

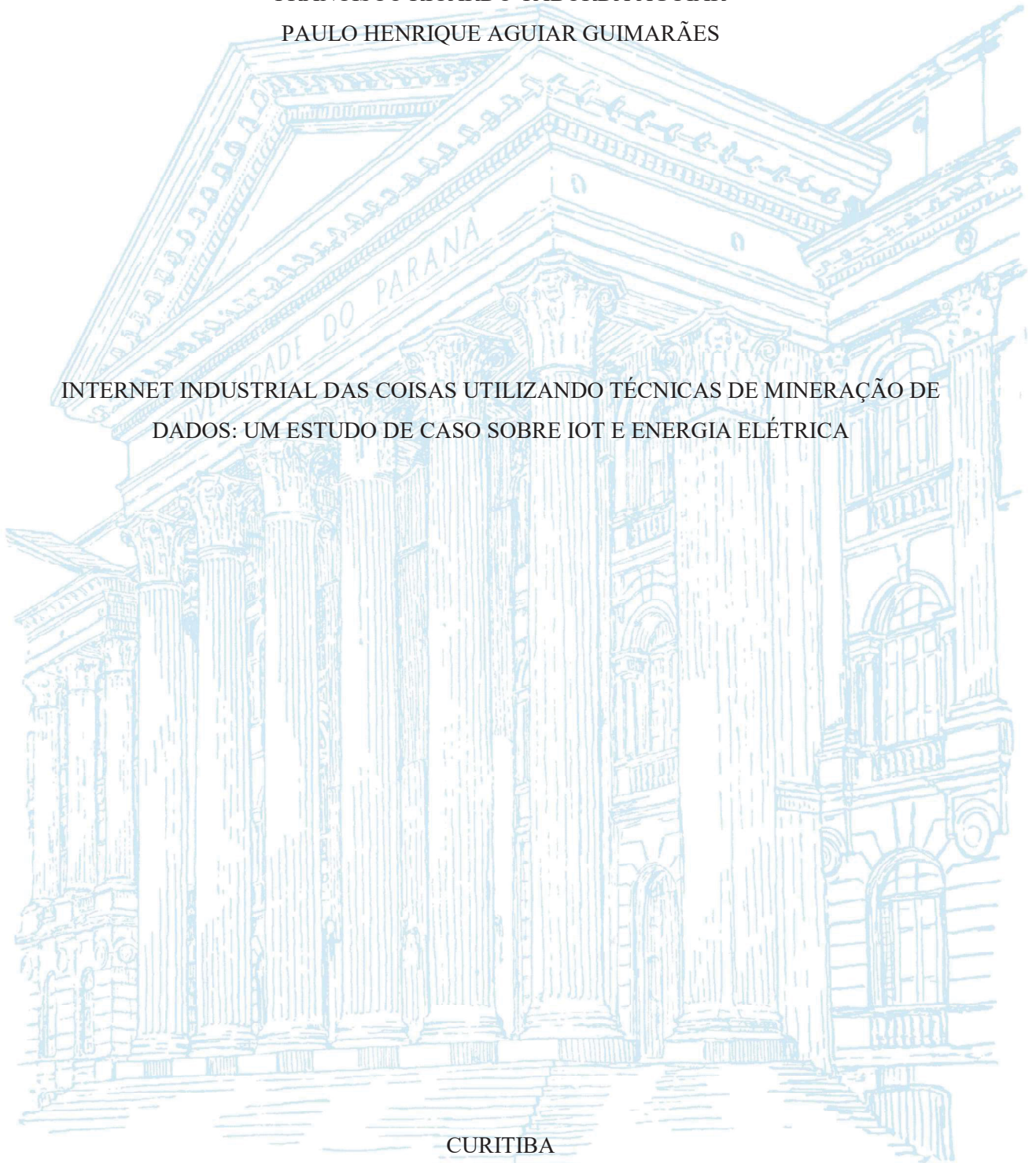
UNIVERSIDADE FEDERAL DO PARANÁ

FRANCISCO RICARDO TABORDA AGUIAR
PAULO HENRIQUE AGUIAR GUIMARÃES

INTERNET INDUSTRIAL DAS COISAS UTILIZANDO TÉCNICAS DE MINERAÇÃO DE
DADOS: UM ESTUDO DE CASO SOBRE IOT E ENERGIA ELÉTRICA

CURITIBA

2020



FRANCISCO RICARDO TABORDA AGUIAR
PAULO HENRIQUE AGUIAR GUIMARÃES

INTERNET INDUSTRIAL DAS COISAS UTILIZANDO TÉCNICAS DE MINERAÇÃO DE
DADOS: UM ESTUDO DE CASO SOBRE IOT E ENERGIA ELÉTRICA

Artigo apresentado ao curso de Pós-Graduação em Engenharia Industrial 4.0, Setor de Tecnologia Mecânica, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Especialista em Engenharia Industrial 4.0.

Orientador: Prof. Pablo Deivid Valle

CURITIBA

2020



INTERNET INDUSTRIAL DAS COISAS UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS: UM ESTUDO DE CASO SOBRE IOT E ENERGIA ELÉTRICA

Paulo Henrique Aguiar Guimarães

Francisco Ricardo Taborda Aguiar

Universidade Federal do Paraná, Av. Cel. Francisco H. dos Santos, 100 - Jardim das Américas, Curitiba - PR, 81530-000
paulohenrique.guima@gmail.com, francisco.ricardot@gmail.com

Resumo. *A Internet Industrial das Coisas (IoT Industrial) tem sido um tópico muito popular sobre os conceitos 4.0. É composta por uma infinidade de dispositivos conectados por software de comunicação. Os sistemas resultantes, e até os dispositivos individuais que o compõem, podem monitorar, coletar, trocar, analisar e agir instantaneamente sobre as informações para alterar inteligentemente seu comportamento ou ambiente - tudo sem intervenção humana. O objetivo desse artigo é apresentar uma das propostas de software de comunicação no ambiente industrial, com IoT, em meio a essa diversidade de protocolos de comunicação e topologias, nos quais divergem de atuador para atuador existente em cada setor industrial. A proposta é utilizar um multimedidor de grandezas elétricas para extrair dados dos atuadores, em tempo real, processar estes dados utilizando técnicas de mineração de dados e apresentá-los em forma de informação por meio de um sistema desenvolvido.*

Palavras chave: *IoT. Grandezas elétricas. Mineração de dados.*

1. INTRODUÇÃO

A Internet Industrial das Coisas já é uma realidade. Um exemplo de um desses sistemas de IoT Industrial seria um parque eólico. Cada turbina eólica está equipada com muitos sensores e softwares que permitem se comunicar com todas as outras turbinas eólicas e até solicitar reparos. Para otimizar a geração de energia, cada turbina eólica ajusta automaticamente suas configurações e comportamento com base nos dados que recebe e processa do sistema, como alterações na velocidade e direção do vento (Industries, RTI 2015). Neste artigo descreve-se com mais detalhes a realidade IoT Industrial no mundo e no Brasil, assim como quais são os obstáculos que impedem o avanço desta tecnologia em determinados setores industriais, nos quais favoreceram em desenvolver um sistema que pode acelerar e despertar o interesse em empresas que necessitam de IoT Industrial.

No entanto, em um ambiente industrial, rodeado de normas, fornecedores, topologias, protocolos de comunicação diferentes e etc., o conceito de IoT Industrial tem sido deixado de lado por conta da falta de confiabilidade, custo e também em alguns casos devido à complexidade de implementação, primeiro pela maneira que os dados são extraídos pelos atuadores e segundo pelo local, no qual estes dados são armazenados.

Dependendo de qual dado deseja-se extrair de um atuador, por exemplo produtividade, é possível utilizar sensores ou até mesmo extrair informações obtidas dos logs gerados pelo próprio software do atuador. Isso se o atuador tiver placa de rede, no qual possibilita a conexão para leitura. Para obter um dado vindo do software do atuador requer compromisso com o fabricante, visto que é necessário compreender a arquitetura do software ou até mesmo solicitar permissão de acesso. Em atuadores sem placas de rede, a obtenção de um determinado dado requer mais trabalho ou ainda é preciso realizar um *retrofit* no equipamento.

Uma vez obtido o dado desejado, o próximo passo é definir o local de armazenagem, isto é, o banco de dados. Existem servidores virtuais Cloud, nuvem, que permitem configurar banco de dados sem precisar de um servidor físico. Apesar da segurança e meios de criptografia existentes atualmente, não são ainda suficientes para convencer empresários a implementarem IoT em seus atuadores utilizando essa abordagem (WEFUSA, 2015). Isto porque transformar uma empresa em 4.0 requer um processo gradual e não exponencial. Essas e outras características de um ambiente industrial não apenas brasileiro, mas como também internacional, é que tem dificultado a implementação de industrial IoT como um simples produto ou serviço.

Um cenário industrial com uma miscelânea de atuadores com diferentes protocolos e níveis de tecnologia é uma realidade. É necessário, portanto, um sistema de extração de dados flexível, isto é, que seja capaz de se adaptar a quaisquer circunstâncias e seja escalonável de acordo com o avanço tecnológico da empresa, ou seja, independente de fabricante, protocolo de rede, de comunicação, de nível tecnológico do atuador e que não necessariamente dependa de um servidor virtual na internet. Não seja invasivo. Um serviço pequeno que atende o grande e o pequeno seja grande.

Este artigo apresenta uma proposta para esse problema. Utilizando um multimedidor de grandezas elétricas em qualquer atuador é possível extrair dados, que atualmente são utilizados para fins de manutenção preditiva do equipamento, e processá-los, por meio das técnicas de mineração de dados, em um sistema capaz de interpretar os valores elétricos em outras informações, tais como produtividade, eficiência energética, análise de redução de custos indiretos,

etc., além de complementar as informações de manutenção como curto entre espiras, bobinas, desequilíbrio de fase (maior consumo de energia e menor vida útil do motor), enrolamentos abertos (sem condutividade) e entre outros. O sistema é projetado a conectar com o multimedidor e os dados obtidos podem ser salvos tanto em servidores *on premise* ou nuvem.

Os princípios de mineração de dados são estabelecidos há muito tempo e bem entendido. No entanto, o ressurgimento da popularidade devido a sucessos recentes no campo de aprendizagem de máquina (Silver et al., 2016) abriu o caminho para esta poderosa ferramenta a ser adotada pela indústria. A metodologia "Indústria 4.0", apresentada em 2013 à feira mundial de Hamburgo (Zühlke, 2014) se transformou em um objetivo único. O termo, cunhado pelo governo alemão para descrever sua visão contínua de fabricação, é multifacetado e muitas vezes aberto à interpretação. Existe um tema claro de manufatura inteligente; que faz uso de tecnologias computacionais avançadas e os avanços em sistemas digitais e processos de aprendizado de máquina apoiam à tomada de decisão, tais como em redes de controle distribuído auto-suficiente capaz de se auto-ajustar e auto-corrigir caso surjam problemas. Através deste trabalho, se desenvolveu uma metodologia adotando técnicas de mineração de dados para tomar decisão com relação às grandezas elétricas medidas do atuador.

A máquina utilizada para este trabalho foi a *MDI FLEX*, Imagem Direta com Micro Espelho, do inglês Micro Mirror Direct Image, destinada à manufatura de placas de circuitos impressos flexíveis.

2. REVISÃO BIBLIOGRÁFICA

Para a realização desse trabalho foi necessário o estudo da literatura para validar os conceitos utilizados, bem como para analisar com melhor visão crítica os resultados obtidos nos ensaios.

2.1 IoT Industrial

A Internet industrial das coisas (IIoT) conecta máquinas e sensores críticos em setores de alto risco, como saúde, energia, transporte e controle industrial. Segundo RTI (Industries, RTI 2015), que é uma das empresas mais influentes em Internet Industrial das Coisas de acordo com estudo realizado pela Appinions e publicado na Forbes, IIoT são sistemas nos quais a falha geralmente resulta em situações de risco de vida ou outras situações de emergência. Por outro lado, os sistemas de IoT tendem a ser dispositivos no nível do consumidor, como termômetros domésticos inteligentes e alimentadores automáticos de animais de estimação. Eles são importantes e convenientes, mas não criam imediatamente situações de emergência.

A segunda perspectiva em (Accenture, 2015) vê a IoT industrial como a infraestrutura que deve ser construída antes que os aplicativos da IoT possam ser desenvolvidos. Em outras palavras, a IoT, até certo ponto, depende da IoT industrial. Por exemplo, muitos eletrodomésticos em rede podem ser classificados como *gadgets* de IoT, como um refrigerador que pode monitorar as datas de vencimento do leite e ovos que ele contém e sistemas de segurança doméstica programáveis remotamente. No lado da Internet industrial, as concessionárias estão permitindo um melhor balanceamento de carga, tomando decisões de gerenciamento de energia até o nível da vizinhança. Os inovadores estão apenas começando a imaginar as possibilidades que podem ser alcançadas tirando proveito de dispositivos e sistemas que podem se comunicar e agir em tempo real, com base nas informações que eles trocam entre si. A medida que a IoT industrial se torna mais bem definida e desenvolvida, aplicativos IoT mais impactantes podem e serão criados.

Há uma diferença, no entanto, entre a disponibilidade dessas tecnologias e capitalizar todo o seu potencial aplicando-os efetivamente dentro de organizações através de toda cadeia de suprimentos e em vários setores. Se a IIoT é para gerar economia e crescimento, os líderes de empresas e países tem que pensar não convencionalmente como as regras de concorrência e produção. A Accenture estima que a IIoT elevará o produto interno bruto real (isto é, ajustado pela inflação) em 1,0% em 2030 sobre as projeções de tendência para 20 grandes economias estudadas. E se essas economias aumentarem seus investimentos em IIoT em 50 por cento e melhorar a subjacente condição propícias em seus respectivos países, isso pode subir de 1,0% para 1,5 por cento naquele ano (Accenture, 2015).

Os estudos da Accenture estimaram que em 2030, Brasil, Rússia, Índia e China (BRICs) terão correspondentes incrementos para o PIB anual de 0,2% em média e, com investimentos aprimorados e medidas, 0,5 por cento. Em particular, a Accenture estima que China parece pronta para ver maior crescimento econômico ganhos da IIoT do que a Rússia, a Índia ou Brasil. Com as medidas de melhoria da IIoT, a China poderia ver seu PIB de US \$ 1,8 trilhão em 2030, aumentando em 1,3%. A economia dos Estados Unidos ganhará US \$ 6,1 trilhões em PIB acumulado até 2030. Ao tomar medidas adicionais, por exemplo, para melhorar a infraestrutura de banda larga do país, esse número pode subir para US \$ 7,1 trilhões. Isso pode significar que o PIB anual dos Estados Unidos é de 2,3% em projeções superiores às tendências em 2030. Incluindo esforços e investimentos adicionais, Alemanha poderia elevar seu PIB acumulado em US \$ 700 bilhões e no Reino Unido 531 bilhões de dólares nos próximos 15 anos. Isso aumentaria o consumo interno bruto alemão e mercado interno bruto do Reino Unido 1,7 por 1,8 por cento, respectivamente, em 2030 sobre as projeções de tendências.

Por outro lado, em relação agora a disponibilidade da tecnologia, realizar todo o potencial da IIoT, empresas e governos precisará superar uma série de obstáculos importantes (WEFUSA, 2015). Chefe deles são segurança e privacidade de dados, que já estão ganhando importância devido ao aumento das vulnerabilidades a ataques, espionagem e violações de dados impulsionadas pelo aumento conectividade e compartilhamento de dados. Segundo WEFUSA, a

cibersegurança se concentrava em número limitado de pontos finais. Com o advento da IIoT, essas medidas não serão mais adequadas, pois os mundos físico e virtual se combinam em larga escala. As organizações precisarão de novas estruturas de segurança que abranjam toda a pilha física cibernética, desde a autenticação e aplicação no nível do dispositivo de segurança, a garantia de todo o sistema, resiliência e modelos de resposta a incidências. Outra barreira crucial é a falta de interoperabilidade entre os sistemas existentes, o que aumentará significativamente a complexidade e os custos de implantações da IIoT.

Está claro que a IIoT deve ser implementada por uma arquitetura cibernética / de segurança física. Um sistema bem pensado. Esta meta pode ser proposta com técnicas não invasivas para interagir remotamente com os atuadores (Simon, Toby 2017). E tendo observado essa lacuna no ambiente industrial o objetivo deste artigo, baseado em IIoT, é extrair os dados do atuador por meio das grandezas elétricas, analisá-los e transformá-los em informação que agregue valor.

2.2 Rede industrial

A extração de dados de um atuador ou sensor requer normalmente um conhecimento prévio de como é a arquitetura de rede do determinado ambiente industrial. As redes industriais são tipicamente distribuídas por natureza e variam consideravelmente em todos os aspectos, incluindo as características da camada de enlace, protocolos de rede usados e topologia. Topologia é um termo genérico para descrever conexões de dispositivos em uma rede. O termo não é exclusivo para rede industrial, embora neste artigo o foco foi somente neste ramo. A topologia de rede, especificamente, é o layout da rede. Em redes industriais, opções de topologia com fio incluem: barramento linear, árvore, estrela e anel. Outras opções de topologia são abertas com conexões sem fio, tais como a identificação por radiofrequência, ou RFID (Radio Frequency IDentification), *bluetooth*, *3G*, etc. (Tanenbaum, 2011). As redes Ethernet e IP (*Internet Protocol*) são onipresentes e podem ser implementadas em qualquer número de topologias - incluindo estrelas, árvores e até topologias de malha completa. Topologias, como anéis, suportam facilmente a redundância necessária, comumente necessária em redes industriais. Uma topologia de barramento representa um domínio de transmissão de mensagens compartilhadas, onde muitos nós estão competindo por uma quantidade finita de largura de banda e confiando na coordenação do tráfego ou na comunicação síncrona para fornecer a melhor conectividade possível (Thomas, 2011). Muitas arquiteturas dos sistemas industrial de controle, ou ICS (*Industrial control systems*), são baseadas em tecnologias subjacentes, como publicação-assinatura e *token-rings* encapsuladas em pacotes com o protocolo de datagrama do usuário, ou UDP (*User Datagram Protocol*), adequados para tecnologias de barramento.

Nas redes empresariais modernas, no entanto, isso é impraticável. A Ethernet fornece um segmento Ethernet dedicado, com largura de banda garantida "primeiro salto" associada a todos os nós, e se tornou uma mercadoria, tornando as topologias em estrela extremamente comuns. Nas redes industriais, Tanenbaum, é mais comum que os computadores de acesso sejam conectados em uma configuração em anel enquanto uma topologia em estrela é usada para conectar-se aos dispositivos finais. Ainda existe uma forte necessidade de topologias de barramento e anel em redes industriais, dependendo do tipo específico de processo de controle que está em operação e dos protocolos específicos que são usados. Em ambientes industriais que dependem da comunicação com fio para garantir a confiabilidade, pode ser proibitivo o custo de implementar topologias de malha nas configurações tradicionais de barramento e anel. As redes de malha tornaram-se o padrão de fato para redes industriais sem fio. Por exemplo, um processo de controle automatizado para higienizar a água pode usar uma topologia de barramento com o protocolo PROFIBUS-PA, enquanto outro processo de controle pode usar Modbus / TCP em uma topologia em anel para controlar sistemas de bombeamento ou filtragem. À medida que se afastam do processo e aproxima-se da rede de negócios, os projetos de TI "típicos" se tornam mais predominantes, a ponto de muitas redes de plantas estarem instaladas.

No entanto, um sistema de internet industrial que depende de todo conhecimento prévio de uma rede industrial para iniciar os primeiros passos, certamente não tende a ser escalável, visto que a medida conforme avança a tecnologia da empresa, a rede industrial caminha junto com esse processo e no entanto é preciso um sistema IoT escalonável, ortogonal, capaz de se adaptar em quaisquer cenário industrial.

2.3 Exposição de imagem direta

O conceito de imagem direta já existe há algum tempo, a partir do início dos anos 80 (Pcb Magazine, 2017). Pioneiros foram a Excellon Automation e a Eocom, em Orange County. Naquela época, segundo o artigo, o setor estava entrincheirado com imagens fotográficas de filme seco e, em pelo menos no mercado norte-americano, houve pouco interesse em ser pioneiro em um conceito tão revolucionário quanto a geração de imagens diretas, juntamente com os problemas relacionados a salas limpas e outros aspectos ambientais. O filme seco foi visto como um aplicação robusta. Como a tecnologia digital começou a penetrar outras aplicações, por exemplo, câmeras digitais, a visão de reduzir custos e melhorar a qualidade da imagem começou a se estabelecer. Mas o equipamento era, e ainda é, muito caro e requer contratos de serviço de alto custo.

No entanto, fabricante após fabricante começou a adotar as ofertas mais recentes de produtos para imagens diretas, nos quais apresentam diferenças de tecnologias intrínsecas, tais como tipos de comprimento de onda através de Led ou Laser, qualidade da exposição de imagem, métodos de centralização e entre outros. Todavia, o processo de exposição é

semelhante, isto é, faz a leitura das coordenadas do arquivo de CAM, manufatura assistida por computador do inglês *Computer-aided Manufacturing*, se determina energia de exposição de acordo com o material e expõe-se a imagem. O sistema de exposição, após ter calculado as coordenadas da imagem junto com as variáveis mencionadas, é movimentado até o término do processo. Neste instante que é possível determinar a quantidade de imagens expostas podendo calcular, por sua vez, a produtividade do atuador.

Neste trabalho analisou-se este processo através da energia elétrica para tomar decisões, utilizando técnicas de mineração de dados, a respeito dos valores medidos.

2.4 Multimedidor de energia

O multimedidor de energia (Fig. 1) é um instrumento destinado à medição de consumo de energia e outras grandezas elétricas tais como corrente, tensão, potências e etc. Uma vez conectado no atuador, os dados extraídos podem ser adquiridos de diversas maneiras, dependendo do tipo de medidor, tais como por meio de protocolo MQTT (*Message Queuing Telemetry Transport*), *bluetooth*, *Wi-fi* e etc.



Figura 1. Multimedidor Fluke 1732 (Fluke, 2020).

2.5. Mineração de dados

Em geral, há um conjunto de dados de exemplos (chamados de instâncias), cada um dos quais compreende os valores de várias variáveis, que na mineração de dados são frequentemente chamadas de atributos. Existem dois tipos de dados, que são tratados de maneiras radicalmente diferentes. Para o primeiro tipo, existe um atributo especialmente designado e o objetivo é usar os dados fornecidos para prever o valor desse atributo para instâncias que ainda não foram vistas (Bramer, 2016). Dados desse tipo são chamados de *labels*, rótulos.

O processo de mineração de dados é uma *pipeline* que contém muitas fases, como limpeza de dados, extração de recursos e design algorítmico (Aggarwal, 2015). O fluxo de trabalho de um aplicativo típico de mineração de dados contém as seguintes fases:

1. Coleta de dados: a coleta de dados pode exigir o uso de hardware especializado, como um rede de sensores, trabalho manual, como a coleta de pesquisas com usuários ou ferramentas de software como um mecanismo de rastreamento de documentos da Web para coletar documentos. Enquanto esta etapa é altamente específico do aplicativo e geralmente fora do domínio do analista de mineração de dados, é extremamente importante porque boas escolhas nesta fase podem impactar significativamente o processo de mineração de dados. Após a fase de coleta, os dados geralmente são armazenados em um banco de dados ou, em geral, um *data warehouse* para processamento.

2. Extração de recursos e limpeza de dados: quando os dados são coletados, eles geralmente não são em um formato adequado para processamento. Por exemplo, os dados podem ser codificados em logs complexos ou documentos de formato livre. Em muitos casos, diferentes tipos de dados podem ser arbitrariamente misturados em um documento de forma livre. Para tornar os dados adequados para processamento, é essencial transformá-los em um formato que seja amigável aos dados algoritmos de mineração, como multidimensionais, séries temporais ou formato semiestruturado. O formato multidimensional é o mais comum, no qual diferentes campos dos dados correspondem às diferentes propriedades medidas que são chamadas de recursos, atributos ou dimensões. É crucial extrair recursos relevantes para o processo de mineração. A fase de extração de recursos geralmente é realizada em paralelo com a limpeza de dados, onde partes ausentes e incorretas dos dados são estimadas ou corrigidas. Em muitos casos, os dados podem ser extraídos de várias fontes e precisam ser integrados em um formato unificado para processamento. O resultado final deste procedimento é uma ótima conjunto de dados estruturados, que pode ser efetivamente usado por um programa de computador. Depois da fase de extração de recursos, os dados podem ser armazenados novamente em um banco de dados para processamento.

3. Processamento analítico e algoritmos: A parte final do processo de mineração é projetar métodos analíticos eficazes a partir dos dados processados. Em muitos casos, pode não ser possível usar diretamente um problema padrão de mineração de dados. O processo geral de mineração de dados é ilustrado na Figura 2, onde o bloco analítico mostra vários blocos de construção que representam o design da solução para uma determinada inscrição.

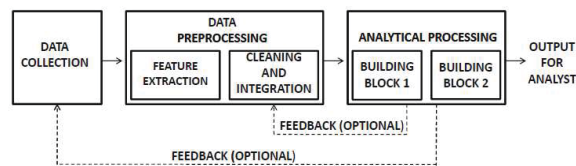


Figura 2: Pipeline do processo de mineração (Aggarwal, 2015).

Muitos problemas de mineração de dados são direcionados para uma meta especializada que às vezes é representada pelo valor de um recurso específico nos dados. Esse recurso específico é referido como o rótulo da classe. Portanto, esses problemas são supervisionados, em que as relações dos recursos restantes nos dados em relação a esse recurso especial são aprendidos. Os dados usados para aprender esses relacionamentos é conhecido como dados de treinamento. O modelo aprendido pode em seguida, ser usado para determinar os rótulos de classe estimados para registros, onde o rótulo está ausente. Portanto, é construído um modelo de treinamento, que é então usado para prever rótulos de classe. Para que esse fluxo possa ser visto facilmente é necessário um software com algoritmos de aprendizado de máquina como o WEKA.

O ambiente de trabalho WEKA é uma coleção de algoritmos de aprendizado de máquina e ferramentas de pré-processamento de dados que inclui vários algoritmos. Ele foi projetado para que seja possível testar rapidamente os métodos existentes em novos conjuntos de dados de maneira flexível. Ele fornece amplo suporte para todo o processo de mineração de dados experimental, incluindo a preparação dos dados de entrada, a avaliação estatística dos esquemas de aprendizado e a visualização dos dados de entrada e do resultado do aprendizado (Witten et. al, 2016).

Além de uma ampla variedade de algoritmos de aprendizado, inclui uma ampla variedade de ferramentas de pré-processamento. Este conjunto de ferramentas diversificado e abrangente é acessado através de uma interface comum para que seus usuários pode comparar diferentes métodos e identificar os mais adequados para o problema em questão. O WEKA foi desenvolvido na Universidade de Waikato, na Nova Zelândia; o nome significa Ambiente Waikato para Análise de Conhecimento do inglês *Waikato Environment for Knowledge Analysis*. O sistema é escrito em Java e distribuído sob os termos da GNU General Public License. Ele roda em praticamente qualquer plataforma e foi testado nos Sistemas Operacionais Linux, Windows e Macintosh.

O WEKA fornece implementações de algoritmos de aprendizado que você pode aplicar facilmente ao seu conjunto de dados. Você pode pré-processar um conjunto de dados, alimentá-lo em um esquema de aprendizado e analisar o classificador resultante e seu desempenho; tudo sem escrever nenhum código de programa.

Alguns algoritmos de aprendizado, no WEKA, incluem do tipo árvore de decisão. Os algoritmos da árvore de decisão (por exemplo, ID3, C4.5 ou J48 e CART) foram originalmente destinados à classificação. A indução da árvore de decisão constrói uma estrutura semelhante ao fluxograma em que cada nó interno (não-folha) denota um teste em um atributo, cada ramo corresponde a um resultado do teste e cada nó externo (folha) indica uma previsão de classe. Em cada nó, o algoritmo escolhe o atributo "melhor" para particionar os dados em classes individuais. Quando a indução da árvore de decisão é usada para a seleção de subconjuntos de atributos, uma árvore é construída a partir dos dados fornecidos. Todos os atributos que não aparecem na árvore são considerados irrelevantes. O conjunto de atributos que aparecem na árvore formam o subconjunto reduzido de atributos (Han et. al, 2012).

O ambiente de trabalho inclui métodos para os principais problemas de mineração de dados: regressão, classificação, clustering, mineração de regras de associação e seleção de atributos. Conhecer os dados é parte integrante do trabalho, e são fornecidos muitos recursos de visualização de dados e ferramentas de pré-processamento de dados. Todos os algoritmos recebem suas entradas na forma de uma única tabela relacional que pode ser lida em um arquivo ou gerada por uma consulta ao banco de dados. Uma maneira de usar o WEKA é aplicar um método de aprendizado a um conjunto de dados e analisar sua saída para aprender mais sobre os dados. Outra é usar modelos aprendidos para gerar previsões sobre novas instâncias.

Embora WEKA reconheça outros formatos de arquivos para importação, como formato de valores separados por vírgula, do inglês *Comma-separated values* (CSV), por exemplo, Witten menciona que o formato final será ARFF, pois é nesse formato que permite definir regras tais como de associação e seleção de atributos, como mencionado. Um arquivo ARFF, Formato de Arquivo Atributo-Relação do inglês *Attribute-Relation File Format*, é um arquivo de texto ASCII que descreve uma lista de instâncias que compartilham um conjunto de atributos.

2.6 Grandezas Elétricas

Também foram vistos dados na literatura relacionados às grandezas elétricas para uma melhor compreensão dos dados medidos pelo multimedidor.

2.6.1 Distorção harmônica total

A distorção harmônica total do inglês *Total harmonic distortion* (THD) é uma figura importante usada para quantificar o nível de harmônicos nas formas de onda de tensão ou corrente. A THD, é uma medida comum do nível de distorção harmônica presente nos sistemas de potência. O THD é definido como a razão entre harmônicos totais e o valor nas frequências fundamentais (Schmilovitz, 2005):

$$THD_v = \frac{\sqrt{\sum_{n=2}^{\infty} V_n^2}}{V_1} \quad (1)$$

Onde V_n é a tensão rms (*root mean square*) ou a amplitude harmônica e $n = 1$ é a frequência fundamental.

$$THD_i = \frac{\sqrt{\sum_{n=2}^{\infty} I_n^2}}{I_1} \quad (2)$$

Onde I_n é a corrente rms (*root mean square*) ou a amplitude harmônica e $n = 1$ é a frequência fundamental.

3. MATERIAIS E MÉTODOS

Foi utilizado o multimedidor *Fluke Energy 1732* (Fig. 1), em testes realizados na Alemanha, na empresa Schmolz Maschinen GmbH. A máquina utilizada para teste foi a MDI FLEX (Micro Mirror Direct Image), que é destinada à exposição de imagens para placas de circuitos impressos flexíveis.

A exposição foi realizada no modo infinito, ou seja, permitindo que o sistema de exposição repita o processo ininterruptamente, assim a quantidade de dados seja maior para análise. Além disso, foi também determinado 400mJ de energia dos LED's, o que é considerado um valor alto para expor imagens simples (Fig. 3), deste modo o sistema de exposição consome mais potência elétrica e, conseqüentemente, se torna de fácil compreensão as grandezas elétricas medidas pelo multimedidor durante o processo e claro, para o teste de mineração de dados.



Figura 3. Imagem utilizada para teste (O autor, 2020).

A imagem de teste foi desenvolvida no software Cam PCB Investigator com 600mm de comprimento e 60mm de largura. O comprimento foi proposital para que o sistema de exposição percorra totalmente a área da mesa e assim quando a última coordenada de exposição for exposta, todo o sistema volta para o ponto de origem, o que por sua vez inicia um novo processo de exposição. Este íntere é o que se pode determinar como tempo de setup do sistema de exposição e além disso, se espera um consumo de energia menor do que durante o processo como já mencionado anteriormente.

A imagem deve ser exportada pelo PCB Investigator no formato Gerber (Ucamco, 2019), Tab. 1, que gerou mais de 25 mil linhas de coordenadas para esta imagem, e assim é possível carregar na máquina MDI, onde converte os dados em coordenadas ou *strips* para o sistema de exposição. As *strips* são frames em pixels que representam áreas de exposição.

Tabela 1. Parâmetros de exportação da imagem Teste no formato Gerber (O autor, 2020).

Parâmetro	Descrição
%FSLAX46Y46*%	Sem Zeros a esquerda, 6 casas decimais e máximo 4 dígitos em X e Y
%MOMM*%	Unidade de exportação: milímetro
%ADD106C,0.02000000*%	Símbolo Circular Dcode 106
%ADD14C,0.04000000*%	Símbolo Circular Dcode 14
%IPPOS*%	Polaridade da Imagem Positiva
%LPD*%	Carrega Polaridade

Na Fig. 4 é possível observar o fluxo dos dados minerados do atuador pelo medidor. O Sistema IoT desenvolvido neste trabalho é composto por gateway, *broker*, database, *analytics* e *front end*. Para visualizar as grandezas elétricas medidas, basta conectar com *broker*, no qual todas as informações são publicadas por meio do protocolo MQTT, por onde é baseado no princípio de publicação de mensagens e assinatura de tópicos, ou "pub / sub" (publish / subscribe). Vários clientes, dispositivos, se conectam a um broker e assinam tópicos nos quais estão interessados. Muitos clientes podem se

inscrever nos mesmos tópicos e manipular as informações como bem entenderem. O *broker* e o MQTT agem como uma interface simples e comum para a qual tudo pode se conectar (Tigli, 2016).

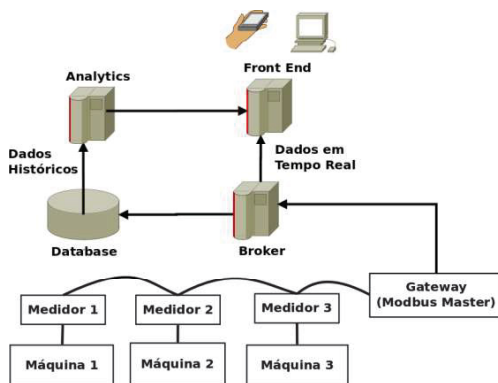


Figura 4. Visão Geral do Sistema desenvolvido para minerar os dados (O autor, 2020).

Essas informações precisam ser analisadas por uma aplicação, no gateway, que determine o comportamento com base em certos valores de grandezas elétricas. A lógica e algoritmo desenvolvido para determinar por exemplo a produtividade em tempo real do atuador, é feita no gateway. Uma vez que a análise é feita, basta fornecer a informação em um aplicativo (front end) para o usuário.

O *gateway* que foi desenvolvido é o componente, tanto hardware como software, no qual estebelece a conexão entre o multimetro e o broker. Hardware por ser um computador físico, mas poderia ser virtual dependendo da possibilidade de fornecimento dos dados do multimetro, mas isso é detalhe intrínscio de cada fabricante.

Neste trabalho foi utilizado WEKA, pois oferece vantagens distintas, como mencionado, para os algoritmos e é rápido, fácil de implementar e acessar; fácil de usar; e requer zero investimento financeiro.

Após os dados das medições serem obtidos é necessário que sejam preparados com a definição de atributos, filtros e classes para que seja possível importar no WEKA corretamente e assim definir o melhor algoritmo e criar o modelo. As classificações foram *exposing*, que indica as instâncias na etapa de exposição, *setup*, na qual se refere as instâncias durante o setup de exposição e por fim a classificação como *done*, onde representam as instâncias do fim da exposição da imagem (Tab. 2).

Tabela 2. Atributos definidos no formato ARFF das 3600 medições (O autor, 2020).

Atributos ARFF	Descrição
@attribute start DATE 'hh:mm:ss'	Tempo Inicial no formato Horas: Minutos: Segundos
@attribute end DATE 'hh:mm:ss'	Tempo Inicial no formato Horas: Minutos: Segundos
@attribute powerA numeric	Potência trifásica A [KW]
@attribute powerB numeric	Potência trifásica B [KW]
@attribute powerC numeric	Potência trifásica B [KW]
@attribute VrmsAN numeric	Tensão rms Fase A com N [V]
@attribute VrmsBN numeric	Tensão rms Fase B com N [V]
@attribute VrmsCN numeric	Tensão rms Fase C com N [V]
@attribute VrmsAB numeric	Tensão rms Fase A com B [V]
@attribute VrmsBC numeric	Tensão rms Fase B com C [V]
@attribute VrmsCA numeric	Tensão rms Fase C com A [V]
@attribute IrmsA numeric	Corrente rms A
@attribute IrmsB numeric	Corrente rms B
@attribute IrmsC numeric	Corrente rms C
@attribute THD V A numeric	Distorção harmônica total da Tensão Fase A [%]
@attribute THD V B numeric	Distorção harmônica total da Tensão Fase B [%]
@attribute THD V C numeric	Distorção harmônica total da Tensão Fase C [%]
@attribute THD I A numeric	Distorção harmônica total da Corrente Fase A [%]
@attribute THD I B numeric	Distorção harmônica total da Corrente Fase A [%]
@attribute THD I C numeric	Distorção harmônica total da Corrente Fase A [%]
@attribute Vrms_Fundamental AN numeric	Tensão rms Fundamental Fase A com N [V]
@attribute Vrms_Fundamental BN numeric	Tensão rms Fundamental Fase B com N [V]

Tabela 3. Continuação da Tabela 2 (O autor, 2020).

Atributos ARFF	Descrição
@attribute Vrms Fundamental CN numeric	Tensão rms Fundamental Fase C com N [V]
@attribute Irms Fundamental A numeric	Corrente rms Fundamental Fase A com N [V]
@attribute Irms Fundamental B numeric	Corrente rms Fundamental Fase A com N [V]
@attribute Irms Fundamental C numeric	Corrente rms Fundamental Fase A com N [V]
@attribute class {exposing,setup,done}	Tipos de classificação: exposing, setup, done

4. RESULTADOS E DISCUSSÃO

Para uma maior base de dados dos ensaios, aumentando assim a confiabilidade dos resultados, a imagem foi exposta durante uma hora por intervalo de medição a cada segundo, totalizando portanto 3600 medições, isto é, instâncias (Tab. 4). Além de que, como mencionado, aplicado uma energia de 400mJ na imagem.

Tabela 4: 40 instâncias no Formato ARFF (O autor , 2020).

ID	Instâncias
1	10:21:37,10:21:38,2522,2934,2517,228,226,230,394,395,396,12,14,12,1,1,1,49,35,53,228,226,230,11,13,11,exposing
2	10:21:38,10:21:39,2060,2493,2065,228,227,231,395,396,397,11,12,11,1,1,1,60,42,65,228,227,231,9,11,9,exposing
3	10:21:39,10:21:40,2192,2597,2204,228,227,231,394,396,397,11,12,11,1,1,1,56,39,60,228,227,231,10,12,10,exposing
4	10:21:40,10:21:41,2514,2934,2516,228,226,230,394,395,396,12,14,12,1,1,1,49,34,53,228,226,230,11,13,11,exposing
5	10:21:41,10:21:42,2248,2659,2257,228,227,230,394,396,397,11,13,12,1,1,1,55,38,59,228,227,230,10,12,10,exposing
6	10:21:42,10:21:43,1995,2426,2011,228,227,231,395,396,397,10,12,11,1,1,1,62,43,66,228,227,231,9,11,9,exposing
7	10:21:43,10:21:44,2520,2918,2520,228,226,230,394,395,397,12,14,12,1,1,1,49,35,53,228,226,230,11,13,11,exposing
8	10:21:44,10:21:45,2457,2857,2452,228,227,230,394,396,397,12,13,12,1,1,1,50,36,55,228,227,230,11,13,11,exposing
9	10:21:45,10:21:46,1819,2246,1834,229,227,231,395,396,398,10,11,10,1,1,1,69,47,73,229,227,231,8,10,8,exposing
10	10:21:46,10:21:47,2496,2911,2502,228,226,230,394,395,397,12,14,12,1,1,1,50,35,53,228,226,230,11,13,11,exposing
11	10:21:47,10:21:48,2520,2935,2522,228,226,230,394,395,396,12,14,12,1,1,1,49,35,53,228,226,230,11,13,11,exposing
12	10:21:48,10:21:49,1945,2381,1959,228,227,231,395,396,397,10,12,10,1,1,1,64,44,68,228,227,231,9,11,9,exposing
13	10:21:49,10:21:50,2303,2700,2311,228,227,230,394,396,397,12,13,12,1,1,1,53,38,58,228,227,230,10,12,10,exposing
14	10:21:50,10:21:51,2524,2924,2522,228,226,230,394,395,396,12,14,12,1,1,1,49,35,53,228,226,230,11,13,11,exposing
15	10:21:51,10:21:52,2149,2593,2154,228,227,231,394,396,397,11,12,11,1,1,1,57,40,61,228,227,230,10,12,9,exposing
16	10:21:52,10:21:53,1816,2348,1836,228,227,231,394,396,397,10,12,10,1,1,1,69,46,73,228,227,231,8,10,8,exposing
17	10:21:53,10:21:54,2504,2923,2514,228,226,230,393,395,396,12,14,12,1,1,1,49,34,53,228,226,230,11,13,11,exposing
18	10:21:54,10:21:55,2519,2939,2525,228,226,230,393,395,396,12,14,12,1,1,1,49,34,52,228,226,230,11,13,11,exposing
19	10:21:55,10:21:56,1927,2368,1946,228,227,231,394,396,397,10,11,10,1,1,1,64,44,68,228,227,231,9,11,9,exposing
20	10:21:56,10:21:57,2312,2712,2320,228,226,230,394,395,396,12,13,12,1,1,1,53,37,57,228,226,230,10,12,10,exposing
21	10:21:57,10:21:58,2525,2925,2521,228,226,230,393,395,396,12,14,12,1,1,1,49,35,53,228,226,230,11,13,11,exposing
22	10:21:58,10:21:59,2133,2560,2137,228,227,230,394,396,397,11,12,11,1,1,1,57,40,62,228,227,230,9,11,9,exposing
23	10:21:59,10:22:00,2112,2538,2127,228,227,230,394,396,397,11,12,11,1,1,1,58,40,62,228,227,230,9,11,9,exposing
24	10:22:00,10:22:01,2517,2952,2521,228,226,230,393,395,396,12,14,12,1,1,1,49,34,53,228,226,230,11,13,11,exposing
25	10:22:01,10:22:02,2326,2745,2332,228,226,230,394,395,397,12,13,12,1,1,1,53,37,57,228,226,230,10,12,10,exposing
26	10:22:02,10:22:03,1912,2350,1931,228,227,231,395,396,397,10,11,10,1,1,1,65,44,69,228,227,231,8,10,8,exposing
27	10:22:03,10:22:04,2522,2920,2519,228,226,230,394,395,396,12,14,12,1,1,1,49,35,53,228,226,230,11,13,11,exposing
28	10:22:04,10:22:05,2523,2921,2518,228,226,230,394,395,396,12,14,12,1,1,1,49,35,53,228,226,230,11,13,11,exposing
29	10:22:05,10:22:06,1821,2250,1835,229,227,231,395,396,397,10,11,10,1,1,1,68,47,73,229,227,231,8,10,8,exposing
30	10:22:06,10:22:07,2417,2835,2425,228,226,230,394,395,397,12,13,12,1,1,1,51,36,55,228,226,230,11,13,11,exposing
31	10:22:07,10:22:08,2519,2946,2508,228,226,229,394,395,396,12,14,12,1,1,1,49,34,53,228,226,229,11,13,11,done
32	10:22:08,10:22:09,2022,2434,2018,229,227,230,394,395,397,10,12,11,1,1,1,61,42,65,229,227,230,9,11,9,setup
33	10:22:09,10:22:10,2002,2221,1824,228,227,230,395,396,397,10,11,10,1,1,1,59,47,74,228,227,230,9,10,8,setup
34	10:22:10,10:22:11,1918,2215,1822,228,227,230,394,396,397,10,11,10,1,1,1,62,47,73,228,227,230,8,10,8,setup
35	10:22:11,10:22:12,1865,2483,1918,229,227,230,394,396,397,10,12,10,1,1,1,72,45,73,229,227,230,8,11,8,setup
36	10:22:12,10:22:13,1876,2303,1861,229,227,230,394,396,397,10,11,10,1,1,1,68,46,73,229,227,230,8,10,8,setup
37	10:22:13,10:22:14,2030,2183,1798,228,227,230,394,396,397,10,11,10,1,1,2,56,47,74,228,227,230,9,10,8,setup
38	10:22:14,10:22:15,1838,2191,1799,228,227,230,394,396,397,10,11,10,1,1,1,66,47,74,228,227,230,8,10,8,setup
39	10:22:15,10:22:16,1779,2184,1800,229,227,230,395,396,397,9,11,10,1,1,1,70,47,73,229,227,230,8,10,8,setup
40	10:22:16,10:22:17,2395,2802,2384,228,226,229,393,395,396,12,13,12,1,1,1,51,36,55,228,226,229,11,12,10,exposing

A Tab.3 representa as primeiras 40 instâncias, ou também pode-se dizer que apartir do instante 10h 21 minutos e 37 segundos até 10h 22 minutos e 17 segundos, do total de 3600 como já mencionado. Cada instância apresenta valores seraparados por vírgula, onde cada valor é atribuído respectivamente de acordo com a Tab. 2. Nessas instâncias é possível

analisar, identificadas na coluna ID (identificação), que da primeira até a trigésima, estas foram classificadas como *exposing*. A trigésima primeira está classificada como *done*. A partir da trigésima segunda até a trigésima nona, a classificação está como *setup*. A quadragésima volta a ser classificada como *exposing* e assim cada instância seguinte apresentará sua própria classificação. Além disso, se nota que nas classificações como *exposing* a potência trifásica é maior do que nas classificações como *setup*, pois nessas não há aplicação de energia de exposição com 400mJ.

Em máquinas ferramentas como a fresadora o consumo de energia é igual à potência multiplicada pelo tempo, a saber a área entre a curva de carga de potência da máquina ferramenta e o tempo do eixo (Lirong, *et al.*, 2015), como mostrado na Fig. 5. Analisando os valores de potência pelo consumo de energia ao produzir uma determinada peça é possível identificar o início e fim do processo e contabilizar em um aplicativo, isto é, pelo consumo de energia determina-se a produtividade do atuador, além de que é possível também identificar etapas do processo.

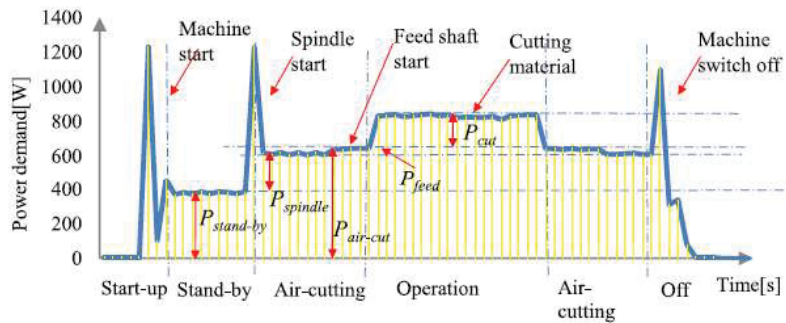


Figura 5: Diagrama esquemático do perfil de potência do processo de fresagem (Lirong, *et al.*, 2015).

Do mesmo modo, como é mostrado na Fig. 6, o teste realizado gerou grandezas elétricas já esperadas como mencionado neste trabalho, no entanto o consumo de energia ao expor a imagem teste (Fig. 3) permitiu identificar o início e fim do sistema de exposição, isto é, podendo contabilizar o número de exposições e disponibilizar essa informação em um aplicativo independente de rede industrial e software do atuador, ou seja, utilizando apenas técnicas de mineração de dados.

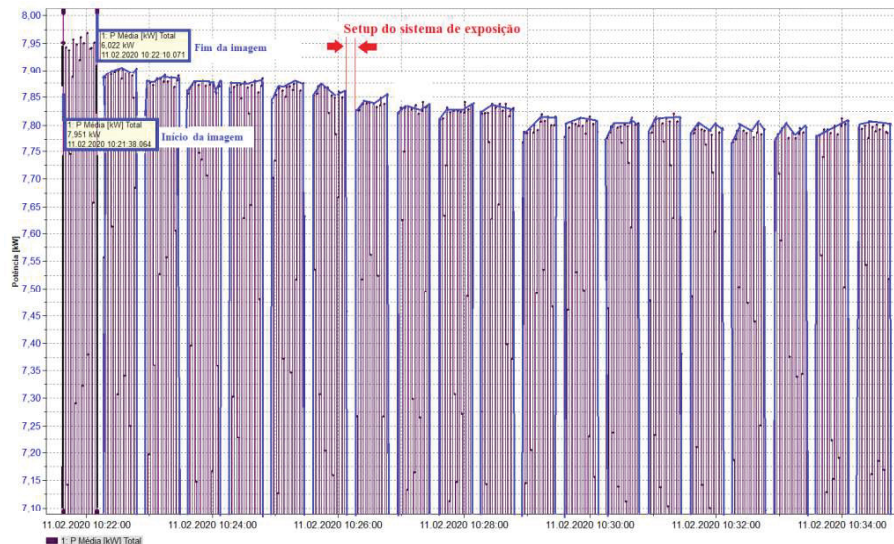


Figura 6: Perfil de potência do sistema de exposição MDI da imagem teste (O autor, 2020).

A primeira etapa do processo de mineração foi preparar os dados para a importação. Uma vez importado apresentou o resultado indicado na Fig. 7, onde ilustra um desequilíbrio de classes.

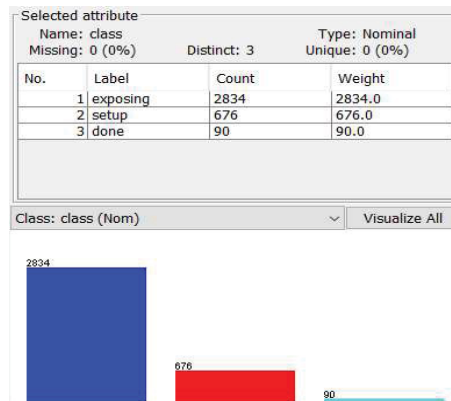


Figura 7: Perfil das classes desequilibradas no WEKA (O autor , 2020).

Quando um problema de classificação binária possui menos dados de uma maneira significativa em uma classe do que na outra, Fig. 7, ao aprender com os resultados de uma medição, as quais a grande maioria das instâncias tem um resultado *exposing* e apenas alguns retornam *done*, alguns algoritmos de aprendizagem de máquina simplesmente aprenderão a ignorar a classe minoritária e a classificar todos os casos na classe majoritária, porque isso produzirá trivialmente alta precisão de classificação (Fig. 8). Esse tipo de modelo de classificação claramente não é útil. Para lidar com classes desequilibradas em problemas de classificação, dois métodos comuns para combater esse problema são a subamostragem, do inglês *undersampling*, da classe majoritária e a superamostragem, do inglês *oversampling*, da classe minoritária, respectivamente (Frank, 2019). O primeiro método, *undersampling*, exclui a quantidade de instâncias das classes majoritária para o número de instâncias da classe minoritária e pode resultar na perda de informações inestimáveis para um modelo. O segundo método, *oversampling*, replica instâncias da classe minoritária, para que as classes tenham o mesmo número de instâncias.

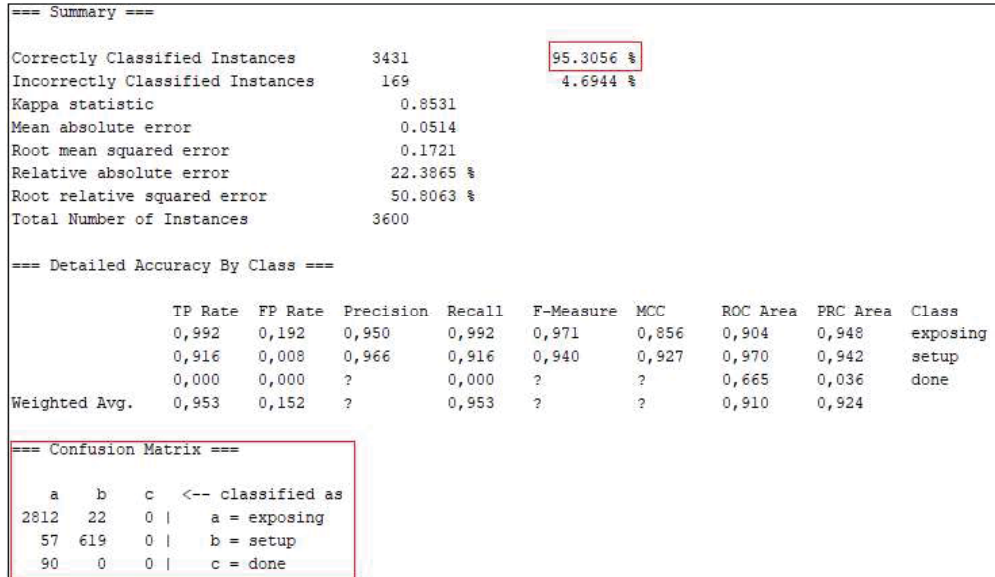


Figura 8: Modelo com algoritmo J48 e sem filtro para equilibrar as classes (O autor, 2020).

Nota-se que na Fig. 8 mostra uma precisão maior que 95% o que é excelente. No entanto, ao analisar a Matrix de confusão do inglês *Confusion Matrix*, nenhuma instância foi classificada como *done*. Esse modelo inicial não atendeu a necessidade do trabalho. Portanto, foi necessário aplicar o filtro *oversampling* nas classes (Fig. 9).

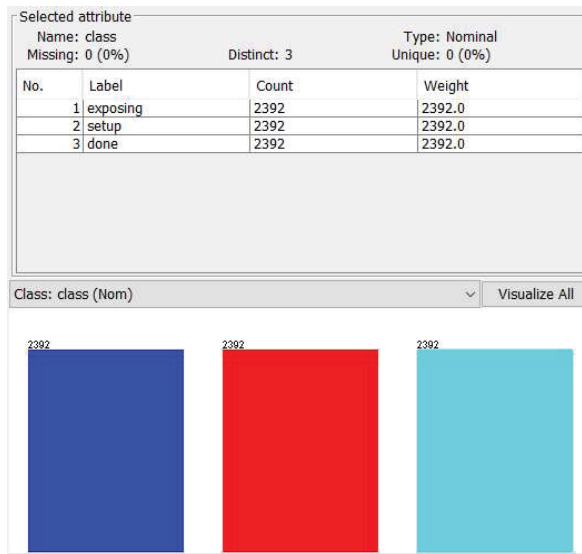


Figura 9: Perfil das classes equilibradas com filtro *oversampling* no WEKA (O autor , 2020).

Com o número de instâncias equilibradas entre as classes, Fig. 9, foi possível testar algoritmos de aprendizado de máquina, no entanto o que teve uma maior precisão foi o J48 com aproximadamente 99%, além do resultado da *confusion Matrix* com todas as instâncias *done* aplicadas corretamente (Fig. 10).

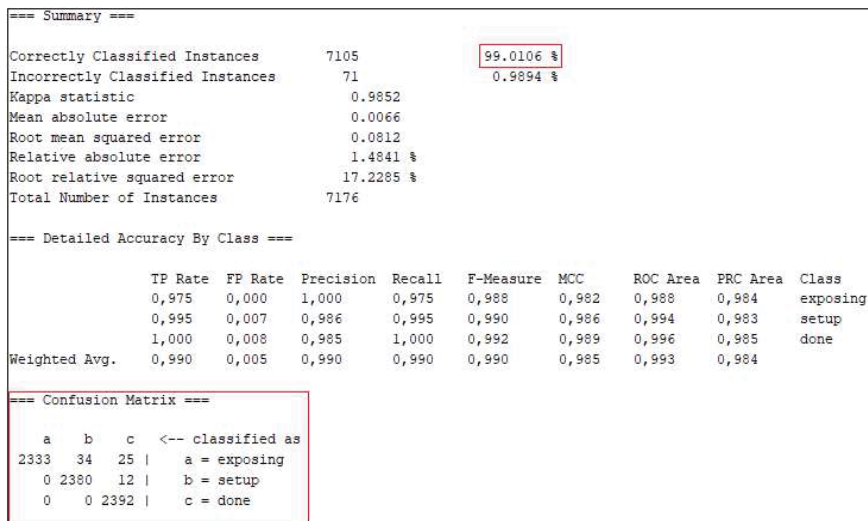


Figura 10: Modelo com algoritmo J48 e com filtro para equilibrar as classes (O autor, 2020).

Para validar o modelo teste resultante, Fig. 10, foi criado 9 instâncias, Tab. 5, semelhantes ao da Tab. 4, para analisar o que foi previsto na classificação. Nota-se, Tab. 4, que cada instância foi iniciada como desconhecida (?), embora é conhecido antecipadamente qual é seu valor indicado na coluna Classe.

Tabela 5: 9 instâncias no Formato ARFF para teste (O autor , 2020).

ID	Instâncias	Classe
1	12:21:39,10:21:40,2192,2597,2204,228,227,231,394,396,397,11,12,11,1,1,1,56,39,60,228,227,231,10,12,10,?	exposing
2	12:21:40,10:21:41,2514,2934,2516,228,226,230,394,395,396,12,14,12,1,1,1,49,34,53,228,226,230,11,13,11,?	exposing
3	12:22:00,10:22:01,2517,2952,2521,228,226,230,393,395,396,12,14,12,1,1,1,49,34,53,228,226,230,11,13,11,?	exposing
4	12:22:07,10:22:08,2519,2946,2508,228,226,229,394,395,396,12,14,12,1,1,1,49,34,53,228,226,229,11,13,11,?	done
5	12:22:47,10:22:48,2496,2929,2500,228,226,230,394,395,396,12,14,12,1,1,1,48,34,52,228,226,230,11,13,11,?	done

Tabela 6: Continuação Tabela 5 (O autor , 2020).

ID	Instâncias	Classe
6	12:23:27,10:23:28,2493,2921,2497,228,226,230,394,395,396,12,14,12,1,1,1,49,34,52,228,226,230,11,13,11,?	done
7	13:22:08,10:22:09,2022,2434,2018,229,227,230,394,395,397,10,12,11,1,1,1,61,42,65,229,227,230,9,11,9,?	setup
8	13:22:09,10:22:10,2002,2221,1824,228,227,230,395,396,397,10,11,10,1,1,1,59,47,74,228,227,230,9,10,8,?	setup
9	13:22:10,10:22:11,1918,2215,1822,228,227,230,394,396,397,10,11,10,1,1,1,62,47,73,228,227,230,8,10,8,?	setup

A Fig. 11 ilustra que todas as instâncias foram previstas corretamente:

```

=== Re-evaluation on test set ===

User supplied test set
Relation:      MdiMachine
Instances:    unknown (yet). Reading incrementally
Attributes:   27

=== Predictions on user test set ===

  inst#  actual  predicted  error prediction
    1    1:?   1:exposing    1
    2    1:?   1:exposing    1
    3    1:?   1:exposing    1
    4    1:?   3:done        1
    5    1:?   3:done        1
    6    1:?   3:done        1
    7    1:?   2:setup       1
    8    1:?   2:setup       1
    9    1:?   2:setup       1

=== Summary ===
    
```

Figura 11: Teste Modelo J48 (O autor, 2020).

6. CONCLUSÃO

Foi possível analisar que utilizando um multimedidor de grandezas elétricas é possível obter informações do atuador, ao analisar os dados obtidos, sem a preocupação com topologias e comunicação de rede, softwares dos atuadores, fabricantes e outras variáveis que complicam a implementação de IoT na indústria. Assim como foi descrito neste artigo, a produtividade, isto é, quantas peças foram produzidas, tempo de setup, ociosidade, falhas durante o processamento e características afins são variáveis possíveis de serem analisadas pelo software apenas interpretando dados das grandezas elétricas medidas.

Embora a eficiência energética da conotação de máquinas ferramentas não foi discutida neste artigo, as pressões ambientais, legais e econômicas exigem empresas manufatureiras para reduzir o consumo de energia e emissão de carbono (Lirong et al., 2015). Especificamente a máquina ferramenta como dispositivo de consumo de energia base no sistema de manufatura, a modelagem de consumo de energia e avaliação da eficiência energética são pré-requisitos para economizar energia na fabricação e também um passo em direção à sustentabilidade. Isso significa que o projeto deste artigo pode ser uma ferramenta auxiliar para identificar se a produção de uma determinada peça foi energeticamente eficiente. Além de que as máquinas ferramentas já poderão vim de fábrica com medidores de energia, o que facilitaria ainda mais a implementação do projeto. Outra possibilidade de implementação seria a previsão de carga em sistemas de energia elétrica baseado em redes neurais artificiais.

O método *oversampling* utilizado neste trabalho, embora teve um resultado positivo, ele duplica exemplos da classe minoritária no conjunto de dados de treinamento e pode resultar em ajustes excessivos em alguns modelos. No entanto, também existe o classificador *CostSensitiveClassifier*, que, quando aplicado no modo padrão, ponderará novamente as instâncias de treinamento para levar em consideração uma determinada matriz de custos de classificação incorreta e, em seguida, usa o classificador criado a partir dos dados ponderados. Os pesos produzidos podem ser usados diretamente por qualquer classificador no Weka que implementa a interface *WeightedInstancesHandler*. Se o classificador base especificado em *CostSensitiveClassifier* não implementar essa interface *WeightedInstancesHandler*, os dados serão amostrados novamente (com substituição) com base nos pesos, normalizando os pesos em uma distribuição de probabilidade usada para amostragem. Como as classes, inicialmente, estavam extremamente desequilibradas, ao utilizar a matriz de custos não foi possível obter um resultado satisfatório. Finalmente, existe o filtro supervisionado *ClassBalancer*, um filtro muito simples que atribui pesos de instância para que cada classe de instâncias tenha o mesmo peso e a soma total de pesos de instância no conjunto de dados permaneça inalterada. Quando o *FilteredClassifier* é aplicado em conjunto com esse filtro e o classificador base não implementa *WeightedInstancesHandler*, os pesos serão

novamente usados para formar uma distribuição de probabilidade para amostragem com substituição. Isso produzirá um conjunto de treinamento em que as duas classes são (aproximadamente) equilibradas.

É claro que muitas empresas não estão prontas para tirar proveito do resultado da economia pela IIoT. Quando a Accenture pesquisou a decisão de mais de 1.400 executivos fabricantes - incluindo 736 CEOs - de alguns das maiores empresas do mundo, a grande maioria (84%) acredita que suas organizações têm a capacidade de criar novos fluxos de renda baseados em serviços de o IIoT. Setenta e três por cento confessam que suas empresas ainda precisam fazer progressos concretos. Apenas 7% desenvolveram uma estratégia abrangente com investimentos a condizer. Da mesma forma, CEOs e executivos expressam confiança notável (96%) que a liderança sênior em suas organizações entende pelo menos algo da natureza da IIoT. Mas muito menos dizem que seus líderes entendem algo a respeito (38%). E mesmo essa porcentagem parece refletir um grau de excesso de confiança, Accenture, em colaboração com o Industrial Internet Consortium (IIC) empreendeu um Fórum Econômico Mundial Industrial Pesquisa de Impacto da Internet entre mais de 90 líderes de mercado, 6 dos que estão ativamente perseguindo iniciativas da IIoT. A grande maioria (88%) disseram que ainda não possuem entendimento dos modelos de negócios subjacentes e implicações a longo prazo da IIoT. Essas opiniões conflitantes sublinham incerteza em torno da IIoT. Nota-se, portanto, que muitas empresas não compreendem todo o seu potencial e que a maioria não está preparada para a fundamental mudança necessária para fornecer produtos híbridos e serviços em uma economia focada em resultados. Se as empresas deixarem de reconhecer o potencial da IIoT para agir, eles precisam melhorar sua compreensão através de experimentação precoce e pilotos com vários parceiros.

Contudo, os sistemas de tecnologia operacional de hoje funcionam amplamente em silos. No entanto, no futuro, um ecossistema digital totalmente funcional exigirá compartilhamento de dados entre máquinas e outros sistemas físicos de diferentes fabricantes. O esforço em direção à interoperabilidade contínua será ainda mais complicado pela longa vida útil de equipamentos industriais típicos, o que requerem uma adaptação ou substituição dispendiosa para trabalhar com as mais recentes tecnologias. Sistemas invasivos, que dependem de adaptação conforme o sistema operacional do atuador ou por meio de protocolo de internet (IP) não parecem ser uma boa proposta. Além disso, outras barreiras e riscos notáveis incluem retorno incerto sobre investimentos em novas tecnologias, tecnologias imaturas ou não testadas, falta de regras de governança de dados além das fronteiras geográficas e uma escassez de talento. Superar esses desafios exigirá liderança, investimento e ações colaborativas entre os principais interessados.

7. REFERÊNCIAS

- Accenture, 2015 “Winning with the Industrial Internet of Things, How to accelerate the journey to productivity and growth” <<https://www.accenture.com>> Acesso em 13/07/2020
- Aggarwal, Charu C., 2015 “Data Mining” <<https://link.springer.com/book/10.1007%2F978-3-319-14142-8>> Acesso em 13/07/2020.
- Bremer, Max, 2016 “Principles of Data Mining” <<https://link.springer.com/book/10.1007%2F978-1-4471-7307-6>> Acesso em 13/07/2020.
- D. Schmilovitz, 2005 “On the definition of total harmonic distortion and its effect on measurement interpretation” <<https://ieeexplore.ieee.org/document/1375138>> Acesso em 13/07/2020.
- Eibe Frank, Mark A. Hall, and Ian H. Witten, 2016 “Data Mining: Practical Machine Learning Tools and Techniques” <https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf> Acesso em 13/07/2020.
- Fluke. 12 Abril 2020 <<https://www.fluke.com/1736-1738>> Acesso em 13/07/2020.
- Frank, Eibe, 2019 <<https://waikato.github.io/weka-blog/posts/2019-01-30-sampling/>> Acesso em 13/07/2020.
- Han, Jiawei; Kamber, Micheline e Pei, Jia, 2015 “Data Mining Concepts and Techniques Third Edition” <<http://myweb.sabanciuniv.edu/data-mining-concepts-and-techniques.pdf>> Acesso em 13/07/2020.
- Iba, Wayne e Langley Pat (1992) “Induction of One-Level decision trees” <<http://lyonesse.stanford.edu//stump.ml92.pdf>> Acesso em 13/07/2020.
- Lirong Zhou, Jianfeng Li, Fangyi Li, Qiang Meng, Jing Li, Xingshuo Xu, 2015. Energy consumption model and energy efficiency of machine tools: a comprehensive literature review Modbus, 22 Abril 2020 <http://modbus.org/docs/PI_MBUS_300.pdf> Acesso em 13/07/2020.
- RTI, 2015 <<http://www.onsite.rti.com/industries>> Acesso em 13/07/2020.
- Silver, D. and Hassabis, D., 2016. AlphaGo: Mastering the Ancient Game of Go with Machine Learning. <<https://hci.iwr.uni-heidelberg.de>> Acesso em 13/07/2020.
- Simon, Toby 2017 Paper Series: No. 46 – Jan. 2017 Global Commission on Internet Governance. “Critical Infrastructure and the Internet of Things” <<https://www.cigionline.org/publications/iot>> Acesso em 13/07/2020.
- Tanenbaum, Andrew S. 2011. Rede de Computadores 5º Edição
- Thomas, Joel, 2011 “Industrial Network Security: Securing Critical Infrastructure Networks for Smart Grid, SCADA, and Other Industrial Control Systems”

G. Paulo, T. Francisco

Internet Industrial das Coisas Utilizando Técnicas de Mineração de Dados: Um Estudo de Caso Sobre Iot e Energia Elétrica

The Pcb magazine February 2017

<<http://www.magazines007.com/pdf/PCB-Feb2017.pdf>> Acesso em 13/07/2020.

Tigli, J.-Y. I3S – University of Nice Sophia Antipolis, 2015-2016 Middleware for Internet of Things – MQTT

WEFUSA, 2015 Industrial Internet of Things: Unleashing the Potential of Connected Products and Services

<<http://www3.weforum.org/docs/WEFUSA>> Acesso em 13/07/2020.

U. A. Bakshi e V.U. Bakshi, 2008 “Electrical Circuits And Machines”, Fourth Revised Edition, pág. 273.

Ucamco, 2019 “The Gerber File Format Specification”

< <https://www.ucamco.com/files/downloads/file/81> > Acesso em 13/07/2020.

Zühlke, D. 2014. Industrie 4.0: From Vision to Reality. Tech News. Vol 26. pp 10-11.