

O uso da mineração de dados no tratamento de alarmes industriais

Erivelto Rocha de Souza, Ronan Assumpção Silva

Resumo: Os estudos referentes ao gerenciamento de alarmes, sobretudo na indústria, vêm crescendo a cada dia. Diferentes técnicas para a categorização e tratamento dos alarmes estão sendo feitas pelo auxílio de ferramentas e métodos de Reconhecimento de Padrões. Sendo essa uma estratégia promissora, apresentaremos um estudo de caso para o controle desses alarmes através da ferramenta Weka. Uma base de dados usada em uma indústria foi empregada e com o algoritmo nativo Naive Bayes Multinomial Text, um resultado superior a 80% de acerto foi obtido na classificação de alarmes de acordo com o risco. Este projeto tem a intenção de mostrar a importância da análise de alarmes industriais e os possíveis ganhos utilizando a análise de dados no diagnóstico desses alarmes. Este trabalho também tem o objetivo de auxiliar os operadores para as tomadas de decisões evitando paradas desnecessárias, apoiando o colaborador na tomada de decisões e solicitando a intervenção da equipe de manutenção na resolução das falhas. Através de uma análise do banco de dados podemos sugerir padrões de comportamento de equipamentos e suas falhas. O banco de dados usado é o registro das ações do dia a dia, que geram informações sobre os equipamentos da indústria em estudo.

Palavras chave: Alarmes, Mineração de Dados, Gerenciamento de Alarmes, Classificação, Reconhecimento de Padrões.

The use of data mining in industrial alarm handling

Abstract: Abstract: Studies on alarm management, especially in industry, are growing every day. Different techniques for categorizing and handling alarms are being done through the aid of Pattern Recognition tools and methods. As this is a promising strategy, we will present a case study for the control of these alarms through the Weka tool. A database used in industry was employed and with the native Naive Bayes Multinomial Text algorithm, a result of more than 80% accuracy was obtained in the classification of alarms according to risk. This project intends to show the importance of the analysis of industrial alarms and the possible gains using data analysis in the diagnosis of these alarms. This work also aims to assist operators in decision making by avoiding unnecessary downtime, supporting the employee in decision making and requesting the intervention of the maintenance team in the resolution of failures. Through a database analysis, we can suggest equipment behavior patterns and their failures. The database used is the record of daily actions, which generate information about the equipment of the industry under study.

Keywords: Alarms, Data Mining, Alarm Management, Classification, Pattern Recognition.

1 Introdução

O gerenciamento de alarmes se tornou uma necessidade, principalmente depois de acontecerem vários acidentes. Um caso clássico apontado pelo *Health and Safety Executive* (HSE) foi a explosão da refinaria da Texaco em Milford Haven no ano de 1994 (Leitão, 2018 p2). Um exemplo de um grande acidente que ocorreu pelo grande volume de alarmes gerado sem condições de avaliação dos operadores, gerando enormes prejuízos às indústrias. Diante de ocorrências como esta, em 1999 o *Engineering Equipment and Materials Users Association* (EEMUA) elaborou um guia de boas práticas para o gerenciamento de alarmes. Nele foi apresentado padrões aceitáveis para o monitoramento e tratamento desses alarmes pelos operadores. Como resultado, esperava-se que eles, os operadores, pudessem tomar uma decisão em tempo hábil, evitando desastres ao meio ambiente, pessoas e patrimônio.

Nos deparamos diariamente com vários alarmes, que servem para evitar acidentes. Apesar disso, nem sempre é claro o aviso ou a ação apropriada nesse cenário. No ambiente industrial, não é diferente. Com os avanços tecnológicos quase tudo o que é realizado é monitorado. Portanto, com um grande volume de alarmes, faz-se necessário a classificação dos alarmes e ações na esperança de preservar equipamentos e em especial a vida dos trabalhadores. No ambiente industrial, em especial, demanda a gestão de um conjunto de alarmes sendo disparado simultaneamente. Neste caso, pode ser ainda mais difícil identificar a ação apropriada, que pode ser diferente daquela tomada quando avisos soam separadamente.

Devido ao grande volume de dados gerados, nos deparamos com a tentativa de encontrar padrões úteis, destaca-se a área de Mineração de Dados. A mineração de dados é área na qual a descoberta de informações úteis em grandes depósitos de dados se dá de forma automática (Tang, Steinbach, Kumar, 2009). Portanto, acredita-se que a partir da Mineração de Dados, seja possível entender a natureza dos alarmes industriais coletados a partir de uma multinacional localizada na região dos Campos Gerais. A partir disso, apresentamos com resultados de experimentos envolvendo a classificação alarmes, o que possibilita uma gestão destes alertas com qualidade, segurança e inteligência, auxiliando os envolvidos nas tomadas de decisões.

1.1 Objetivo Geral

Apresentar um estudo de caso, cuja Classificação em Mineração de Dados auxilia o entendimento dos alarmes industriais armazenados em uma base de dados com grande volume de dados.

1.2 Objetivos Específicos

- Analisar como a mineração de dados pode ajudar no gerenciamento de alarmes.
- Identificar padrões de alarmes nas indústrias quanto a ação recomendada pela interpretação de um ou mais alarmes.
- Propor os benefícios da mineração de dados na diminuição dos alarmes gerados.

1.3 Motivação

Como resultado desse trabalho buscamos responder:

- Apresentar uma possível forma de realizar o gerenciamento de alarmes.
- Como a Mineração de Dados pode auxiliar no diagnóstico e seleção de alarmes.

Buscando apresentar um caminho para o tratamento de alarmes, espera-se que o estudo possa oferecer alternativas para a melhoria de um ambiente industrial, principalmente no que se refere ao tópico de alarmes, suas particularidades, seus riscos e formas de prevenção a partir deles.

2 Revisão Literária

Este capítulo tem o objetivo de esclarecer dois tópicos relacionados neste trabalho: alarmes industriais e reconhecimento de padrões, principalmente na técnica de classificação. Encerra-se o capítulo com uma seção que discute a importância desses elementos, sobretudo, a importância de automatizar a análise dos alarmes bem como a ação diante de avisos isolados ou em conjunto.

2.1 Alarmes Industriais

Inúmeros incidentes são provocados anualmente por sobrecarga de alarmes e/ou informações aos operadores (Leitão, 2008), diante desses acontecimentos a EEMUA elaborou um guia de boas práticas para o gerenciamento de alarmes. Cada vez mais procuramos métodos para manter o número de alarmes por operador que sejam gerenciáveis.

Com isso o operador ganha tempo para se concentrar no que é realmente importante, a segurança da planta e das pessoas. Segundo Silva Contreras, 2017, caso exista uma informação que não requeira a execução de uma atividade, esta não deve ser configurado como um alarme.

O sistema de alarme é uma ferramenta vital e produtiva para a gestão de controle de processos industriais (Araújo, 2010), cada vez mais os estudiosos buscam maneiras de garantir uma eficiência no gerenciamento de alarmes devido a sua importância para as indústrias, pessoas e meio ambiente.

2.2 Weka

As técnicas de mineração de dados são organizadas para agir sobre grandes bancos de dados com o intuito de descobrir padrões úteis e recente que poderiam, de outra forma, permanecer ignorados (Tang, Steinbach, Kumar, 2009). Essas técnicas estão presentes em quase todos os segmentos, neste trabalho utilizaremos uma ferramenta de classificação e seus algoritmos, chamada Weka.

O Weka é um software de mineração de dados, com aplicação em Java desenvolvido pela universidade de Waikato na Nova Zelândia, contém ferramentas para preparação, classificação, regressão, clustering, mineração de regras de associação e visualização de dados. (site Weka), um produto com licença GNU.

Weka é uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Ele contém ferramentas para preparação de dados, classificação, regressão, agrupamento, mineração de regras de associação e visualização (WAIKATO, 2019).

Escolhemos este software por sua credibilidade e reconhecimento no universo acadêmico, com uma comunidade sempre atualizando e desenvolvendo algoritmos sempre crescendo a ferramenta de mineração de dados. Utilizaremos para os nossos experimentos o algoritmo Naive Bayes MultiNominal Text, utilizando como opção de teste a referência cruzada que conseguimos um melhor desempenho do algoritmo.

2.2.1 Naive Bayes Multinomial Text

O algoritmo Naive Bayes MultiNomial Text, utiliza o teorema de Bayes. Thomas Bayes foi o primeiro a fornecer uma equação apresentando resultados probabilísticos de um evento A acontecer se outro evento B acontecer, conforme a Equação 1.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Equação 1: Teorema de Bayes

No caso do algoritmo de classificação de texto, de maneira resumida, se tem um atributo com valores textuais e um atributo meta. Cada observação textual presente na base de dados passa por uma etapa de guardar cada palavra encontrada, de modo que não haja a repetição da palavra para um mesmo registro ou até mesmo considerando os demais registros. Como

resultado, essas palavras encontradas são representadas como um vetor de palavras, e passa-se a ver a probabilidade relacionada entre a presença dessas palavras e o atributo meta.

3 Metodologia

Os alarmes tem grande importância para o funcionamento das indústrias. A manutenção e análise destes possibilita o trabalho com segurança, qualidade e respeito ao meio ambiente, diminui o desperdício de dinheiro, aumenta a disponibilidade dos equipamentos, facilita a manutenção de baixo custo e também protege a vida dos trabalhadores. Neste capítulo é apresentado um método para a análise dos alarmes industriais, que procura reconhecer a gravidade do alarme pelas suas características.

O método utilizado neste trabalho está dividido em algumas partes, com o objetivo de facilitar o entendimento do leitor. Este método faz uso da Mineração de Dados para a classificação dos alarmes industriais. Na Figura 1, descreve-se o método como um processo com 5 etapas distintas.

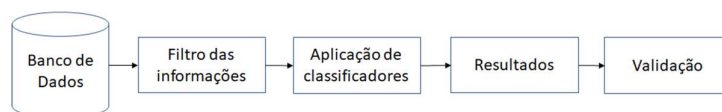


Figura 1. Diagrama das etapas da metodologia.

Na etapa identificada como Banco de Dados, são coletados os dados correspondentes à natureza do problema. Para isso, estaremos utilizando uma base de dados de uma empresa X contendo alarmes reais gerados durante o expediente. Os sinais gerados pelo Sistema Digital de Controle Distribuído (SDCD), são responsáveis pelo funcionamento dos equipamentos industriais. Na Figura 2 é apresentado um exemplo de acionamento de bomba de transferência para um tanque contendo água. A bomba de água (BB), cuja função é esvaziar o tanque (TQ), precisa ter conhecimento de ao menos duas variáveis, tais como: i) o transmissor de nível (LT) está operacional e afere que o nível de água no tanque é suficiente e, ii) a válvula (HV) localizada logo após a bomba (BB) está aberta para o escoamento do líquido. Desta maneira, um alarme é gerado no caso de ativar o funcionamento da bomba (BB), sem que cada condição imposta esteja satisfeita. Portanto, cada equipamento utilizado neste processo é monitorado e alimenta uma base de dados com registros observados em tempo real.

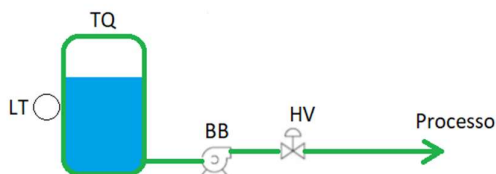


Figura 2. Diagrama Básico

A classificação dos alarmes referentes a uma bomba são diferentes de outros equipamentos que constam na base de dados. Entretanto, cada registro obrigatoriamente é associado a uma classe (LL, LO, STATE, HI, HH ou IOBAD). Para facilitar o entendimento deste sistema de classificação de registros, opta-se por dar continuidade ao exemplo do tanque de água. Com foco no transmissor de nível (LT), cada situação é classificada como:

- a) LL: Alerta LOW-LOW. Significa que o líquido no tanque é muito baixo, pois ocupa entre 0 a 15% do volume total.
- b) LO: Alerta LOW. É acionado quando o líquido total no tanque ocupa entre 15 a 30% do volume total e, portanto, está baixo.
- c) STATE: Alerta periódico para avaliar a disponibilidade e comunicação do equipamento com o SCDC.
- d) HI: Alerta HIGH, serve para indicar que o nível do líquido no tanque está alto, pois ocupa entre 70 a 85 % do volume total.
- e) HH: Alerta HIGH-HIGH. Este aviso demonstra que o líquido no tanque encontra-se num nível muito alto.
- f) IOBAD: Este alerta é acionado sempre que a comunicação com o equipamento falha.

Prosseguindo a segunda etapa da Figura 1, o processo aplica o filtro das informações com o objetivo de padronização. Considera-se da base dados, apenas as características úteis ao processo de classificação dos alarmes. É importante mencionar que os registros pertencentes a base de dados são definidos por projeto, mas nem todos eles precisam fazer parte da extração de conhecimento. Por exemplo, se há um campo com observações genéricas pode ser que atrapalhe na padronização dos registros da base de dados e ainda prejudique a extração de conhecimento por parte do algoritmo de classificação. Neste caso, deve-se tomar uma decisão de remover o atributo ou distribuir esses dados em atributos mais significativos.

Este projeto, que envolve profissionais de múltiplos conhecimentos, pode requerer ajustes diversos. Por exemplo, quais valores registrados em cada equipamento devem corresponder aos alertas LL, LO, STATE, HI e HH. Além disso, dada a natureza e importância de cada equipamento, o tempo de ausência de resposta deve estar corretamente ajustado para classificar o alerta como IOBAD.

Na etapa de aplicação de classificadores, busca-se entender como as características podem ser usadas para a previsão das classes definidas (LL, LO, STATE, HI, HH e IOBAD). Nesta etapa, ferramentas de Mineração de Dados como Weka podem ser empregadas. Uma das tarefas mais comuns e promissoras para o reconhecimento de padrões em uma base de dados é a classificação. Considerando que a base de dados possui muitas características expressas como texto, poucos algoritmos de classificação são capazes de extrair padrões nestas condições se comparados ao número de algoritmos que tratam de dados numéricos. Um dos algoritmos capaz de trabalhar com dados textuais é o Naive Bayes, mas o método não depende deste ou qualquer outro algoritmo de classificação em particular. A escolha do algoritmo nesta etapa se dá principalmente pelas características dos dados. Para relembrar os conceitos de Mineração de Dados, Classificação e Naive Bayes o leitor pode se dirigir ao capítulo 2.

A etapa de resultados refere-se às informações disponibilizadas pela ferramenta utilizada pela classificação. Cada ferramenta pode apresentar diferentes informações, entretanto, na Mineração de Dados, é comum a obtenção da matriz de confusão e a porcentagem de instâncias classificadas corretamente. Das informações que dependem do algoritmo de classificação para serem corretamente interpretadas destacam-se as regras obtidas, que podem ser simples regras como “SE, ENTÃO” extraídas a partir de uma árvore de decisão como a C4.5 até os modelos probabilísticos, cujo Naive Bayes se enquadra.

Finalmente, na etapa de validação verifica-se a qualidade das informações encontradas. O objetivo principal desta etapa é comprovar que os dados coletados na etapa de banco de dados estão corretamente ajustadas. Somente após a validação das informações é que poderá ser sugerido no futuro um sistema automático de controle de alarmes e acionamento de equipamentos com a precisão necessária.

Validação: a partir dos relatórios realizaremos a verificação das informações e validação das informações dos resultados alcançados. Espera-se que modelando o problema na forma de uma base para a classificação e posteriormente propondo um protótipo para tratativa dos alarmes, possa-se sugerir sistemas inteligentes para esse contexto, bem como pode inspirar a tratativa de alarmes como e sugerir trabalhos futuros, tudo em prol da segurança e para a evolução do conhecimento.

Na indústria, é comum fazer a coleta de dados dos eventos por meio de um SDCD, porém, não há uma configuração universal de equipamentos que proporcione segurança e confiabilidade na tomada de decisão após a ocorrência de alertas. Sendo assim, neste capítulo foi apresentado um método para validação dos alarmes industriais.

O método está dividido em 5 etapas, o que facilita o entendimento e o detalhamento das partes no processo. Um dos aspectos mais importantes neste método está na fase de coleta e organização dos dados, aqui chamada de etapa de banco de dados. Tamanho é o impacto da aferição dos equipamentos, que o método deve ser reproduzido sempre que algum ajuste de equipamento afete diretamente na classificação dos alarmes, responsável por traduzir o evento do equipamento na gravidade do alerta.

Somente após a aprovação de todas as etapas deste método é que começa a ser possível a modelagem de um sistema automático para o acionamento de equipamentos baseado em alarmes. Portanto, a correta aferição e funcionamento dos equipamentos, a correta classificação dos registros de alarme, o uso de classificador apropriado, o entendimento das informações do classificador e, finalmente, a aprovação de todas as etapas deste modelo proposto se faz necessário. Para uma maior confiabilidade neste processo, justifica-se o emprego de classificação como uma tarefa de mineração de dados. Esta tarefa passa a ser atraente para avaliar a integridade, qualidade e gravidade dos registros numa base de dados de eventos e alertas.

4 Experimento

Nesta sessão de experimento são descritas as ferramentas, as configurações, detalhamento e discussão acerca do método de trabalho proposto para a tratativa de alarmes industriais. Primeiramente, apresentamos na seção de materiais, a base de dados considerada nesta proposta, bem como a ferramenta e o algoritmo de classificação. Esta seção encerra-se com a discussão acerca da validade do método para o reconhecimento da gravidade do alarme dada a descrição do equipamento.

4.1 Materiais

Inicia-se esta seção com a descrição da base de dados. A base de dados empregada neste experimento contém 1.048.573 registros. São 9 características descritas a seguir:

- a) Composto: Localiza a área que aconteceu o alarme

- b) Tipo: Descrição do tipo de falha que ocorreu
- c) Descrição: Nome do equipamento monitorado
- d) Área: Determina a área da fábrica
- e) Data: Dia e hora que ocorreu o alarme
- f) Valor: Valor de referência para a variável
- g) ValorAlarm: Valor configurado para atuar o alarme
- h) UnidEng: Define o valor da variação Unidade de Engenharia
- i) Bloco: Complemento de composto para definir local de instalação

Dessas características (atributos) é importante mencionar que o “Tipo” é nosso atributo meta para a classificação. Dos vários tipos de alarmes encontrados, propusemos considerar que eles fossem agrupados entre as categorias LL, LO, STATE, HI, HH e IOBAD, conforme está detalhado no capítulo 3, que descreve a Metodologia. Infelizmente, o atributo valor não é sempre preenchido nos registros, o que dificulta ainda mais a análise das informações na busca por padrões de ação baseando-se nos alertas produzidos pelos equipamentos.

Na busca de padrões, é importante que a base de dados seja confiável e que os valores preenchidos sejam condizentes com os valores esperados. Sendo assim, como filtragem dos atributos, optou-se por remover os atributos Composto, Área, Data, Valor, ValorAlarm, UnidEng e Bloco. O atributo “Descrição” é um atributo cujo valor é textual. Logo, há de se recorrer a um algoritmo de classificação que associe o conteúdo do texto ao tipo de alarme encontrado. A explicação de como esse ponto será satisfeito será detalhado na próxima seção.

4.2 A ferramenta e o algoritmo de classificação

Como ferramenta, foi adotada a Weka, produzida pela universidade de Waikato e detalhada na seção de Revisão Bibliográfica. Essa ferramenta possui algoritmos diversos de classificação e algumas opções referentes a experimentos. A que se melhor se adequa a essa proposta é a validação cruzada, também conhecida como *cross validation* e *k-fold*. Este tipo envolve a repetição da classificação por *k* vezes, permitindo que seja estimada a média e o desvio-padrão, por exemplo. Para este experimento, foi usado a validação cruzada com 10 repetições, também conhecida como 10-fold.

Dentre as opções de algoritmo encontradas, optou-se por usar o Naive Bayes Multinomial text. No contexto desta proposta, este algoritmo organiza a característica “Descrição” em um vetor de palavras. Em seguida, busca-se relacionar essas palavras ao “Tipo” de alarme registrado.

4.3 Resultados e discussões

O gerenciamento de alarmes possui grande importância para garantir a segurança as instalações, pessoas e meio ambiente. Com os avanços da tecnologia e o aumento na quantidade de equipamentos monitorados, o aumento do número de alarmes é inevitável, para tanto precisamos garantir o controle e auxiliar os usuários dos sistemas a trabalharem de forma seguro.

Quando trabalhamos com uma quantidade grande de variáveis a mineração de dados atinge bons resultados, a ferramenta Weka tem grande aceitação de na mineração de dados e credibilidade com os seus usuários.

Desta forma quando juntamos a mineração de dados no controle do gerenciamento de alarmes, vamos propor uma solução para otimizar os custos gerados com os acidentes em indústrias, assim podemos conservar vidas e o meio ambiente.

Ao realizar a análise do banco de dados, utilizando o classificador weka e o algoritmo Naive Bayes Multinomial Text, com 1.048.573 de instâncias obtivemos um resultado de 85,94% de classificação assertivas usando uma referência cruzada. Apresentamos a distribuição dos nossos atributos separando conforme tabela 1.

A frequência independente de uma classe	
-2	24751.0
-1	171780.0
0	260657.0
1	176694.0
2	40163.0
3	374534.0

Tabela 1 - Quantidade de atributos

O algoritmo realiza os testes juntando o teorema de bayes com a sistema de laplace e distingue cada um desses atributos individualmente localizando partes em probabilísticas de ocorrer uma falha e mostra na matriz de confusão o quanto o sistema acertou dos testes realizados, conforme tabela 2.

a	b	c	d	e	f	Classificação
11635	635	8058	610	3811	1	a = -2
396	126039	12083	28463	405	4393	b = -1
46	7471	230028	4760	4323	14028	c = 0
890	27264	9921	135810	1546	1262	d = 1
928	2000	3902	3739	29534	5	e = 2
0	1342	4335	725	0	368131	f = 3

Tabela 2 - Matriz de confusão

Na tabela descreve como separamos os nossos alarmes do nosso banco de dados separando os atributos da seguinte forma LL = -2, LO=-1, RESTO=0, HI= 1, HH= 2 e IOBAD= 3. Onde -2 corresponde a um alarme de sinal muito baixo, -1 corresponde a um alarme sinal baixo, 0 corresponde a alarmes não codificados como importantes, 1 corresponde a alarme de sinal alto, 2 corresponde a um sinal muito alto e o 3 corresponde a perda de sinal ou equipamento em falha, utilizamos essa sistemática para classificar nossos alarmes.

5 Conclusão

Conclui-se que a utilização de algoritmos de classificação, apresenta um acerto de superior a 85,94% de acerto para a base de dados utilizada. E demonstrou que podemos aplicar a mineração de dados para o gerenciamento de alarmes para industriais, sempre aprimorando os dados e com possibilidades de aplicar outros algoritmos para obter um valor aceitável de acertos. Identificamos, na nossa base de dados, um número aceitável de acertos no gerenciamento de alarmes industriais.

Conseguindo apresentar alguns benefícios que a mineração de dados pode auxiliar no tratamento de alarmes, elaborando padrões para verificar o tipos de alarmes e auxiliar o operador na tomada das decisões.

Assim esse é o início dos estudos do uso da mineração de dados no gerenciamento de alarmes e como futura pesquisas ficam a utilização online desses dados juntamente com o machine learning aprimorando as definições e equalizando os números de alarmes.

6 Referências

Araújo, E. V. **Gerenciamento de alarmes em plantas industriais: Conceito, normas e estudo de caso em um forno de aquecimento de blocos**. 2010. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Minas Gerais, Belo Horizonte.

DATE, C.J. 1941, **Introdução a sistema de banco de dados** / C.J.Date: tradução de Daniel Vieira - Rio de Janeiro: Elsevier, 2003 - 9 edição

EEMUA (2007), **Alarm systems: A guide to design, management and procurement**, Standard.

ISA (2009), **Ansi/isa-18.2 -2009, Management of alarm systems for the process industries**, Standard, International Society of Automation and American National Standards Institute.

Silva Contreras, Isabelle, **Gerenciamento de Alarmes em Plantas Industriais** - Isabelle Silva Contreras. – Rio de Janeiro: UFRJ/Escola Politécnica, 2017.

Leitão, Gustavo. **O sistema de alarme é uma ferramenta vital e produtiva para a gestão de controle de processos industriais**. Gustavo Bezerra Paz Leitão .Natal-RN - 2008

Leitão, Gustavo. **Classificação on-line de situações anormais em operação de processos industriais baseada em processamento de alarmes e variáveis de processos**-Gustavo Bezerra Paz Leitão. - Natal-RN 2018.

TAN, Pang - Ning; STEINBACH, Michael; KUMAR, Vipin **Introdução ao DATAMINING Mineração de Dados**, Rio de Janeiro: Editora Ciência Moderna Ltda., 2009.

WAIKATO, U. O. **WEKA**. <http://www.cs.waikato.ac.nz/ml/weka/> , acessado Maio de 2019.