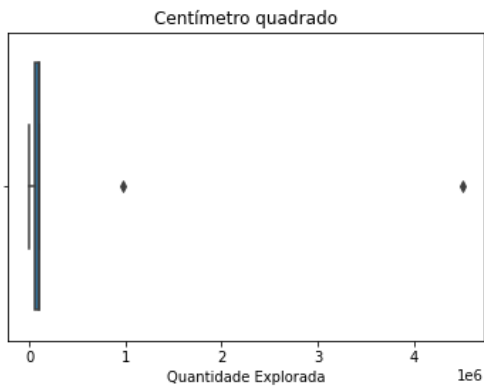
Tabela 2. Quantitativo de instâncias para o atributo *Situação Cadastral* face aos dados ausentes de *Data de Aprovação do PRA*.

Situação Cadastral	Ocorrência
Ativa	7.449
Cadastramento indevido	31
Encerramento de atividades	1.096
Suspensão para Averiguações - dados cadastrais inconsistentes e/ou recadastramento	339

Em termos de verificação de *outliers*, foi considerado especificamente o atributo *Quantidade Explorada*, único que apresenta efetivamente dados numéricos passíveis de utilização em termos de cálculo. Diante disso, inicialmente foi realizada uma avaliação junto ao atributo Unidade de Medida, uma vez que a base possui registros que utilizam diferentes unidades de medida. A Tabela 3 apresenta as unidades de medida que foram identificadas junto à base de dados. Foram identificadas 37 unidades de medida diferentes, entre as quais foram identificados registros de *outliers* em 12, conforme apresentado na Figura 2. Frente à isso, para fins de análise de dados considerando a utilização do KDD enquanto processo de geração de informação, a ocorrência de *outliers* deve ser tratada em concordância com os objetivos de geração de informação e métodos de mineração de dados a serem utilizados, sendo possível a remoção, substituição ou ainda segmentação de dados (AMARAL, 2016).

Tabela 3. Ocorrências de unidade de medida.

Unidade de Medida
'Metro Cúbico', 'Ton. por ano', 'Tonelada', 'Unidade', 'Litro', 'Mil Méetros Cúbicos', 'Milhão de Méetros Cúbicos', 'Tonelada por metro cúbico', 'Metro Quadrado', 'Hectare', 'Centímetro quadrado', 'Metro', 'Metro Cúbico/ha', 'quilogramas', 'Milheiro', 'Bloco', 'Amostra', 'Gramma', 'Gramma por Metro Cúbico', 'Tubo', 'Real', 'Estéreo', 'Metro de Carvão', 'Km <sup>2</sup> por ano', 'Kg por barco ano', 'Litro/ha', 'Estéreo/ha', 'Dia', 'Anuário', 'Carcassa', 'Milhões unidades', 'Livro', 'Rolo', 'Megawatt-hora', 'Tora', 'Miligrama', 'Peça'



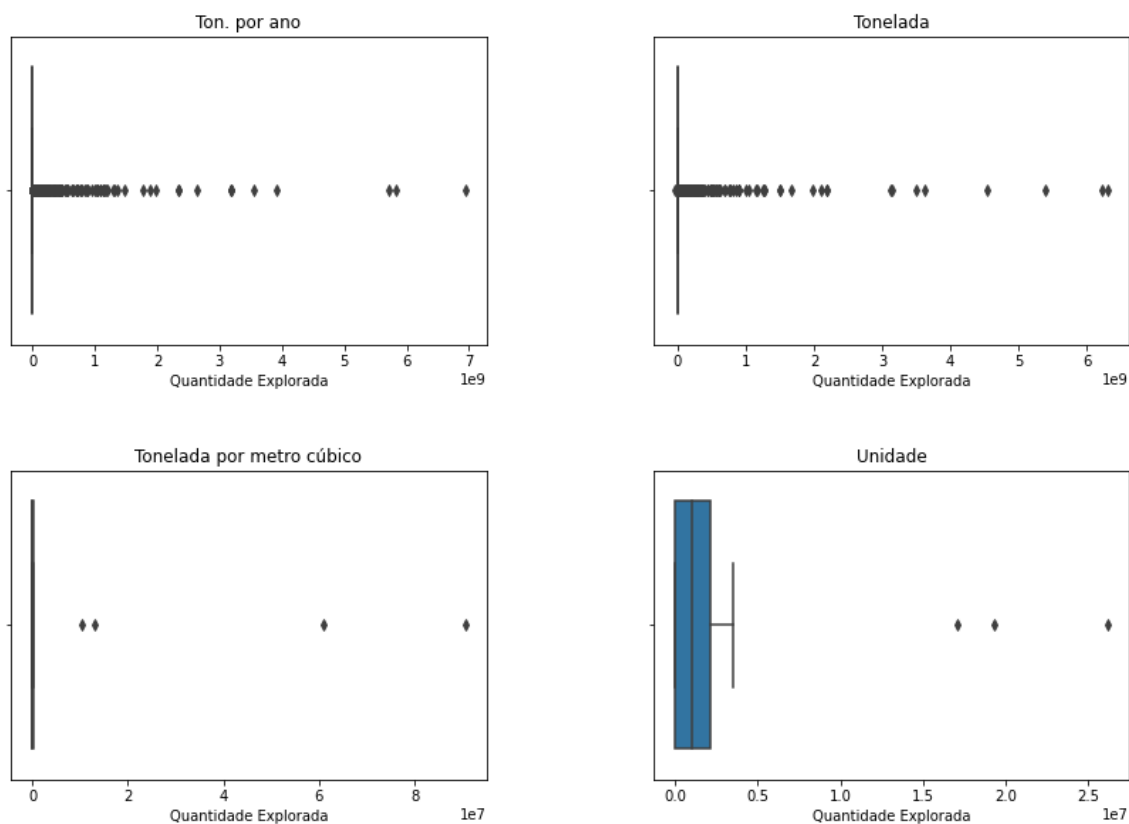


Figura 2. Identificação de *outliers* através do uso de *boxplot*.

Para os atributos categóricos e descritivos foram realizadas avaliações em termos de integridade e homogeneidade. A Tabela 4 apresenta as instâncias únicas para os atributos *Ano de início da exploração* e *Ano de término da exploração*. É possível notar que existem ocorrências que evidentemente destoam dos valores esperados para tais atributos, tais como valores numéricos extremamente baixos ou consideravelmente altos. Diante disso, é necessário realizar uma verificação junto ao órgão responsável pela disponibilização dos registros sobre a assertividade destes campos.

Tabela 4. Ocorrências de possíveis valores inconsistentes.

Atributo	Valores únicos
Ano de início da exploração	0, 1, 2, 3, 10, 11, 12, 79, 80, 81, 90, 96, 101, 102, 104, 106, 109, 116, 200, 201, 207, 307, 405, 406, 410, 509, 602, 603, 604, 606, 609, 620, 901, 910, 912, 1009, 1019, 1086, 1101, 1111, 1199, 1211, 1309, 1311, 1409, 1501, 1504, 1507, 1510, 1605, 1693, 1694, 1704, 1800, 1865, 1885, 1888, 1900, 1901, 1902, 1903, 1906, 1913, 1916, 1918, 1923, 1930, 1934, 1935, 1936, 1937, 1938, 1939, 1940, 1942, 1943, 1944, 1945, 1946, 1947, 1948, 1949, 1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2044, 2070, 2112, 2206, 2212, 2306, 2309, 2311, 2401, 2405, 2411, 2505, 2507, 2801, 3007, 3101, 3105, 5279

---

Ano de término da exploração	0, 2, 3, 5, 8, 10, 11, 12, 20, 47, 49, 50, 52, 120, 201, 202, 205, 210, 214, 215, 225, 1012, 1015, 1053, 1057, 1111, 1112, 1200, 1203, 1235, 1900, 1920, 1929, 1930, 1936, 1938, 1950, 1954, 1960, 1970, 1971, 1974, 1977, 1980, 1981, 1982, 1986, 1988, 1989, 1991, 1992, 1993, 1994, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2092, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2121, 2124, 2125, 2126, 2127, 2128, 2139, 2140, 2143, 2144, 2146, 2147, 2148, 2149, 2150, 2152, 2155, 2157, 2158, 2165, 2178, 2180, 2199, 2200, 2201, 2205, 2209, 2210, 2211, 2219, 2222, 2226, 2228, 2230, 2235, 2237, 2249, 2263, 2266, 2290, 2293, 2300, 2304, 2308, 2338, 2368, 2422, 2424, 2460, 2467, 2469, 2525, 2580, 2612, 2640, 2660, 2672, 2792, 2823, 2870, 2872, 2874, 2900, 2939, 2985, 2986, 2999, 3000, 3010, 3050, 4991, 4999, 9999
------------------------------	---

---

Finalmente foi verificado o atributo Nome do Produto, o qual contém o nome descritivo do minério ao qual cada um dos registros se refere. Com base em observações realizadas sobre o conjunto de instâncias, é possível concluir que tal atributo tem sua alimentação realizada de forma livre, não seguindo um padrão pré-definido. A Tabela 5 apresenta um exemplo retirado da base de dados. Neste sentido, para propósitos de análise, em conformidade com o KDD, na etapa de pré-processamento torna-se necessária a realização de rotinas de tratamento que proporcionem uma integração com base em cálculo de similaridade, por exemplo.

Tabela 5. Não homogeneidade de valores em um mesmo produto.

---

<b>Instâncias</b>
brita
brita e agregados
agregados, brita
BRITA
extração brita
pedra brita

---

#### 4. CONCLUSÃO

Este artigo apresentou um estudo voltado à avaliação da qualidade de dados abertos disponibilizados sobre extração e tratamento de minérios. No estudo foi possível observar de forma clara a necessidade e importância da realização de estudos exploratórios que busquem verificar fatores como integridade, completude e corretude de dados. Na base de dados em questão, foram notáveis os aspectos referentes à não homogeneidade de dados no que se refere ao nome do produto, o que pode ser um atributo crucial para diferentes tipos de análise neste conjunto de dados. Além

disso, a ocorrência de dados ausentes ou ainda possíveis dados inconsistentes se apresenta como um problema desafiador para a garantia da assertividade dos resultados em procedimentos analíticos, neste caso, baseados no uso do KDD.

### 3. REFERÊNCIAS

- AMARAL, F. **Aprenda Mineração de Dados: teoria e prática**. Rio de Janeiro: Alta Books, 2016.
- DEAN, J. **Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners**. New Jersey, USA: Wiley, 2014.
- GOMES, C. F. S.; RIBEIRO, P. C. C. **Gestão da Cadeia de Suprimentos Integrada à Tecnologia da Informação**. São Paulo: Cengage Learning, 2013.
- HUGOS, M. H. **Essentials of Supply Chain Management**. Third edition. New Jersey, USA: John Wiley & Sons, 2011.
- PORTAL BRASILEIRO DE DADOS ABERTOS. **Extração e tratamento de Produtos Mineraiis**, 2021. Disponível em: <https://dados.gov.br/dataset/extracao-e-tratamento-de-produtos-mineraiis>. Acesso em: 13 Jun, 2021.
- SIMCHI-LEVI, D.; KAMINSKY, P. SIMCHI-LEVI, E. **Cadeia de Suprimentos Projeto e Gestão**. Porto Alegre: Bookman, 2010.

### 4. DIREITOS AUTORAIS

Os autores são os únicos responsáveis pelo conteúdo do material impresso incluído no seu trabalho.