



INSTITUTO TECNOLÓGICO VALE



**Programa de Pós-Graduação em Instrumentação, Controle e
Automação de Processos de Mineração (PROFICAM)
Escola de Minas, Universidade Federal de Ouro Preto (UFOP)
Associação Instituto Tecnológico Vale (ITV)**

Dissertação

**PROPOSTA DE PLATAFORMA INTEGRADA PARA DIGITAL TWIN COM USO DE
APRENDIZADO DE MÁQUINA**

Luiz Guilherme Menezes Barata

**Ouro Preto
Minas Gerais, Brasil
2019**

Luiz Guilherme Menezes Barata

**PROPOSTA DE PLATAFORMA INTEGRADA PARA DIGITAL TWIN COM USO DE
APRENDIZADO DE MÁQUINA**

Dissertação apresentada ao Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração da Universidade Federal de Ouro Preto e do Instituto Tecnológico Vale, como parte dos requisitos para obtenção do título de Mestre em Engenharia de Controle e Automação.

Orientador: Prof. Saul Emanuel Delabrida
Silva, D.Sc.

Ouro Preto
2019

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

B226p Barata, Luiz Guilherme Menezes .
Proposta de Plataforma Integrada para Digital Twin com Uso de Aprendizado de
Máquina. [manuscrito] / Luiz Guilherme Menezes Barata. - 2019.
78 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. Saul Emanuel Delabrida Silva.

Dissertação (Mestrado Profissional). Universidade Federal de Ouro Preto.
Programa de Mestrado Profissional em Instrumentação, Controle e Automação de
Processos de Mineração. Programa de Instrumentação, Controle e Automação de
Processos de Mineração.

Área de Concentração: Engenharia de Controle e Automação de Processos
Minerais.

1. Representação descritiva - Digital Twin. 2. Inteligência artificial - Mineração
de dados - Detecção de Anomalias. 3. Aprendizado do computador (Machine
Learning). 4. Internet das coisas - Indústria 4.0. I. Barata, Luiz Guilherme Menezes
. II. Silva, Saul Emanuel Delabrida. III. Universidade Federal de Ouro Preto. IV.
Título.

CDU 681.5:622.2

Bibliotecário(a) Responsável: Maristela Sanches Lima Mesquita - CRB:1716

Mestrado Profissional em Instrumentação, Controle e Automação de Processos de
Mineração - PROFICAM

Proposta de Plataforma Integrada para Digital Twin com Uso de Aprendizado de
Máquina

Luiz Guilherme Menezes Barata

Dissertação defendida e aprovada em 02 de setembro de 2019 pela banca examinadora
constituída pelos professores:

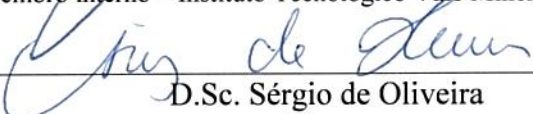


D.Sc. Saul Emanuel Delabrida Silva
Orientador – Universidade Federal de Ouro Preto (UFOP)

D.Sc. Andrea Gomes Campos Bianchi
Participação por videoconferência
Membro interno – Universidade Federal de Ouro Preto (UFOP)



D.Sc. Gustavo Pessin
Membro interno – Instituto Tecnológico Vale Mineração (IFMG)



D.Sc. Sérgio de Oliveira
Membro externo – Universidade Federal de São João del-Rei (UFSJ)

Para Lara e Sofia.

Agradecimentos

Às minhas pequenas Lara e Sofia, que com gestos de compreensão como “O papai está estudando para o mestrado agora” me fortaleceram e apoiaram de forma doce e objetiva no desenvolvimento de todo este projeto.

À Érica pelo incentivo e paciência em todos os momentos e, principalmente, pela orquestração de tudo ao seu alcance para garantir o sucesso desta empreitada.

Ao meu orientador Professor Saul Emanuel Delabrida Silva, pelo tempo dedicado, assertividade, competência e, principalmente, pela sensibilidade para entender as limitações que impactaram o trabalho.

À Vale por propiciar um ambiente de incentivo à pesquisa e apoio para evolução profissional.

Aos meus amigos do time de Inovação em TI Industrial da Vale, por compartilharem seus conhecimentos e experiências.

À minha família, pela firme motivação durante todo o percurso e por ceder cada minuto do tempo precioso de seu convívio.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES), Código de Financiamento 001; do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG); e da Vale SA.

"Do. Or do not. There is no try."
(Yoda)

Resumo

Resumo da Dissertação apresentada ao Programa de Pós Graduação em Instrumentação, Controle e Automação de Processos de Mineração como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

PROPOSTA DE PLATAFORMA INTEGRADA PARA DIGITAL TWIN COM USO DE APRENDIZADO DE MÁQUINA

Luiz Guilherme Menezes Barata

Setembro/2019

Orientador: Saul Emanuel Delabrida Silva

Equipamentos industriais estão sujeitos a diferentes condições operacionais durante seu ciclo de vida que tornam complexo o uso da modelagem e simulação clássicas. A convergência digital abriu caminho para desenvolvimento de tecnologias de simulação capazes de se adaptarem e preverem situações e prescrever ações, evoluindo para o conceito de *Digital Twin*. Este trabalho descreve experiências no desenvolvimento de aplicações práticas da Indústria 4.0 em transportadores de correia: um modelo SVM para predição de falhas; instalação de sensores de baixo custo para inspeção e monitoramento remotos de vibração; e a integração de dados de processos e equipamentos com serviços de computação na nuvem com *streaming*. Os resultados obtidos com cada uma das aplicações são analisados, de forma conjunta, frente aos requisitos do *Digital Twin* para propor uma arquitetura integrada para implementação dessa tecnologia utilizando uma plataforma na nuvem, independente de sistemas ou soluções existentes nas empresas.

Palavras-chave: Digital Twin, Detecção de Anomalias, Indústria 4.0.

Macrotema: Usina; **Linha de Pesquisa:** Tecnologias da Informação, Comunicação e Automação Industrial; **Tema:** Aumento de Produtividade na Usina/Pelotização; **Área Relacionada da Vale:** Tecnologia Operacional.

Abstract

Abstract of Dissertation presented to the Graduate Program on Instrumentation, Control and Automation of Mining Process as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

DIGITAL TWIN PLATFORM USING MACHINE LEARNING

Luiz Guilherme Menezes Barata

September/2019

Advisor: Saul Emanuel Delabrida Silva

Industrial equipment are subject to different operational conditions during their lifecycle which makes the use of classic modeling and simulation techniques more complex. The digital convergence paved the way for the development of simulation technology capable of adapting to operational changes therefore predicting occurrences and prescribing actions accurately, evolving to the Digital Twin concept. This work presents practical Industry 4.0 experiences for conveyor belts: an Support Vector Machine model to predict failures; low cost sensors for remote inspection and monitoring of vibration; and integration of process and equipment data to cloud storage via streaming. The results of each experience are analyzed against the Digital Twin requirements to define an integrated architecture to implement the technology in a cloud platform, independent of software or solution provider.

Keywords: Digital Twin, Anomaly Detection, Industry 4.0.

Macrotheme: Processing Plant; **Research Line:** Information, Communication and Industrial Automation Technologies; **Theme:** Productivity Increase of Processing/Pelletizing Plant; **Related Area of Vale:** Operational Technology.

Lista de Figuras

Figura 1.1	Circuito singelo de alimentação de uma usina de beneficiamento de minério de ferro. Fonte: Autor.	18
Figura 1.2	Gráfico da utilização física agregada das correias transportadoras. Fonte: Autor.	19
Figura 2.1	Sistemas e/ou software utilizados conforme a fase do ciclo de vida. Fonte: (RÍOS <i>et al.</i> , 2015)	22
Figura 2.2	Evolução das tecnologias de simulação. Fonte: (ROSEN <i>et al.</i> , 2015)	23
Figura 2.3	Evolução da convergência entre os mundos físico e digital. Fonte: (TAO e ZHANG, 2017)	23
Figura 2.4	Potencial impacto econômico da digitalização na indústria da mineração. Fonte: adaptado de World Economic Forum (2017)	24
Figura 2.5	Curva <i>hype cycle</i> para tecnologias emergentes. Fonte: (GARTNER, 2018a)	25
Figura 2.6	Arquitetura de um coletor de dados de um PIMS utilizando OPC. Fonte: (ASPEN TECHNOLOGY, 2016)	26
Figura 2.7	O PIMS inserido no contexto da norma ISA95. Fonte: Autor.	27
Figura 2.8	Exemplo de aplicações típicas do PIMS. Fonte: Autor.	28
Figura 2.9	Variação das funcionalidades MES conforme os tipos de manufatura. Fonte: (KLETTI, 2007)	30
Figura 2.10	Relatório apresentando dados consolidados de eventos de equipamento em processo contínuo de beneficiamento de minério de ferro. Fonte: Autor.	31
Figura 2.11	Relação entre os cinco Vs conforme a definição de <i>Big Data</i> do IDC. Fonte: (ADDO-TENKORANG e HELO, 2016)	32
Figura 2.12	Visão geral de um ESB. Fonte: (KEEN <i>et al.</i> , 2004)	37
Figura 2.13	Support Vector Machine. Fonte: (RASCHKA e MIRJALILI, 2017).	39
Figura 2.14	Roletes de carga suportando uma correia. Fonte: (NASCIMENTO, 2018)	41
Figura 2.15	Proposta de serviços para uma plataforma de DT. Fonte: (TAO <i>et al.</i> , 2018)	43
Figura 2.16	Tendência de pesquisa pelo termo “Digital Twin” na Internet nos últimos 10 anos. Fonte: Google Trends	44
Figura 3.1	Fases do modelo de referência CRISP-DM. Fonte: (CHAPMAN <i>et al.</i> , 2000)	47
Figura 3.2	Distribuição dos estados da correia transportadora. Fonte: Autor.	48

Figura 3.3	Fluxo de dados da solução para leitura de vibração de equipamentos. Fonte: Documentação Interna da Vale	51
Figura 3.4	<i>Framework</i> de IIoT <i>Analytics</i> . Fonte: adaptado de Moura <i>et al.</i> (2017) . . .	52
Figura 3.5	Integrações do PIMS que provêm dados de processos para sistemas corporativos da Vale. Fonte: Autor	53
Figura 3.6	Integração do PIMS para <i>backfilling</i> do <i>data lake</i> e <i>advanced analytics</i> . Fonte: adaptado de documentação da Vale	54
Figura 3.7	Integração do PIMS para <i>streaming</i> de dados para o <i>data lake</i> . Fonte: adaptado de documentação da Vale	54
Figura 4.1	Matrizes de confusão dos modelos treinados com os dados originais de 93 variáveis (acima) e PCA (abaixo). Fonte: Autor	56
Figura 4.2	Arquitetura em alto nível para implementação da solução de DT. Fonte: Autor	58
Figura 4.3	Hierarquia dos arquivos e esquema de dados proposto para o <i>data lake</i> . Fonte: Autor	60
Figura 4.4	Esquema de mensagem de integração proposto baseado no padrão <i>AutomationML</i> para criação de uma DTI. Fonte: Autor	63
Figura 4.5	Esquema típico de um modelo de classificação no Azure Machine Learning Studio. Fonte: (PAOLANTI <i>et al.</i> , 2018)	64

Lista de Tabelas

Tabela 2.1	Tecnologias e ferramentas normalmente utilizadas pelas aplicações BDA . .	33
Tabela 2.2	Comparação entre os bancos de dados corporativos e <i>data lake</i> . Fonte: adaptado de Fang (2015)	34
Tabela 2.3	Recursos mínimos de um ESB. Fonte adaptado de Keen <i>et al.</i> (2004)	38
Tabela 4.1	Acurácia atingida pelo modelo SVM com os dados originais. Fonte: Autor .	55
Tabela 4.2	Acurácia atingida pelo modelo SVM com <i>datasets</i> de transição do estado. Fonte: Autor	57
Tabela 4.3	Métricas para o <i>streaming</i> de dados. Fonte: documentação de projeto Vale .	57

Lista de Siglas e Abreviaturas

ADLS Azure Data Lake Store

API Application Programming Interface

APM Asset Performance Management

BD Big Data

BDA Big Data Analytics

BLE Bluetooth Low Energy

BOL Begin of Life

BNP Back-propagation Neural Networks

CBM Monitoramento Baseado na Condição

CMMS Computerized Maintenance Management System

CPPS Cyber Physical Production System

CPS Cyber Physical System

CRISP-DM Cross Industry Standard Process for Data Mining

DT Digital Twin

DTE Digital Twin Environment

DTI Digital Twin Instance

DTP Digital Twin Prototype

EOL End of Life

ERP Enterprise Resource Planning

ESB Enterprise Service Bus

ETL Extract Transform Load

HDFS Hadoop File System

IDC International Data Corporation

IIoT Internet das Coisas Industriais

IoT Internet das Coisas

ISA The International Society of Automation

JMS Java Message Service

JSON JavaScript Object Notation

KPI Key Process Indicator

LDA Linear Discriminant Analysis

LIMS Laboratory Information Management System

MES Manufacturing Execution System

MESA Manufacturing Enterprise Solutions Association

ML Machine Learning

MOL Middle of Life

MTS Multivariate Time Series

NASA National Aeronautics and Space Administration

ODBC Open Database Connectivity

OPC OLE for Process Control

TO Tecnologia Operacional

PCA Principal Component Analysis

PIMS Process Information Management System

CLP controlador lógico programável

PLM Product Lifecycle Management

REST Representational State Transfer

SDCD sistema digital de controle distribuído

SOA Service-Oriented Architecture

SVM Support Vector Machine

TCLD transportador de correia de longa distância

TI Tecnologia da Informação

VANT Veículo Aéreo Não Tripulado

WPD Wavelet Packet Decomposition

XML eXtensible Markup Language

Sumário

1	Introdução	16
1.1	Problema	17
1.2	Objetivos Gerais	17
1.2.1	Objetivos Específicos	17
1.3	Justificativa	18
1.4	Organização do Trabalho	19
2	Referencial Teórico e Trabalhos Relacionados	20
2.1	<i>Digital Twin</i>	20
2.2	Fontes de Dados	25
2.2.1	<i>Process Information Management System</i> (PIMS)	25
2.2.2	<i>Manufacturing Execution System</i> (MES)	28
2.2.3	Internet das Coisas Industriais	30
2.3	<i>Big Data</i>	30
2.3.1	<i>Data Lake</i>	32
2.4	Integração de Sistemas	35
2.4.1	Arquitetura Orientada a Serviços	35
2.4.2	Barramento de Integração	36
2.4.3	<i>AutomationML</i>	37
2.5	Modelo para Detecção de Anomalias	38
2.5.1	<i>Support Vector Machine</i> (SVM)	38
2.6	Transportador de Correia	40
2.7	Trabalhos Relacionados	42
2.7.1	Soluções para Detecção de Anomalias	42
2.7.2	Soluções para <i>Digital Twin</i>	42
3	Desenvolvimento	45
3.1	Metodologia	45
3.2	Modelo para Detecção de Anomalias	46
3.2.1	Entendimento do Processo	46
3.2.2	Compreensão dos Dados	47

3.2.3	Preparação dos Dados	48
3.3	Experiências com Aplicações Industriais	50
3.3.1	Internet das Coisas Industriais	50
3.3.2	Integração dos Dados de Processos e Equipamentos	52
4	Resultados e Discussão	55
4.1	Avaliação do Modelo SVM	55
4.2	Integração do PIMS com <i>Data Lake</i>	57
4.3	Arquitetura Integrada de <i>Digital Twin Environment</i>	58
4.3.1	Base de Conhecimento	59
4.3.2	Camada de Integração	60
4.3.3	Serviço de Detecção de Anomalias	63
5	Conclusão	65
6	Trabalhos Futuros	66
6.1	Serviço de Visualização	66
6.2	Serviço de Simulação	66
6.3	Serviço de Otimização	67
6.4	Outros Trabalhos	67
	Referências Bibliográficas	68
	Apêndices	75

1. Introdução

O conceito de um gêmeo para equipamentos não é novo e foi amplamente utilizado pela NASA nos programas espaciais. Durante o programa Apollo, foram desenvolvidos dois veículos sendo que o veículo em terra era utilizado no treinamento preparatório e, durante as missões espaciais, era utilizado para simular precisamente as condições da missão e permitir análise de solução para situações críticas no espaço.

Empresas de tecnologia como Amazon, Apple e Google utilizam o conceito de *Digital Twin* ao implementar uma cópia virtual do perfil de cada consumidor de seus serviços. Partindo de modelos demográficos genéricos, essas empresas coletaram dados de seus consumidores e aplicaram técnicas de análise avançada para criar o perfil virtual de cada um, que é aprimorado constantemente a cada clique, comentário, compra ou página visitada (GE DIGITAL, 2018). O emprego dessa tecnologia suporta a estratégia dessas empresas para aumentar suas receitas através de melhores experiências dos consumidores ao usar os seus produtos.

Propõe-se que o mesmo pode ser realizado no setor industrial dada a crescente digitalização na indústria, o aumento do volume de dados devido à Internet das Coisas Industriais e a convergência entre a Tecnologia Operacional (TO) e a Tecnologia da Informação (TI), isto é, uma quarta revolução que a indústria está passando, chamada de Indústria 4.0.

Um processo ou equipamento industrial está sujeito a condições operacionais diversas, ao desgaste dos equipamentos e falhas operacionais que impactam o desempenho geral de uma planta com o passar dos anos. Uma abordagem para aumentar o rendimento dos ativos operacionais em busca de resultados é o desenvolvimento de um modelo preditivo capaz de identificar e adaptar-se às condições operacionais, isto é, um *Digital Twin*.

Modelos de simulação tipicamente utilizados na indústria de mineração são desenvolvidos na fase de engenharia detalhada e podem não representar com exatidão uma operação unitária ou equipamento industrial em funcionamento há muitos anos por não considerar fatores que impactam nos parâmetros operacionais como, por exemplo, desgaste de peças, manutenções, condições climáticas, variabilidade dos produtos ou insumos e melhorias executadas.

Diante dessa realidade, o *Digital Twin* apresenta, em um primeiro momento, uma oportunidade de evolução do método clássico para simulação da operação de equipamentos tendo em vista a operação autônoma de uma planta industrial. Portanto, é importante analisar o processo de criação de algoritmos para detecção de anomalias de equipamentos na mineração juntamente com outras experiências de aplicações da Indústria 4.0 para identificar requisitos para uma solução de *Digital Twin* focada na produtividade de uma planta de tratamento de minério de ferro.

Os resultados parciais deste trabalho foram aceitos no 14º Simpósio Brasileiro de Automação Inteligente em um artigo intitulado “Experiência com Aprendizado de Máquina para Construção de um Digital Twin para a Indústria da Mineração” que foi apresentado em Outubro de 2019.

1.1. Problema

É grande a gama de tecnologias disponíveis para desenvolvimento de soluções computacionais com aplicações tanto para o ambiente de TI quanto o de TO, o que dificulta a definição de uma arquitetura que atenda aos requisitos de uma plataforma de *Digital Twin* escalável para atender a necessidade de uma empresa de mineração de grande porte.

Os sistemas de chão de fábrica possuem uma quantidade massiva de dados de sensores dos equipamentos industriais armazenados em bancos de dados históricos. Apesar dos dados estarem disponíveis, ainda são pouco integrados aos sistemas corporativos e tecnologias comumente utilizados em níveis táticos e estratégicos das empresas, normalmente em soluções arquitetadas pelo departamento de Tecnologia da Informação como por exemplo: computação na nuvem e Internet das Coisas. Portanto, é necessário definir uma arquitetura para integração dos sistemas de chão de fábrica em uma plataforma escalável de *Digital Twin* cujo objetivo seja aumentar a utilização de um equipamento industrial.

1.2. Objetivos Gerais

Apresentar uma arquitetura para plataforma integrada de *Digital Twin*, em um ambiente genérico independente da solução de fornecedores dos diversos sistemas de uma empresa, a partir das tecnologias de IIoT, *Big Data Analytics* e da experiência no desenvolvimento de um algoritmo para detecção de anomalias sob um conjunto de dados de sensores e informações oriundos de fontes no ambiente da TO.

1.2.1. Objetivos Específicos

- Conceituação original do *Digital Twin* a fim de se identificar os requisitos necessários de um *Digital Twin* capaz de endereçar o desafio de melhorar a utilização dos equipamentos das empresas no setor de mineração;
- Analisar os conceitos e técnicas de Internet das Coisas Industriais, *Big Data Analytics*, integração de sistemas e algoritmos para detecção de anomalias para definição de uma plataforma integrada de *Digital Twin*;
- Avaliar o processo de desenvolvimento de um modelo SVM para detecção de anomalias de um transportador de correias de minério de ferro;
- Avaliar aplicações de Internet das Coisas Industriais e integração de dados operacionais com a nuvem no contexto da plataforma integrada de *Digital Twin*.

1.3. Justificativa

As empresas do setor de mineração deverão repensar a forma como as informações são geradas e processadas. As plantas de mineração essencialmente deverão integrar planejamento, controle e suporte a decisões em toda a cadeia produtiva para otimizar volume, custo, investimento e segurança (DELLOITE, 2017).

Os transportadores de correia são equipamentos numerosos dentro uma usina de beneficiamento de minério de ferro, portanto, a sua utilização física tem impacto direto na produtividade da planta. Uma usina de beneficiamento de minério de ferro pode possuir uma ou mais linhas de tratamento de minério entre seus diversos processos de tratamento e, quando o transporte entre processos diferentes possui somente uma linha, este é chamado de circuito singelo, como exemplo destacado na Figura 1.1. Em circuitos singelos, a parada de uma correia transportadora pode causar uma parada geral da usina se os processos de tratamento posteriores à correia não possuírem estoque para manter sua operação.

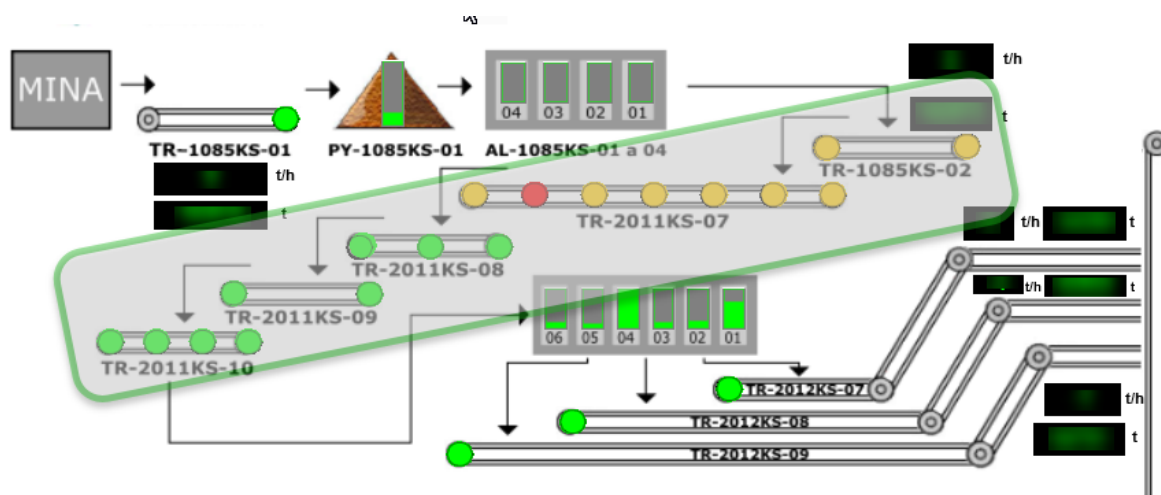


Figura 1.1: Circuito singelo de alimentação de uma usina de beneficiamento de minério de ferro. Fonte: Autor.

Um transportador de correia de um circuito singelo, portanto, impacta na utilização ou produtividade de todo o circuito, por exemplo: o circuito singelo apresentado na Figura 1.1 obteve uma utilização de 71,2% no intervalo de um ano devido principalmente a uma correia que faz circuito (ocorreram mais de 7.000 paradas, sendo que a maior parou o equipamento por 18 horas). Reforçando o exemplo anterior, a Figura 1.2 apresenta o gráfico da utilização física agregada em 2018 das correias transportadores do circuito singelo em questão e nota-se uma variabilidade considerável no mês de Junho.

A produtividade do circuito poderia aumentar se a operação na sala de controle for capaz de prever a ocorrência de eventos de parada para que uma decisão mais correta possa ser tomada, por exemplo: alterar os *set points* de operação para reduzir a carga ou programar uma parada (dependendo do tipo de evento previsto).

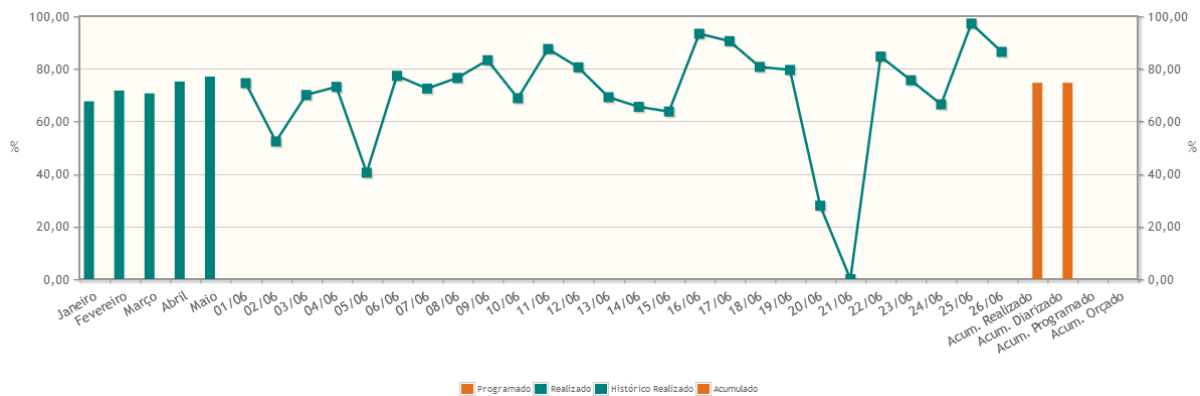


Figura 1.2: Gráfico da utilização física agregada das correias transportadoras. Fonte: Autor.

Essas informações não podem ser ignoradas, principalmente em face da grande variação negativa nos preços das commodities (em especial, o minério de ferro) nos últimos anos. O cenário global no mercado de minério de ferro para os próximos anos pressiona as empresas a continuar reduzindo seus custos operacionais, aumentar produtividade dos ativos e controlar a oferta de seus produtos com foco no aumento da margem de venda.

1.4. Organização do Trabalho

Este trabalho está organizado da seguinte forma:

1. **Capítulo 2:** apresenta o conceito de Digital Twin para o setor industrial desde a concepção do tema e aplicações ou protótipos identificados na literatura;
2. **Capítulo 3:** apresenta a metodologia deste trabalho, o processo de desenvolvimento do modelo SVM e as experiências industriais com aplicações de Internet das Coisas Industriais e integração de sistemas;
3. **Capítulo 4:** discorre sobre os resultados obtidos com o modelo SVM e a integração de sistemas para evolução de uma plataforma de Digital Twin;
4. **Capítulo 5:** apresenta as principais conclusões a respeito do projeto desenvolvido;
5. **Capítulo 6:** apresenta propostas de trabalhos futuros que contribuirão para a implementação de uma plataforma integrada de Digital Twin;

2. Referencial Teórico e Trabalhos Relacionados

Este capítulo apresenta um detalhamento do conceito de *Digital Twin* (DT) que fundamenta a base para o estudo deste trabalho, e também descreve o referencial teórico das tecnologias analisadas nos experimentos que basearam o desenvolvimento da arquitetura integrada proposta no decorrer do trabalho. Ao longo da seção são detalhados: as principais fontes de dados consideradas para uma plataforma de DT, o conceito de *Big Data* e *data lake* (importantes para a fundação do DT), as tecnologias de integração aplicadas no desenvolvimento da arquitetura do DT, o conceito de *Support Vector Machine* e uma breve explicação sobre os componentes de transportadores de correias. Por fim, são apresentados os trabalhos identificados na literatura que são relacionados ao desenvolvimento de modelos para detecção de anomalias em equipamentos industriais e ao desenvolvimento de plataformas integradas de DT com alguma aplicação de aprendizado de máquina.

2.1. *Digital Twin*

O *Digital Twin* (DT) é uma representação virtual (fiel e precisa) de uma entidade, processo ou sistema físico (por exemplo, equipamentos, produtos, etc.) e, dentro do conceito de sistema de produção cyber-físico (*Cyber Physical Production System*, CPPS), tem como objetivos o monitoramento, a análise e/ou a simulação de condições operacionais dos equipamentos com foco em objetivos específicos dos setores de manufatura como: aumentar produtividade, vida útil dos ativos ou reduzir custos operacionais. O termo *Digital Twin* foi descrito primeiramente pelo professor Dr. Michael Grieves em uma apresentação de 2002 para a indústria sobre *Product Lifecycle Management* (PLM) na Universidade de Michigan, sendo que, na época, era chamado de conceito ideal de PLM e que possuía todos os elementos básicos de um DT (GRIEVES e VICKERS, 2016). Em seu artigo, define o DT como um conjunto de construções das informações virtuais que descrevem por completo um produto manufaturado desde o nível micro-atômico ao nível macro-geométrico e estabelece dois tipos de DT conforme a seguir:

- ***Digital Twin Prototype (DTP)***: descreve o protótipo de um ente físico e contém as informações necessárias para descrever e produzir uma versão física que duplica a virtual. As informações associadas são relacionadas às especificações para a criação da versão física, tais como modelos 3D e a lista de materiais;
- ***Digital Twin Instance (DTI)***: descreve um ente físico específico e que permanece conectado a este durante todo seu ciclo de vida. As informações contidas na versão virtual podem ser modelos 3D após construção, especificação dos materiais atuais ou passados (caso houver algum serviço de reparo) e dados capturados de sensores podendo ser correntes ou em tempo real, históricos ou preditivos.

Grieves e Vickers (2016) também apresentam o *Digital Twin Environment* (DTE) como sendo o ambiente onde são realizadas operações de simulação independentes da localização da entidade física, podendo ser do tipo preditiva, cujo objetivo é prever estado futuro ou desempenho; ou interrogativa, aplicada somente sobre o DTI, cujo objetivo é monitorar o estado atual ou analisar o passado.

Shafte *et al.* (2012) introduzem o termo *Digital Twin* nos *roadmaps* tecnológicos da NASA e propostas para exploração espacial sustentável e um conceito de DT para a próxima geração de aviões militares e veículos espaciais. Os autores também apresentam a visão futura para os sistemas de engenharia baseados em simulação da NASA a partir do desenvolvimento de DT e o considera o terceiro desafio técnico a ser enfrentado pela NASA, em ordem de prioridade, ficando atrás de sistemas adaptativos para missões avançadas e sistemas integrados de simulações para simulações completas de missões.

Com foco no setor aeronáutico, Ríos *et al.* (2015) abordam os aspectos, sumarizados a seguir, que considera impactar na criação de um DT e os principais softwares em cada fase do ciclo de vida (Figura 2.1):

- **Identificação:** identificador global único que permite que se realize uma associação entre o ativo físico e sua representação virtual, e que possibilita a integração de todas as informações do objeto oriundas dos mais diferentes sistemas;
- **Ciclo de Vida:** definição das fases do ciclo de vida para que se possa implementar a gestão dos dados e informações gerados, por exemplo, *Begin of Life*, *Middle of Life* e *End of Life*;
- **Configurações:** processo de gestão da configuração para garantir que as informações e documentação da versão digital correspondem à versão física, dado que é comum um produto passar por diferentes configurações ao longo do seu ciclo de vida;
- **Informações:** armazenamento e análise avançada para processamento dos dados gerados pela versão física ao longo de seu ciclo de vida, como o volume de dados é substancial surge a necessidade de adoção de técnicas de *big data*;
- **Modelos:** integração dos diferentes modelos da versão física para cada fase do seu ciclo de vida para conferir maior fidelidade na relação entre as versões física e virtual.

Rosen *et al.* (2015) apresentam o DT como sendo um dos pilares para a implementação de processos ou plantas autônomas, impulsionado pela Internet das Coisas, alavancando a transformação digital no conceito da Indústria 4.0. Em seu trabalho, os autores entendem que o conceito de *Digital Twin* seria o próximo passo das tecnologias de modelagem e simulação (Figura 2.2), sendo que tais tecnologias evoluíram de aplicações individuais e restritas a especialistas na década de 1960, depois ferramentas de simulação passaram a ser um padrão a partir de 1985 e nos anos 2000 utilizam-se sistemas de design e engenharia totalmente baseados em

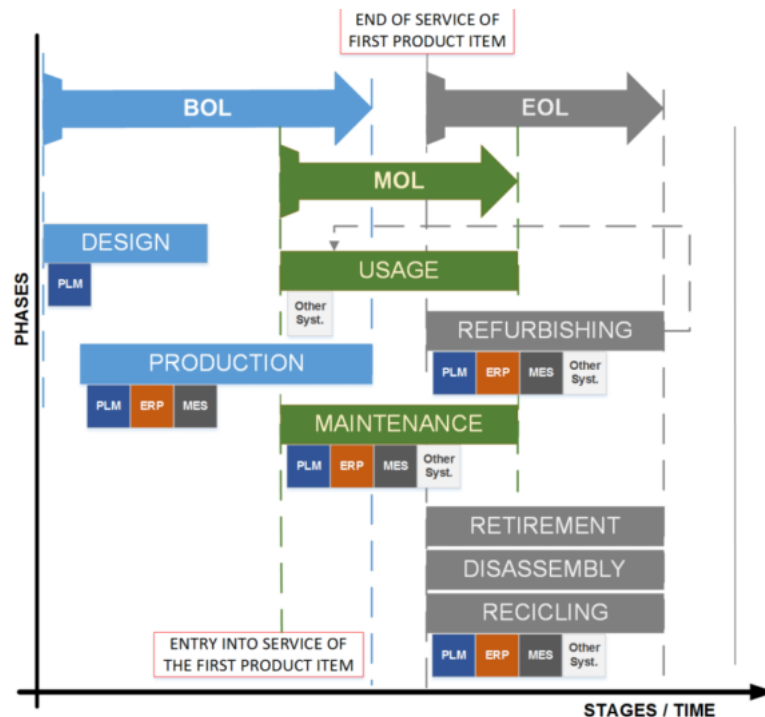


Figura 2.1: Sistemas e/ou software utilizados conforme a fase do ciclo de vida. Fonte: (RÍOS *et al.*, 2015)

simulação. Os autores citam o caso da Airbus Industrie, que desenvolveu um gêmeo físico de suas aeronaves, chamado *Iron Bird*, para simular, validar e otimizar sistemas vitais em tempo de projeto; mas depois converteu componentes para modelos virtuais, ou seja, o começo da criação de um DT.

Nesse contexto, fica evidente o desafio de sincronismo automático dos parâmetros dos equipamentos em operação com modelos virtuais devido ao desgaste de materiais, manutenções e etc., e acredita-se que novas ideias originadas de sistemas cyber-físicos (CPS) levarão a novos conceitos de manufatura que dependerão de integrações simples com modelos de simulação para atingir objetivos específicos das indústrias.

Tao e Zhang (2017) apresentam brevemente a evolução da convergência entre físico e digital na Figura 2.3 no âmbito da Indústria 4.0 e coloca três pontos que devem ser observados para que a convergência ocorra com sucesso.

- Como construir um modelo digital coexistindo com o ativo físico e com a mesma relevância para o chão-de-fábrica;
- Consistência entre os ativos físicos e o modelo digital;
- Convergência dos dados dos ativos físicos e modelo digital gerando informações para a produção.

Sob a ótica da digitalização e da Indústria 4.0, a captura, armazenamento e análise de dados será cada vez mais importante. Os ambientes de TI e TO nas corporações serão as bases

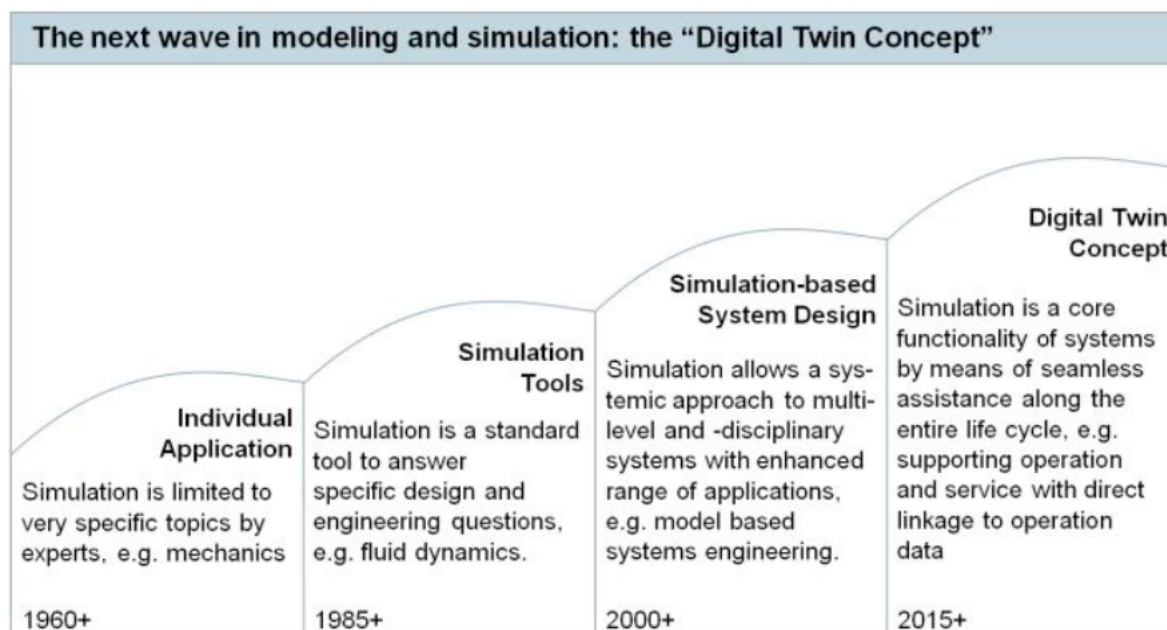


Figura 2.2: Evolução das tecnologias de simulação. Fonte: (ROSEN *et al.*, 2015)

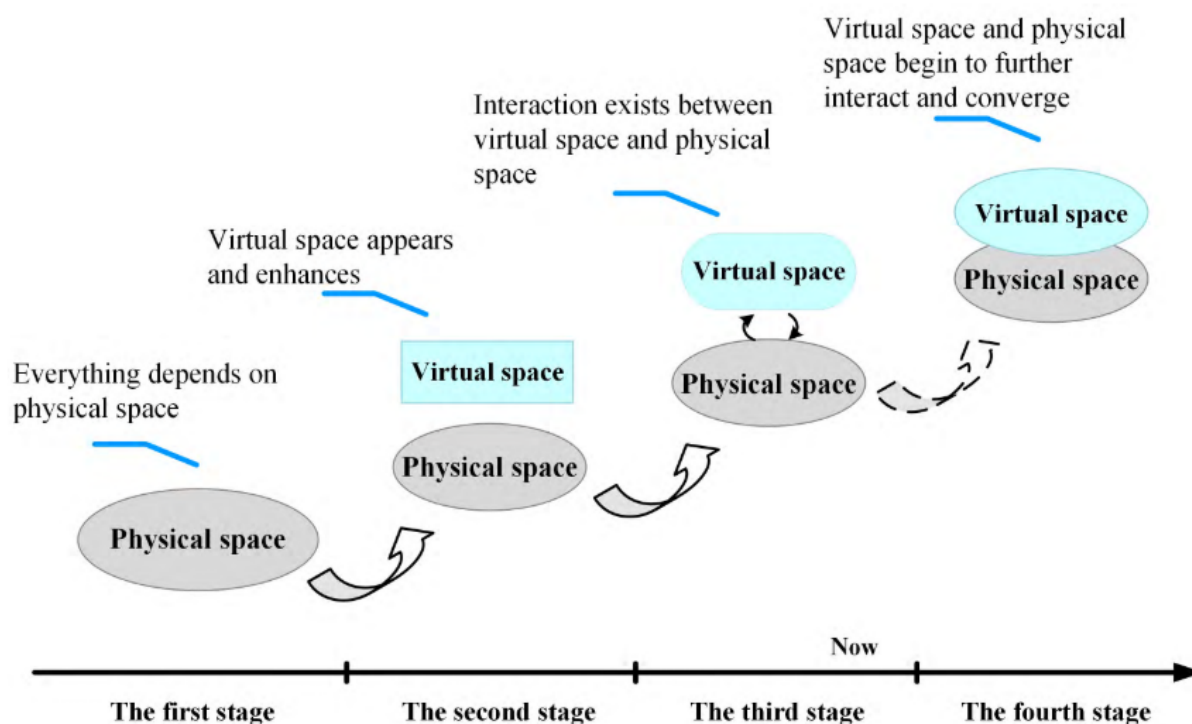


Figura 2.3: Evolução da convergência entre os mundos físico e digital. Fonte: (TAO e ZHANG, 2017)

para o *Digital Twin*, ou seja, as fontes dos dados serão os sensores (inteligentes ou não), sistemas de controle (por exemplo: controlador lógico programável - CLP, sistema digital de controle distribuído - SDCCD) e sistemas embarcados no chão de fábrica até os sistemas de gestão (por exemplo: PIMS, LIMS, MES, ERP) (UHLEMANN *et al.*, 2017). Um estudo apresentado em World Economic Forum (2017) mostrou um potencial impacto econômico de 321 bilhões de

dólares americanos até 2025 com a digitalização da indústria de mineração (Figura 2.4). As frentes de *smart sensors*, operações autônomas, plataformas integradas e análise avançada de dados têm relação direta com a implementação de Digital Twin.

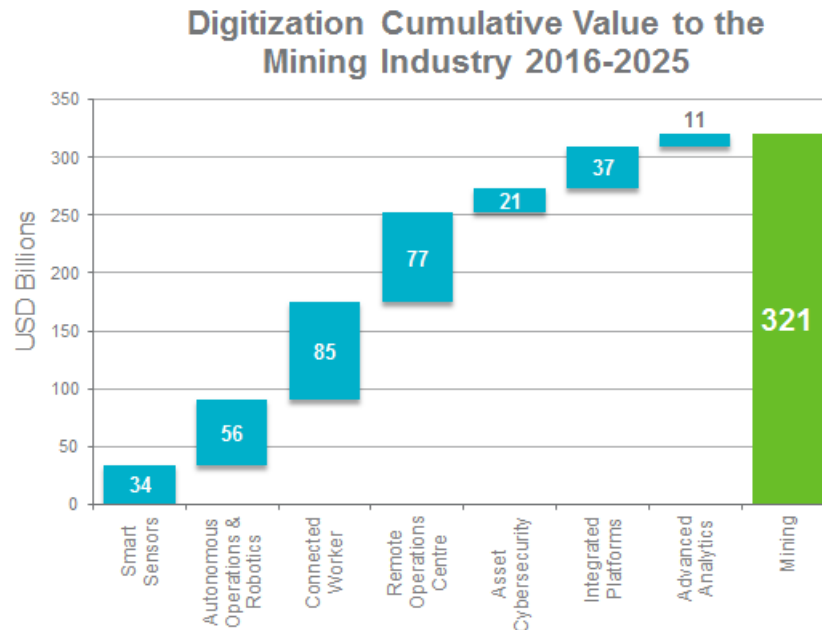


Figura 2.4: Potencial impacto econômico da digitalização na indústria da mineração. Fonte: adaptado de World Economic Forum (2017)

Os ganhos de produtividade decorrem, não limitados a esses exemplos, de melhor utilização dos ativos (advinda da otimização da programação de manutenções) ou correção dos parâmetros de operação dos equipamentos industriais (quando detectada mudança de especificação de produtos ou insumos).

Gartner (2017) estima que até 2021 metade das grandes empresas industriais implementará *Digital Twin* resultando em um ganho de até 10% em produtividade e incluiu a tecnologia pela primeira vez na curva *hype cycle* de tecnologias emergentes (Figura 2.5) como parte da tendência de plataformas digitais ao lado de Internet das Coisas e estima que atingirá maturidade entre 5 e 10 anos. Uma pesquisa publicada por Gartner (2018b) identificou que 48% das organizações implementando projetos de IoT estão implementando ou pretendem implementar DT em 2018 e o número de organizações utilizando DT crescerá três vezes até 2022.

Diante do que foi identificado na literatura a respeito de *Digital Twin* no setor industrial, pode-se considerar que o DT possui três elementos significativos:

- **Modelo:** estruturação dos componentes do DT, tipicamente encontrada em forma de hierarquia;
- **Analytics:** algoritmos de análise avançada de dados (por exemplo, aprendizado por máquina) com finalidade de descrever, prever ou prescrever a operação da entidade física, podendo, inclusive, ser a fundação para uma operação autônoma;

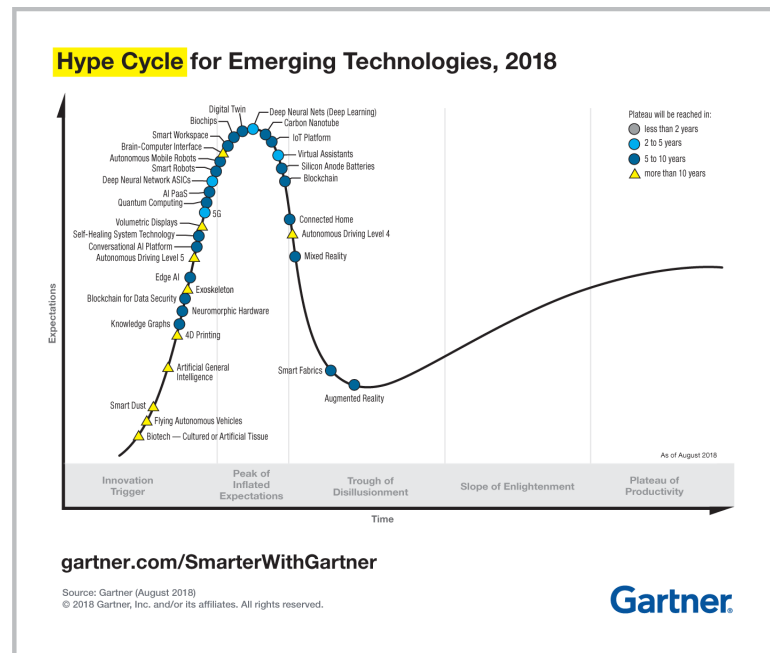


Figura 2.5: Curva *hype cycle* para tecnologias emergentes. Fonte: (GARTNER, 2018a)

- **Base de conhecimento:** conjunto de todos os dados (por exemplo, séries temporais de sensores) e resultados derivados dos algoritmos de análise avançada.

Os elementos mencionados acima formam a base de uma DTI e o objeto deste trabalho explora o elemento de *analytics* ao lado de experiências com aplicações industriais para refinamento dos requisitos (no que diz respeito a este elemento) em uma proposta de arquitetura de sistema para *Digital Twin*.

2.2. Fontes de Dados

2.2.1. Process Information Management System (PIMS)

No final da década de 1990, a maioria das indústrias no Brasil possuíam operações pouco integradas formando ilhas de informação, ou seja, os dados eram gerados pelos sensores e sistemas de controle em cada operação de forma não padronizada e utilizados de formas diferentes em relatórios operacionais ou de produção. Tal situação apresentava dois problemas para a indústria na época (BECERRA e PALACIOS, 2003):

- Ausência de uma base de dados comum e padronizada implicava em dados duplicados e/ou conflitantes;
- Ausência de conectividade ou integração entre o sistema de planejamento corporativo, em nível estratégico, e o sistema de controle dos processos, operando em tempo real.

O PIMS foi desenvolvido com objetivo primário de integrar os dados dos sistemas de controle de processos e eliminar as ilhas de informação existentes em uma planta industrial. A tecnologia implementada pelo PIMS integra, em tempo real, dados automáticos e manuais dos processos industriais e permitiu maior integração de pessoas, outras tecnologias, procedimentos e estratégias (BECERRA e PALACIOS, 2003). O sistema, então, possibilitou o aumento de produtividade das plantas industriais dado que as informações passarem a ser geradas e disponibilizadas de forma mais fácil e uniforme para as pessoas corretas e no momento adequado.

Esses sistemas, atualmente, têm a capacidade de coletar e armazenar milhares de dados em *tags*, ou pontos, disponíveis nas mais diversas fontes de dados (CLP, SDCD, LIMS, MES, entradas manuais, etc.). Os softwares de PIMS possuem alta resiliência devido à sua arquitetura de coleta de dados e funcionalidades de alta disponibilidade dos servidores do banco de dados histórico-temporal. Os coletores de dados são tipicamente conectados à rede de supervisão e controle, próximos das fontes de dados no chão de fábrica, e utilizam a tecnologia *OLE for Process Control* (OPC) para conectar-se a um servidor que fornecerá os dados. Estes coletores possuem mecanismos de *buffering* dos dados coletados em tempo real para armazenar os dados por um período curto de tempo na eventual ocorrência de uma falha de comunicação com o servidor do banco de dados histórico-temporal (Figura 2.6). (ASPEN TECHNOLOGY, 2014, 2016, 2017)

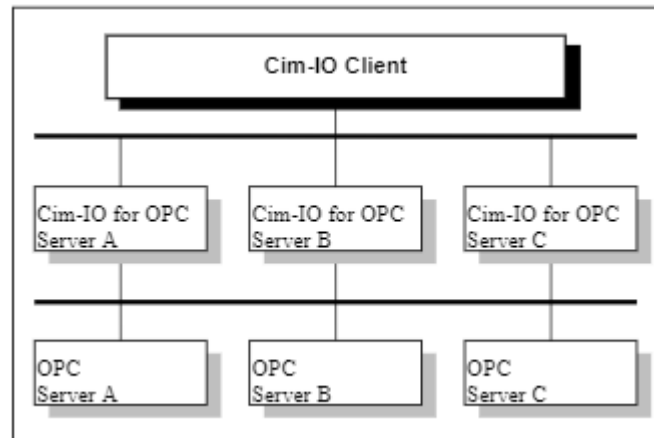


Figura 2.6: Arquitetura de um coletor de dados de um PIMS utilizando OPC. Fonte: (ASPEN TECHNOLOGY, 2016)

O servidor do banco de dados histórico-temporal executa algoritmos de compressão sobre cada dado recebido dos coletores distribuídos pelo chão-de-fábrica (ASPEN TECHNOLOGY, 2017). Esta funcionalidade, aliada à redução do custo dos discos rígidos e armazenamento na nuvem, permite que dados fiquem disponíveis *online* por um período de tempo grande, em alguns casos ultrapassando 20 anos.

Segundo Becerra e Palacios (2003), as aplicações do PIMS são das mais variadas e em diferentes níveis de uma corporação (desde o analista e operador ao supervisor, gerente e diretor), exemplificando:

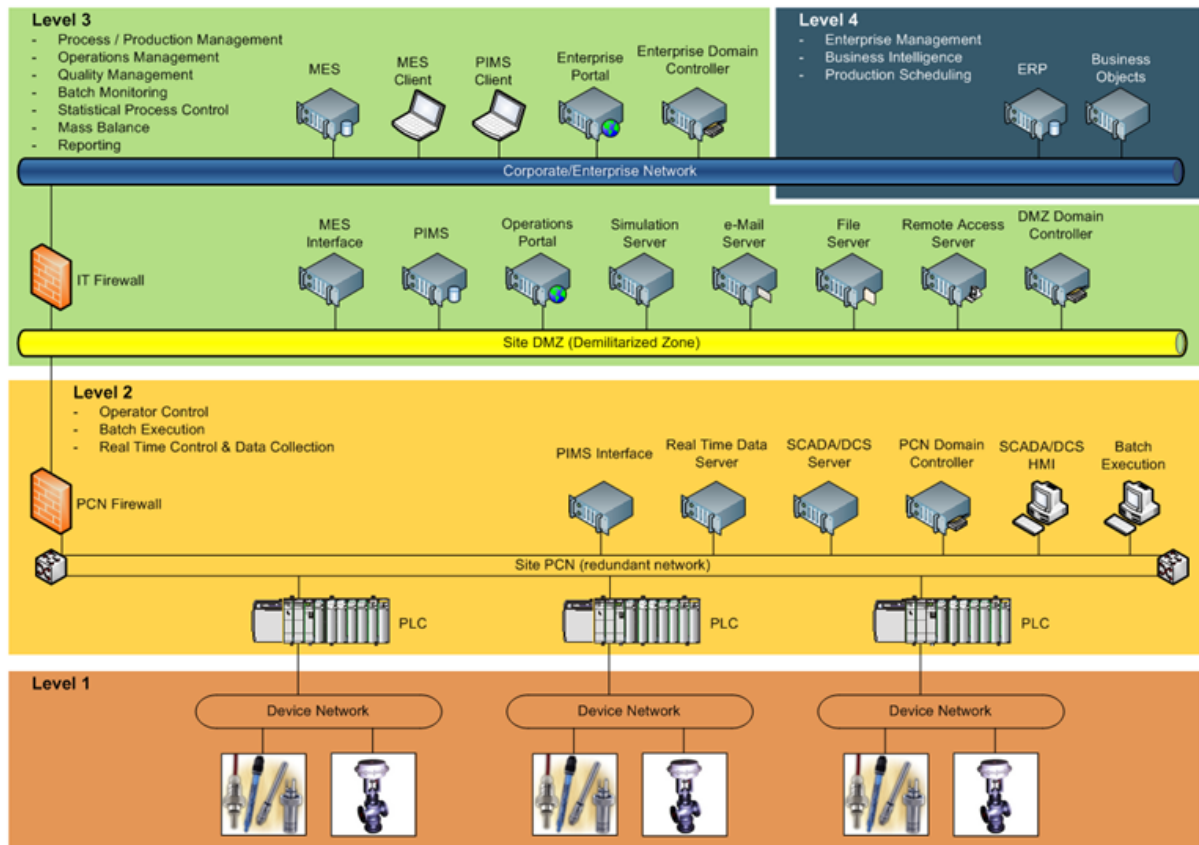
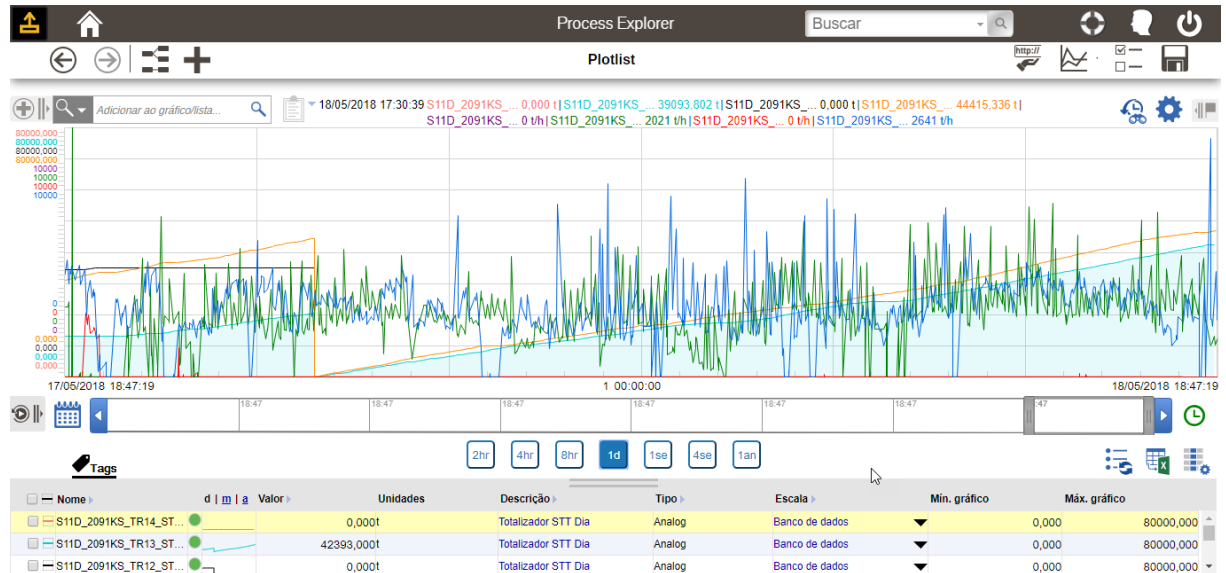


Figura 2.7: O PIMS inserido no contexto da norma ISA95. Fonte: Autor.

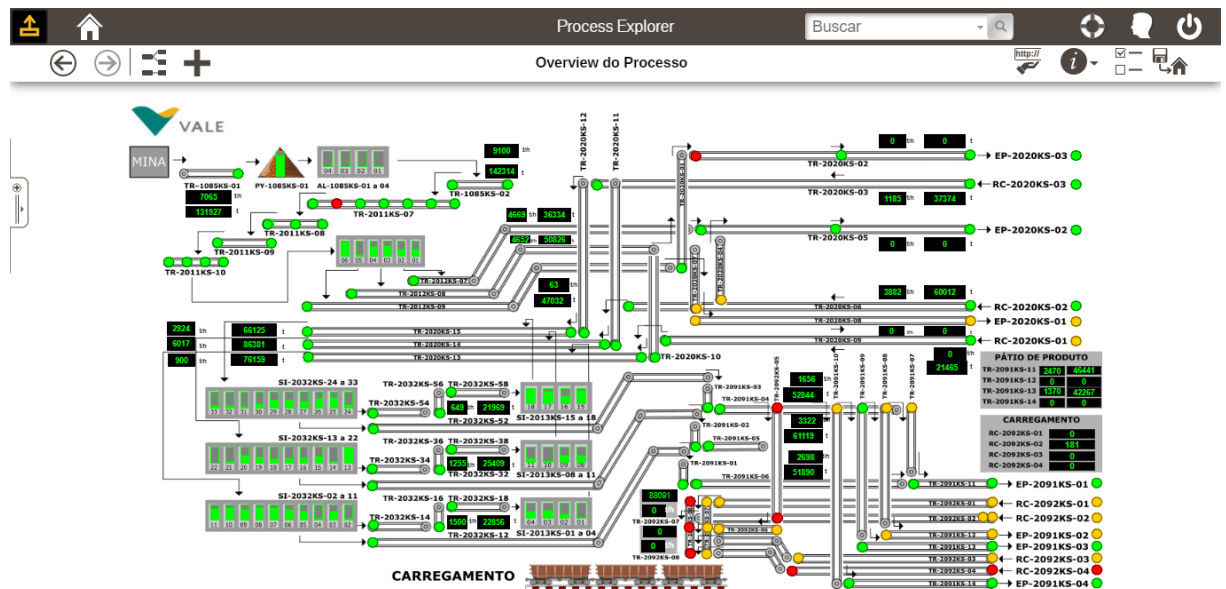
- Engenharia de processos: a facilidade de acesso aos dados históricos possibilita, por exemplo, monitorar os processos para identificar períodos ótimos de operação e definir modelos de operação que os operadores devem seguir com foco na otimização da produção e/ou aumento da durabilidade dos equipamentos;
- Manutenção: a equipe de manutenção, antes isolada da operação da planta, passa a ter acesso em tempo real á informações dos equipamentos em operação sem necessidade de deslocar-se para uma sala de controle para coletar dados para suas análises. O PIMS torna-se uma ferramenta essencial para análises de causa raiz de problemas com os equipamentos de processo;
- *Key Process Indicator* (KPI): a centralização dos dados em uma única base de dados permite o cálculo de indicadores de desempenho da planta como um todo em tempo real, não mais limitado à uma operação da planta, e a correlação de milhares de variáveis entre si co objetivo .
- Integração entre sistemas: as informações são disponibilizadas de forma mais simples para sistemas corporativos (ERP). Kolz (2000) cita um exemplo de aplicação do Monitoramento Baseado na Condição (CBM) onde o ERP é configurado para disparar uma ordem de manutenção a partir da totalização de carvão produzida na planta. Em contra-

partida à manutenção pelo calendário, a CBM otimiza produção e reduz custos da planta pois aumenta a disponibilidade do equipamento ao reduzir a frequência de manutenções no mesmo.

A Figura 2.8 demonstra algumas aplicações típicas do PIMS encontradas nas operações da Vale.



(a) Gráfico de tendência



(b) Tela sinótica

Figura 2.8: Exemplo de aplicações típicas do PIMS. Fonte: Autor.

2.2.2. Manufacturing Execution System (MES)

Manufacturing Execution System, ou sistema de gestão da produção, é um software (pacote comercial ou desenvolvido sob medida) utilizado para apoiar o controle dos processos

de gestão em uma operação industrial, como: insumos e material, capacidade dos equipamentos, qualidade, estoque, rastreabilidade, otimização, etc. O MES fornece ferramentas que permitem que as operações das plantas otimizem a produção através da redução dos ciclos da produção, da qualidade dos produtos, melhoria na gestão dos estoques ou aumento de produtividade (HELO *et al.*, 2014).

Segundo apresentado em Kletti (2007), os processos produtivos podem ser classificados em discretos, contínuos e *make-to-order*, segundo a descrição a seguir.

- Produção discreta: a produção é organizada em ordens a partir de uma série de operações que, em alguns casos, pode ser agrupadas em linhas de montagem. A disponibilidade dos recursos é uma variável essencial para permitir um melhor sequenciamento das ordens de produção;
- Produção contínua: as indústrias de processo contínuo produzem grande quantidade de um determinado produto e, tipicamente, tem pequena variabilidade de produtos. Os equipamentos e/ou circuitos de produção são organizados em linha e operam em regime 24 x 7, o que justifica o menor número de produtos tendo em vista o custo para mudança da linha de produção. A variável mais importante nesse tipo de indústria é a utilização dos equipamentos;
- *Make-to-order*: essa indústria utiliza uma lista de materiais que são processados em células de produção independentes, significando que o tempo de transição muitas vezes não é crítico.

As funcionalidades e requisitos para as aplicações de um MES, portanto, mudam conforme o tipo de manufatura das empresas. A Figura 2.9, descrita por Kletti (2007) demonstra em alto nível como a abrangência das funcionalidades de um MES pode mudar de um tipo de manufatura para outro.

Existem diversas implementações de MES sendo que as mais conhecidas são para as indústrias de produção discreta e contínua. Para a produção discreta, o MES objetiva ser um sistema online de informações para efetivo controle da produção. Já para a produção contínua, o MES seria formado em sua maior parte pelos sistemas de controle e supervisão dos processos e equipamentos da planta (KLETTI, 2007).

No esforço de padronização das funcionalidades e requisitos do MES destacam-se duas organizações sem fins lucrativos: Manufacturing Enterprise Solutions Association (MESA) e The International Society of Automation (ISA); com membros de empresas de manufatura, fornecedores de software, integradores de sistemas, academia e estudantes (MANUFACTURING ENTERPRISE SOLUTIONS ASSOCIATION, 2018; THE INTERNATIONAL SOCIETY OF AUTOMATION, 2018). No contexto do presente trabalho, considera-se mais importante os padrões definidos pela norma ISA95 que dispõe sobre a integração de dados entre sistemas de controle e sistemas corporativos.

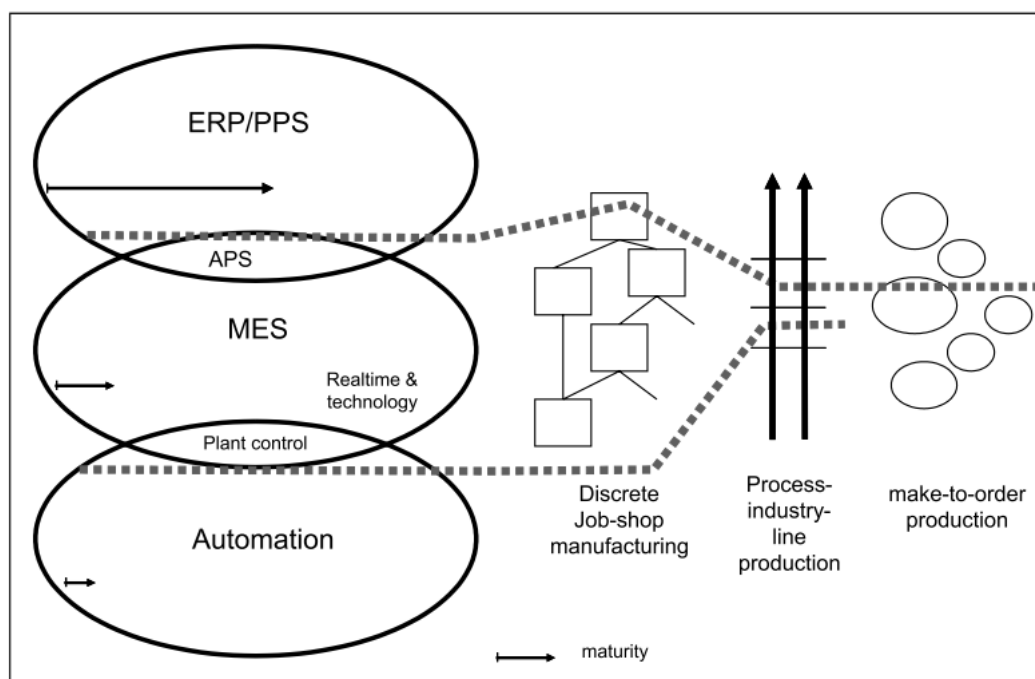


Figura 2.9: Variação das funcionalidades MES conforme os tipos de manufatura. Fonte: (KLETTI, 2007)

Em algumas corporações os dados relacionados aos eventos de paradas e produção dos equipamentos são registrados no MES, pois dependem de entradas manuais dos operadores de salas de controle para classificar o evento conforme o tipo da parada (Figura 2.10). Essa informação é importante para o DT, no âmbito deste projeto, porque o modelo pode correlacionar com dados oriundos do PIMS e ERP para aumentar a assertividade do resultado.

2.2.3. Internet das Coisas Industriais

2.3. *Big Data*

O conceito de *Big Data* (BD) refere-se ao crescimento em larga escala dos dados em praticamente todos os campos nas últimas décadas e prevê-se que o volume de dados gerados crescerá muito mais nos anos seguintes devido à expectativa de aumento no número de dispositivos conectados gerando dados. Vários conceitos de BD podem ser encontrados na literatura, conforme Addo-Tenkorang e Helo (2016), tais como a próxima fronteira para inovação, competição e produtividade ou a nova geração de tecnologias e arquiteturas desenvolvidas para que empresas possam explorar economicamente o grande volume de dados gerado pelas mais variadas fontes através da coleta, descoberta, armazenamento e análise em alta velocidade (este último, apresentado pelo IDC, sendo o mais aceito). A definição apresentada pelo IDC por Addo-Tenkorang e Helo (2016) descreve cinco atributos de BD conforme abaixo.

- **Variedade:** número cada vez maior de fontes de dados disponíveis para aquisição e pro-

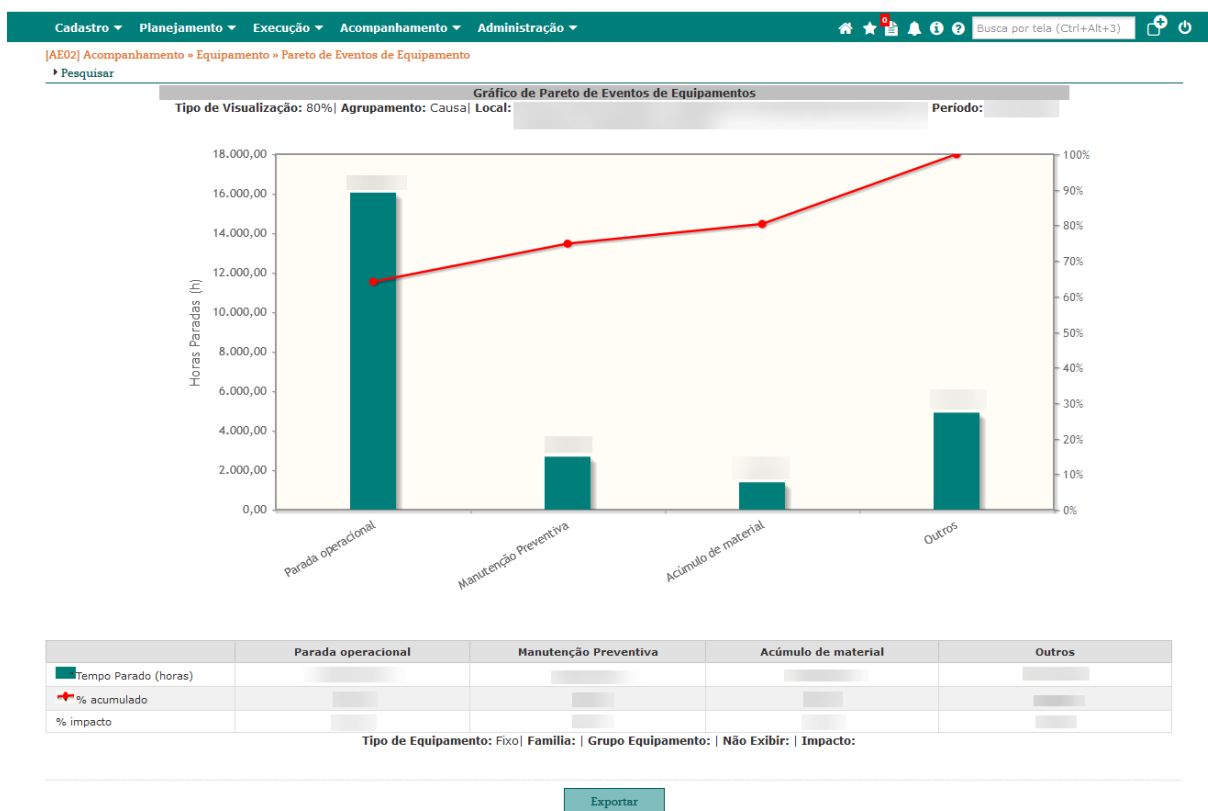


Figura 2.10: Relatório apresentando dados consolidados de eventos de equipamento em processo contínuo de beneficiamento de minério de ferro. Fonte: Autor.

cessamento relacionado ao desenvolvimento de tecnologias como IoT, dispositivos inteligentes, *wearables*, entre outros;

- **Velocidade:** aumento crescente na frequência com que os dados são coletados das fontes;
- **Volume:** aumento na quantidade de dados sendo coletados ou escritos nos sistemas de bancos de dados existentes;
- **Veracidade:** indica incerteza ou imprecisão dos dados. As empresas buscam implementar algoritmos para automatizar a preparação dos dados antes da análise;
- **Valor Agregado:** valor ou o benefício que a indústria obtém ao implementar as tecnologias de BD;

A relação entre os cinco atributos está representada na Figura 2.11, que pretende demonstrar que deve-se maximizar o valor agregado processando grandes volumes de dados, com variedade cada vez maior, em alta velocidade e com garantia maior da veracidade, tudo isto em um curto período de tempo.

Dubey *et al.* (2016) apresentam, em seu artigo, razões para considerar que BD e Big Data Analytics (BDA) foram alvo de discussões nos meios acadêmico e industrial como, por

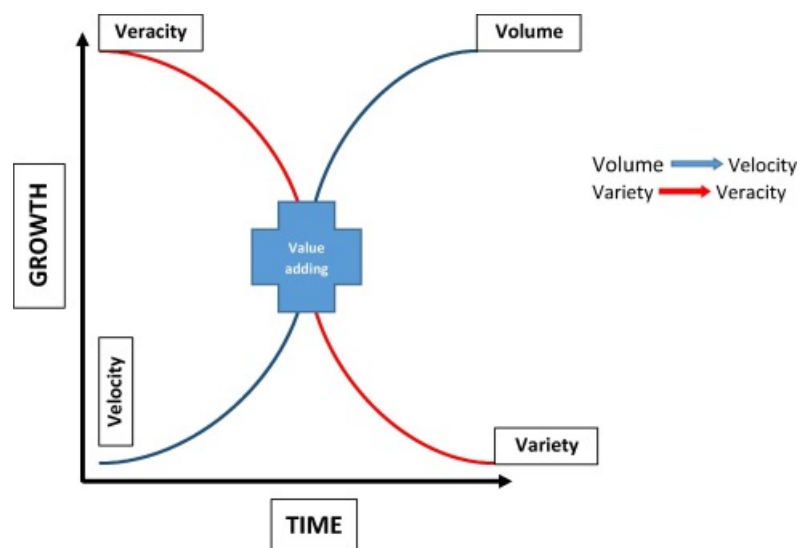


Figura 2.11: Relação entre os cinco Vs conforme a definição de *Big Data* do IDC. Fonte: (ADDO-TENKORANG e HELO, 2016)

exemplo, a convergência entre IoT, computação na nuvem e ativos inteligentes. No mesmo trabalho, consideram cinco aspectos com potencial para causar alguma ruptura no campo da gestão das operações industriais:

1. BD e análise preditiva (*predictive analytics*);
2. Manufatura aditiva;
3. Veículos autônomos;
4. Ciência dos materiais;
5. Cadeia de suprimentos “sem fronteiras”;

Na indústria de processos contínuos, algumas equipes de TO consideram que *Big Data* existe há algum tempo partindo do princípio que um grande volume de dados são coletados dos sensores no chão de fábrica e processados em tempo real pelos sistemas de controle.

Em Morariu *et al.* (2018) é apresentada uma aplicação de BDA (*map-reduce*) para descobrir padrões de consumo energético de equipamentos através de regressão linear. A partir desta regressão é possível prever o consumo do equipamento e aplicar mudanças nos parâmetros de operação para otimizar o consumo ou alterar a linha de produção.

2.3.1. *Data Lake*

O termo *data lake* foi cunhado em 2010 por James Dixon para distinguir a abordagem para gestão de dados em *data marts* ou *warehouses* e Hadoop. *Data lakes* normalmente são planejados como um local de armazenamento central em uma organização, empresa ou instituição

Tabela 2.1: Tecnologias e ferramentas normalmente utilizadas pelas aplicações BDA

Tecnologia ou Ferramenta	Definição
Computação na nuvem	Utilização de uma infraestrutura de serviços de computação e processamento de dados baseado em servidores e/ou <i>cluster</i> de servidores com a finalidade de garantir maior disponibilidade, escalabilidade e menor custo operacional para aplicações. (ADDON-TENKORANG e HELO, 2016)
Apache Hadoop	Biblioteca de software que fornece uma plataforma para processamento distribuído de conjuntos de dados muito grandes em <i>clusters</i> de servidores utilizando modelos de programação simples. (THE APACHE SOFTWARE FOUNDATION, 2018c)
MapReduce	Modelo de programação para processamento paralelo de conjuntos de dados muito grandes baseado em duas funções: <i>map</i> e <i>reduce</i> . A primeira função converte os dados em pares de chave/valor; a segunda função aplica uma ou mais operações sobre os valores de uma determinada chave. (MORARIU <i>et al.</i> , 2018)
Apache Spark	Sistema de computação em <i>cluster</i> para análise de dados em larga escala com APIs de alto nível, por exemplo, Java, Python e R. (THE APACHE SOFTWARE FOUNDATION, 2018a)
Apache Kafka	Plataforma distribuída de integração de dados baseada em um barramento <i>publish/subscribe</i> com capacidade de <i>streaming</i> e tolerante à falhas. (THE APACHE SOFTWARE FOUNDATION, 2018b)

onde qualquer tipo de dados de qualquer tamanho pode ser copiado em qualquer taxa utilizando qualquer método de importação (inicial, batelada, *streaming*) em seu formato original (nativo, *raw*) (MATHIS, 2017).

Fang (2015) apresenta o *data lake* como uma metodologia apoiada por um repositório massivo de dados baseado em tecnologias de baixo custo que melhoram a captura, refinamento, arquivamento e exploração de dados dentro de uma organização, um conceito atrelado fortemente ao Apache Hadoop que se tornou popular devido à relação custo/benefício frente aos desafios postos pelo *big data*. O autor compara os principais aspectos dos bancos de dados corporativos (*Enterprise Data Warehouse*) e dos *data lakes*, apresentado na Tabela 2.2, e denota uma tendência para as empresas migrarem suas arquiteturas para à medida que ganham mais conhecimento prático.

Um *data lake* não possui um esquema de dados bem definido para permitir a ingestão rápida de dados ao evitar tarefas complexas de modelagem dos dados ou esforços de integração. No entanto, não é possível trabalhar com os dados sem que exista um esquema em algum momento, que pode ser no armazenamento dos dados (*on-load*), no consumo dos dados (*on-read*)

Tabela 2.2: Comparação entre os bancos de dados corporativos e *data lake*. Fonte: adaptado de Fang (2015)

Aspecto	<i>Enterprise Data Warehouse</i>	<i>Data Lake</i>
Processamento	Centenas a milhares de usuários concorrentes; Análise interativa dos dados; Capacidade avançada de gestão do banco; Processamento em bateladas.	Processamento de dados em larga escala; Está aumentando capacidade de processamento paralelo para suportar cada vez mais usuários concorrentes.
Esquema	Definido na criação dos banco de dados; Requer maior trabalho no início, mas oferece desempenho, segurança e integração.	Normalmente definido na leitura; Oferece muita agilidade e facilidade na captura de dados ao custo de maior trabalho no final do processo; Funciona bem com tipos de dados cujos valores não são conhecidos.
Escala	Grande volume de dados a um custo moderado.	Volumes de dados extremos a baixo custo.
Acesso	Dados são acessados com linguagem SQL e ferramentas de BI; Pesquisa por busca.	Dados acessados por programas criados por desenvolvedores; Pesquisa por varredura.
Consulta	SQL	Programação
Custo	Uso eficiente de CPU e I/O.	Baixo custo de processamento e armazenamento.
Complexidade	Joins complexos.	Processamento complexo.

ou sob-demanda *on-need*) (FANG, 2015). Walker e Alrehamy (2015) relatam que a ausência de um esquema de dados ou gestão de metadados para pode transformar o *data lake* em um *data swamp*, tornando os dados incompreensíveis e mais difíceis de serem interpretados e assimilados. Os autores sugerem a transmissão dos metadados juntamente com os dados para aumentar sua compreensibilidade, mas tal implementação pode tornar a solução rígida como um sistema de banco de dados corporativo tradicional.

A tecnologia de *data lakes* está em franco desenvolvimento, seja baseado em componentes do Apache Hadoop ou software proprietário, por várias empresas a exemplo de Google, Microsoft, Amazon e SAP (MATHIS, 2017). A plataforma Azure, disponibilizada pela Microsoft a nível global para a Vale, possui várias soluções que poderiam ser utilizadas como *data lake* e este trabalho foca no Azure Dara Lake Store (ADLS). Este serviço oferecido na plataforma Azure é um sistema de arquivos totalmente gerenciável, elástico, escalável e seguro, compatível com o sistema de arquivos Hadoop (HDFS) e semântica do Cosmos (sistema de arquivos utilizado em projeto da Microsoft) e que foi otimizado para aplicações de BDA demandando alto nível de paralelismo para leitura/escrita, processamento de dados e baixa

latência no acesso (RAMAKRISHNAN *et al.*, 2017). Uma parte fundamental do ADLS é a adição de um *namespace* hierárquico para armazenamento que organiza objetos/arquivos em uma hierarquia de diretórios para acesso eficiente a dados. Uma convenção de nomenclatura de armazenamento de objetos comum usa barras no nome para imitar uma estrutura de diretórios hierárquica. Operações como renomear ou excluir um diretório tornam-se operações únicas de metadados atômicos no diretório, em vez de enumerar e processar todos os objetos que compartilham o prefixo de nome do diretório (MICROSOFT, 2018).

2.4. Integração de Sistemas

Um dos principais desafios no desenvolvimento de uma plataforma é a integração dos seus serviços com os sistemas existentes em uma empresa. No âmbito deste trabalho, este desafio extrapola o ambiente corporativo, onde as tecnologias de integração são mais implementadas, porque demanda integração com sistemas de controle e supervisão da planta para a disponibilização de dados operacionais dos equipamentos ou processos que deseja-se implementar um DT. Por conta disso, é necessário analisar os principais conceitos, ferramentas e práticas empregadas para a integração de aplicações complexas e heterogêneas com um foco nos requisitos identificados para uma plataforma de DTE.

2.4.1. Arquitetura Orientada a Serviços

Endrei *et al.* (2004) introduz a arquitetura orientada a serviços (SOA, do inglês Service-Oriented Architecture) como um paradigma de integração de sistemas para abordar os problemas da heterogeneidade (co-existência de diferentes aplicações, arquiteturas, tecnologias e fornecedores de TI) e mudança (diferentes requisitos, entregas mais rápido devido ao cenário mais competitivo) acompanhando a evolução de uma organização empresarial típica dos anos 80 para um novo ecossistema de negócios mais distribuído. Para endereçar estes problemas, tal arquitetura precisa implementar serviços com as seguintes características: acoplamento fraco e independentes de local e protocolo. Erl (2008) estabeleceu princípios do paradigma da SOA para o desenvolvimento de serviços com interoperabilidade, ou seja, independentes da implementação das aplicações envolvidas em uma integração:

- **Contrato de Serviço Padrão:** considerado fundamental para SOA, estabelece o propósito, funcionalidades, tipos ou modelos de dados e políticas dos serviços;
- **Acoplamento Fraco:** relacionado à dependência entre o contrato, implementação e consumidores dos serviços. Um acoplamento fraco permite maior independência na evolução da lógica e implementação ao mesmo tempo que garante interoperabilidade com consumidores;

- **Abstração:** relacionado ao acoplamento ao permitir que se omita o maior nível de detalhes de um serviço, impacta na governança do serviço;
- **Reutilização:** altamente recomendada na SOA, define que serviços sejam desenvolvidos em um nível corporativo com contexto funcional agnóstico, ou seja, pode ser consumido por diferentes aplicações sem a necessidade de modificações;
- **Autonomia:** relacionado à capacidade dos serviços de executar consistentemente e de forma confiável independente do ambiente em que se encontram;
- **Ausência de Estado:** relacionado à independência das solicitações ou sessões de um serviço das anteriores;
- **Descoberta:** relacionado à interpretação do serviço pelos consumidores, ou seja, as informações dos contratos devem ser claras suficiente para promover a sua descoberta e reutilização;
- **Composição:** relacionado à ausência de estado, implica em desenvolver serviços com objetivos específicos e bem definidos para tarefas ou requisitos pequenos, e a composição desses serviços atende a requisitos de negócio complexos.

Serviços baseados nos princípios acima reduzem o impacto para seus consumidores nas situações onde há uma atualização do sistema. No entanto, em uma eventual substituição do sistema, os consumidores seriam afetados porque normalmente ocorrem mudanças no contrato e dificulta a descoberta do serviço. Para endereçar este problema (e também reduzir a complexidade para os departamentos de TI), são implementados Barramentos de Integração que expõem os serviços de uma organização de forma mais abstrata.

2.4.2. Barramento de Integração

Barramento de integração (ESB, do inglês Enterprise Service Bus) não é um produto em si, mas uma prática de arquitetura de integração para implementações da SOA que estabelece um barramento de mensagens a nível corporativo que combina a infraestrutura de mensageria com transformação e roteamento de conteúdo em uma camada lógica de integração entre consumidores e provedores de serviços. O principal objetivo de um ESB é permitir o desenvolvimento e gestão das lógicas de negócios independentes da infraestrutura, rede e provisionamento dos serviços (ENDREI *et al.*, 2004), ou seja, permitir que aplicações distintas entre si, distribuídas pela organização, comuniquem-se de forma simples, transparente e centralizada conforme demonstrado na visão geral apresentada na Figura 2.12.

Keen *et al.* (2004) define, conforme apresentado na Tabela 2.3, os recursos mínimos que um ESB deve implementar para prover uma infraestrutura consistente com os princípios da SOA.

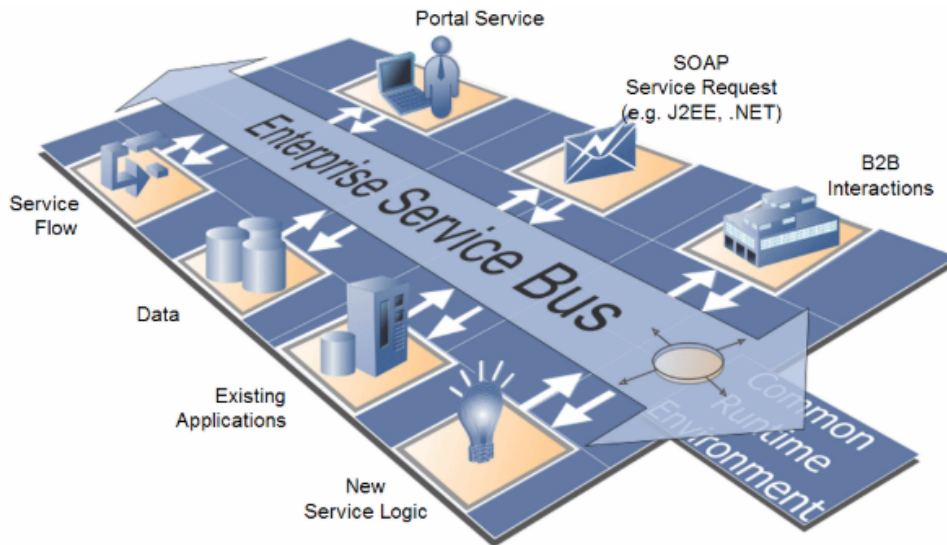


Figura 2.12: Visão geral de um ESB. Fonte: (KEEN *et al.*, 2004)

2.4.3. *AutomationML*

O processo de engenharia de sistemas de produção é complexo porque envolve diversas disciplinas da engenharia que criam diferentes artefatos em diferentes ferramentas sem a devida integração de dados entre si, o que aumenta o volume de trabalho humano em cada etapa do processo. Para endereçar o problema de integração entre as aplicações, deve-se considerar um padrão formatação atendendo aos requisitos a seguir e que normalmente levam à adoção do XML (LÜDER e SCHMIDT, 2015):

- O formato das mensagens podem ser adaptados conforme o caso de uso e flexíveis para extensões;
- A representação dos dados deve ser eficiente;
- A representação dos dados deve ser baseada em padrões internacionais;

AutomationML é um padrão de formato aberto (IEC 62714), baseado em XML, desenvolvido por um consórcio de empresas e universidades com foco na integração sem perda de dados. Este padrão busca integrar os diferentes sistemas utilizados nas etapas de um projeto de engenharia, desde a engenharia básica ao comissionamento de sistemas de produção, sendo capaz de transmitir dados de topologia, mecânicos, elétricos, pneumáticos, hidráulicos, funcionais e de controle de processos. A utilização do padrão vem crescendo recentemente, considerando os artigos publicados na literatura, e sua adoção para integração no âmbito de DT é relevante pois permite modelar os dados para a criação ou atualização das instâncias conforme metodologia apresentada no trabalho de Schroeder *et al.* (2016).

Kovalenko *et al.* (2018) descrevem uma metodologia representação baseada em ontologia para o *AutomationML* com objetivo de promover maior adoção de tecnologias de semântica

Tabela 2.3: Recursos mínimos de um ESB. Fonte adaptado de Keen *et al.* (2004)

Categoria	Recursos	Motivo
Comunicação	- Roteamento - Endereçamento - Ao menos um padrão de mensagem - Disponibilizar ao menos um protocolo de transporte	Endereça a transparência de localização e permite substituição de serviços.
Integração	- Diferentes estilos de integração e adaptadores - Transformação entre protocolos	Suporta a integração em ambientes heterogêneos e permite a substituição de serviços.
Interação	- Definição da interface dos serviços - Modelo de serviços de mensageria - Substituição da implementação do serviço	Separar o código da aplicação dos protocolos específicos dos serviços e suas implementações.
Gerenciamento	- Capacidade de administração	Controle sobre o endereçamento e nomenclatura dos serviços.

na engenharia de sistemas ciber-físicos de produção. Fischer *et al.* (2017) apresentam um ambiente de simulação como serviço capaz de criar modelos automaticamente utilizando o *AutomationML* para integração de dados através de interfaces REST em uma arquitetura seguindo o paradigma SOA. No trabalho, os autores identificam que os arquivos de dados tendem a um tamanho grande devido ao *overhead* do formato, mas não impactam na utilização do serviço.

2.5. Modelo para Detecção de Anomalias

2.5.1. *Support Vector Machine* (SVM)

Support Vector Machine é uma técnica de aprendizado supervisionado, considerada como uma extensão do *perceptron*, tipicamente empregada em problemas de classificação e regressão. A técnica consiste em estabelecer hiperplanos que maximizam as margens, i.e., as distâncias entre os limites de decisão e as amostras do conjunto de dados de treino que estão próximas desses limites (chamadas de *support vectors*) (RASCHKA e MIRJALILI, 2017). A Figura 2.13 descreve, em alto nível, a técnica considerando um espaço de duas variáveis, nota-se que podem ser traçados diversos limites de decisão neste espaço e o resultado do algoritmo de treino é um plano (linha tracejada no gráfico à direita) que maximiza a distância para os *support vectors* (destacados com círculos no mesmo gráfico).

A razão para maximizar as margens como função objetivo é porque ela tende apresentar menor erro de generalização enquanto que modelos com margens menores são mais susceptíveis

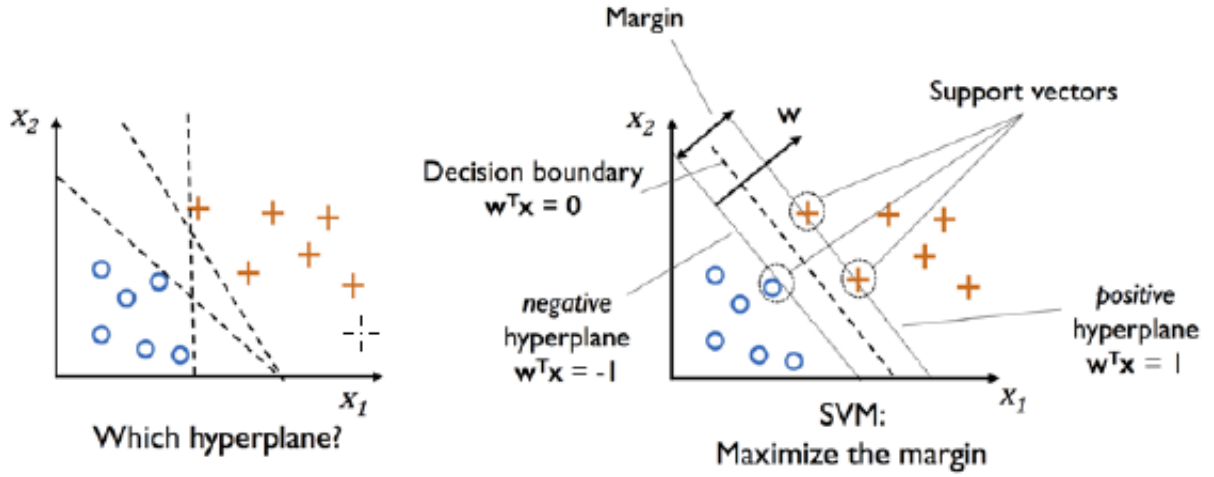


Figura 2.13: Support Vector Machine. Fonte: (RASCHKA e MIRJALILI, 2017).

a *overfitting* (quando um modelo apresenta um bom resultado com os dados de treino e não generaliza suficientemente quando alimentado com dados de teste, não vistos antes) (RASCHKA e MIRJALILI, 2017; WIDODO e YANG, 2007).

Dado que $\mathbf{x}_i$ ($i = 1, 2, \dots, n$), sendo n o número de amostras. Assume-se que as amostras possuem uma classe positiva ($y_i = 1$) e uma classe negativa ($y_i = -1$). Considerando linearidade nos dados, pode-se descrever o hiperplano $f(\mathbf{x}) = 0$ como na Equação 2.1, onde $\mathbf{w}$ é um vetor de dimensão N e b é um escalar e ambos definem a posição do hiperplano.

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = \sum_{i=1}^n w_i x_i + b = 0 \quad (2.1)$$

Um hiperplano distinto deve satisfazer às condições $f(x_i) = 1$ se $y_i = 1$, $f(x_i) = -1$ se $y_i = -1$; que pode ser representado na forma completa:

$$y_i f(\mathbf{x}) = y_i (\mathbf{w}^T \mathbf{x} + b) \geq 1, i = 1, 2, \dots, n \quad (2.2)$$

Então, o hiperplano ótimo que separa as duas classes pode ser obtido resolvendo o problema de otimização apresentado na Equação 2.3 sujeito às condições na Equação 2.4, onde ξ_i é a medida da distância entre a margem e as amostras no lado incorreto, C é o parâmetro de penalização do erro ($C > 0$), e ϕ é a função de núcleo.

$$\min_{\mathbf{w}, b, \xi} \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i \quad (2.3)$$

$$\begin{cases} y_i (\mathbf{w}^T \phi(x_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \quad (2.4)$$

O emprego de SVM em problemas de classificação não-lineares requer a utilização da função de núcleo ϕ , que basicamente mapeia os dados em um espaço de dimensão maior onde

a classificação linear é possível. As funções de núcleo comumente utilizadas estão listadas na Equação 2.5.

Linear	$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \mathbf{x}_j$	(2.5)
Polinomial de grau g	$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i \mathbf{x}_j + r)^g$	
<i>Radial Basis Function</i>	$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \ \mathbf{x}_i - \mathbf{x}_j\ ^2), \gamma > 0$	
Sigmoide	$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(-\gamma \mathbf{x}_i \mathbf{x}_j + r), \gamma > 0$	

Em uma aplicação de monitoramento de equipamentos e previsão de falhas, o modelo é treinado para identificar um padrão nas variáveis dos sensores e classificá-los conforme a ocorrência de falhas no conjunto de dados. Equipamentos industriais possuem muitos sensores sendo que alguns podem não ser significativos ou conter informação suficiente sobre a condição do equipamento, portanto, é necessário realizar uma análise das variáveis para aumentar a acurácia do classificador baseado em SVM (WIDODO e YANG, 2007).

2.6. Transportador de Correia

Os transportadores de correia são equipamentos utilizados quando se pretende transportar grânéis sólidos continuamente, capazes de atingir altas capacidades de transporte com elevada disponibilidade, excelente custo de operação e manutenção e menos danosos ao meio ambiente, quando comparado a outros meios de transporte (GELAIS *et al.*, 2016).

A norma brasileira NBR 6177 apresenta o seguinte conceito de transportador de correia: “Arranjo de componentes mecânicos, elétricos e estruturas metálicas, consistindo em um dispositivo horizontal ou inclinado (ascendente ou descendente) ou em curvas (côncavas ou convexas) ou ainda, uma combinação de quaisquer destes perfis, destinado à movimentação ou transporte de materiais a granel através de uma correia contínua com movimento reversível ou não que se desloca sobre os tambores, roletes ou mesa de deslizamento, segundo uma trajetória pré-determinada pelas condições de projeto, possuindo partes ou regiões características de carregamento e descarga”.

Esse sistema eletromecânico pode ser composto pelos elementos apresentados a seguir (FRUCHTBAUM, 2013; MARCONDES, 2011).

Correia

A correia é o elemento responsável por acomodar o material e transportá-lo, cujas especificações consideram as características do material (granulometria, densidade, abrasividade, temperatura, etc.), a capacidade pretendida, tipos de rolete e tensão.

Rolete

O rolete é um conjunto formado pelos rolos e sua estrutura e que tem a função de suportar a correia. Os rolos são cilindros de aço que giram em torno de seu próprio eixo, fazendo com que a correia se desloque sobre eles.

Os roletes podem ser classificados conforme sua aplicação no sistema de transporte: carga, impacto, retorno, transição e auto-alinhantes.

- **Rolete de Carga:** conjunto que suporta a carga imposta pelo peso da correia e do material transportado. Nas plantas da Vale no Brasil, os roletes de carga são tipicamente instalados de forma inclinada para diminuir a perda de material fino de minério (quando aplicável) e aumentar a carga transportada pelo TCLD. Os roletes na configuração inclinada normalmente contêm três rolos;
- **Rolete de Impacto:** conjunto instalado na seção em que o material é descarregado sobre a correia, com rolos revestidos com material capaz de absorver o impacto e proteger a correia contra danos;
- **Rolete de Retorno:** conjunto instalado na parte inferior da correia e suporta a carga da mesma durante o retorno;
- **Rolete de Transição:** conjunto instalado próximo aos tambores e possui regulagem de ângulo, utilizado para diminuir a curvatura da correia à medida que se aproxima dos tambores;
- **Rolete Auto-alinhante:** conjunto responsável por manter o alinhamento da correia em qualquer um dos lados (carga ou retorno).

Os rolos que compõem os roletes tipicamente possuem dois rolamentos, sendo importante considerar o tempo de vida útil dos rolamentos dos rolos na especificação do sistema de transporte (LODEWIJKS, 2004).

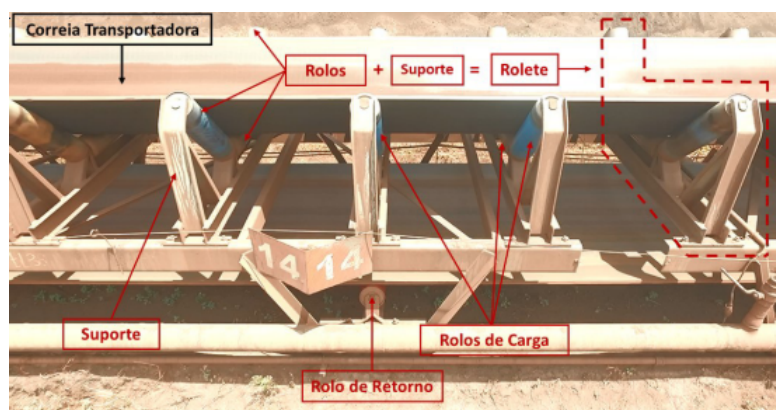


Figura 2.14: Roletes de carga suportando uma correia. Fonte: (NASCIMENTO, 2018)

Tambor

Os tambores são elementos essenciais para o funcionamento dos transportadores de correia porque transmitem o movimento imposto pelo sistema de acionamento, do eixo de saída dos redutores de velocidade para a correia (MARCONDES, 2011).

Segundo Marcondes (2011) os tambores podem ser classificados como de descarga, desvio, acionamento, esticamento e retorno.

2.7. Trabalhos Relacionados

2.7.1. Soluções para Detecção de Anomalias

Widodo e Yang (2007) apresenta uma extensa revisão da literatura de 1996 a 2006 sobre aplicações de modelos baseados em SVM especificamente para o monitoramento da condição e detecção de falhas nos mais diversos equipamentos: rolamentos, motores de indução, bombas, compressores, entre outros. Nota-se que o SVM normalmente não é aplicado puramente nos *datasets* sendo acompanhado de técnicas para redução de dimensões (PCA) e seleção de variáveis (árvore de decisões ou algoritmo genético).

Chen *et al.* (2011) apresentam em seu artigo um caso de aplicação do modelo baseado em SVM para prever falhas em uma turbina de uma usina termoeletrica tendo como fonte de dados 29 sensores de temperatura, pressão e velocidade da turbina. Os autores selecionaram SVM pelos seguintes fatores, além da maximização das margens: uso de núcleos (permite que se trabalhe em espaços não-lineares) e ausência de mínimos locais. No mesmo artigo, Chen *et al.* (2011) demonstram que o modelo de aprendizado baseado em SVM obteve resultado superior a outros dois modelos de referência (*Linear Discriminant Analysis* e *Back-propagation Neural Networks*) independente da aplicação das técnicas de redução de dimensões, análise de correlação e árvore de decisões.

Li *et al.* (2013) demonstram a aplicação de WPD e SVM para detecção de anomalias em roletes de transportadores de correia em uma mina de carvão com a instalação de um acelerômetro no conjunto para coletar o sinal de vibração. O sinal é decomposto com WPD e a energia de cada frequência alimenta um modelo SVM (com núcleo *Radial Basis Function*) para identificar as falhas em roletes. O modelo treinado obteve uma acurácia de 91,67% e tempo de computação perto de 0,5 segundos, o que pode incorrer em atraso na detecção de falhas, mas dentro de parâmetros aceitáveis considerando a velocidade média de operação do transportador de correia nos testes.

2.7.2. Soluções para Digital Twin

Tao *et al.* (2018) propõem uma aplicação de DT para um transformador baseado em uma plataforma de DT segmentada em serviços dedicados a funções distintas e especializadas,

representada na Figura 2.15. Os serviços do DT do transformador seriam capazes de calcular um indicador operacional do equipamento em geral, consumo atual e projetado de energia do transformador, previsão da qualidade da energia convertida. Estes indicadores, no entanto, somente poderiam ser calculados mediante a instalação de diversos sensores adicionais que normalmente não estão disponíveis em equipamentos mais antigos.

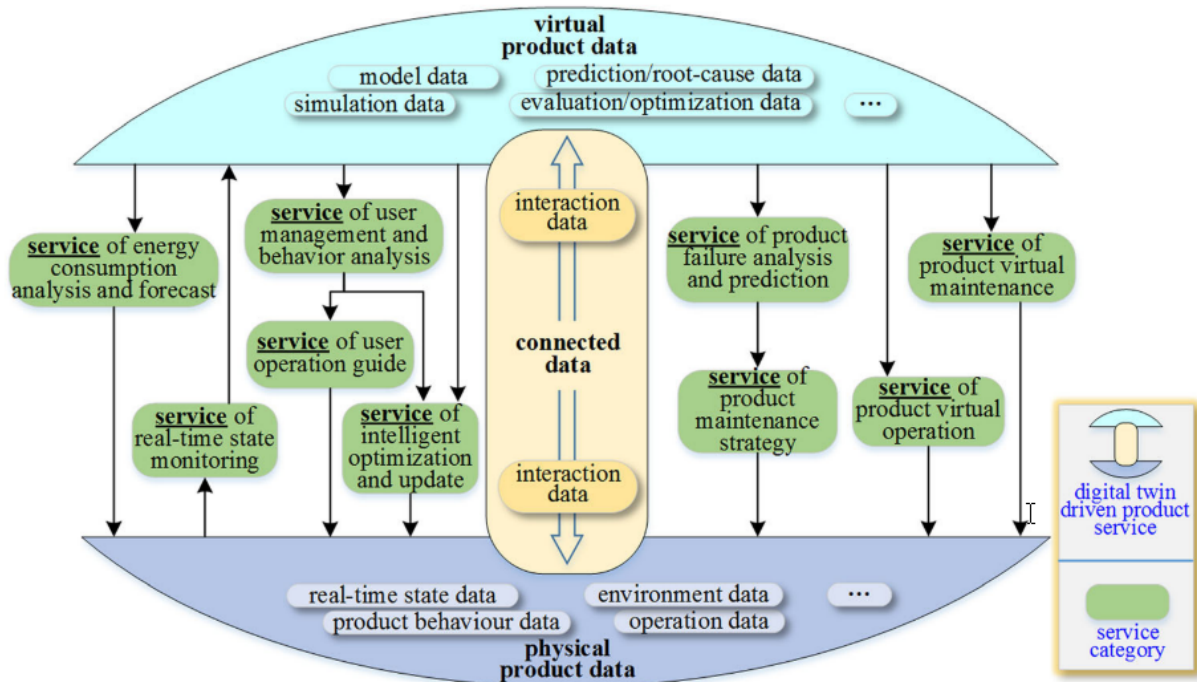


Figura 2.15: Proposta de serviços para uma plataforma de DT. Fonte: (TAO *et al.*, 2018)

Recentemente, Luo *et al.* (2019) exemplificaram a aplicação do DT em um equipamento e relataram a importância do algoritmo de previsão de falhas para se atingir os objetivos de precisão na simulação e auto-avaliação do DT. Em sua experiência foi implementada uma rede neuro-fuzzy, mas não foram avaliados outros algoritmos para endereçar os principais desafios de implementação do DT.

Xu *et al.* (2019) propõem a utilização de um DT existente gerando grande volume de dados para: contornar os problemas de *datasets* de tamanho limitado, com distribuições diferentes e desbalanceados; alimentar um algoritmo de *Deep Transfer Learning* com objetivo de aumentar significativamente a acurácia da previsão de falhas.

Cerrone *et al.* (2014) discorrem sobre o processo de simulação de falhas em peças de prova retiradas de uma determinada frota de veículos e consideraram a aplicação do conceito de DT para endereçar o problema de ambiguidade no diagnóstico ou previsão do caminho da falha nas peças, ou seja, a “personalização” de cada peça e elaboração de um modelo computacional para previsão das falhas.

O DT pode ser uma parte ou todo o CPS porque possui todas as informações e conhecimento do sistema físico, como mostra Schroeder *et al.* (2017) através de seu estudo de caso ao sustentar sua metodologia baseada no padrão *AutomationML* para modelar o DT em alto

nível e facilitar a integração com outros sistemas. O emprego de *AutomationML* é de interesse relevante, porque este padrão permite armazenar informações de engenharia segundo o paradigma da orientação a objetos e, conseqüentemente, modelar componentes físicos e lógicos como objetos englobando diferentes aspectos.

Prinsloo (2019) descreve brevemente um estudo de caso de DT na indústria de mineração, combinando dois softwares de simulação com dados coletados em tempo real de sensores por uma plataforma de IoT, em que simula a descarga de carvão por carros em um alimentador de sapatas da britagem para prever o entupimento do sistema e gerar alertas.

O Grupo Voith desenvolveu uma solução de DT em nuvem para transportadores de correia que recebe dados da instrumentação instalada no equipamento e simula a operação a partir nas propriedades físicas do sistema de transporte (por exemplo: potência dos acionamentos, espessura, comprimento, largura, etc.). A empresa informa que a solução calcula um indicador de eficiência energética para *benchmarking* além de prever falhas de operação, porém não foram encontradas evidências de aplicações da solução. (MINING.COM, 2018; VOITH, 2019; ZIEGLER, 2019)

Pela pesquisa na literatura, nota-se que DT é um conceito relativamente novo e pouco aplicado na indústria de base e transformação e que o interesse por este tipo de solução começou a crescer há três anos como pode ser observado na Figura 2.16. Verificou-se que empresas envolvidas com TO cada vez mais procuram evoluir suas soluções (sistemas de controle, gestão de ativos, aplicativos, etc.) baseadas no conceito de DT (ABB, 2019; GE DIGITAL, 2019; SIEMENS, 2019; VOITH, 2019), porém os casos de construções de DT são pouco divulgadas dado o caráter estratégico e vantagem competitiva.

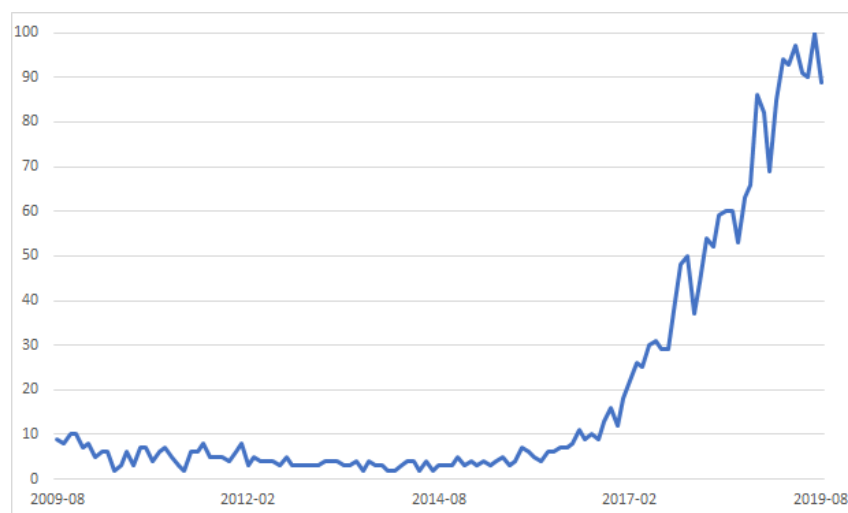


Figura 2.16: Tendência de pesquisa pelo termo “Digital Twin” na Internet nos últimos 10 anos.
Fonte: Google Trends

3. Desenvolvimento

Este capítulo apresenta as experiências desenvolvidas no decorrer do projeto com objetivo de validar na prática os requisitos estudados para um DT no sentido de elaborar a proposta de arquitetura integrada. Inicialmente, descreve-se a metodologia considerada para o desenvolvimento do projeto, seguido pela apresentação do modelo CRISP-DM, quando são detalhadas as etapas aplicadas ao desenvolvimento do modelo de detecção de anomalias do transportador de correia. Em seguida, são apresentadas as aplicações de IIoT e integração de dados do chão de fábrica, importantes fontes de dados para uma plataforma de DT.

3.1. Metodologia

Uma proposta de arquitetura para construção de DT possui desafios onde se destacam a comunicação entre diversas fontes de dados e a sincronia das informações dos componentes do DTI. O desenvolvimento da plataforma a partir de metodologias tradicionais, baseado em atividades sequencias, pode não ser o mais produtivo uma vez que as partes que integram a plataforma têm tecnologias distintas e seguem linhas evolutivas independentes. Dessa forma, propõe-se que o desenvolvimento seja calcado em metodologia ágil, com uma abordagem de melhoria contínua segmentada em cada parte, com foco na solução integrada.

O presente trabalho focou no conhecimento do processo de desenvolvimento de um modelo de detecção de anomalias para validação dos requisitos estudados e identificação de restrições práticas a fim de se propor uma arquitetura integrada para a plataforma do DT. Importante mencionar que o presente trabalho contribui em parte para um projeto maior de APM da empresa. Portanto, abaixo apresentam as principais atividades para atingir os objetivos propostos:

- Caracterização do problema, relevância e consequências para a indústria da mineração, por meio de entrevistas com especialistas e revisão de bibliografia especializada;
- Levantamento do estado da arte do DT na indústria de manufatura;
- Desenvolvimento de um algoritmo para detecção de anomalias de uma correia transportadora a partir de séries temporais multivariável (MTS);
- Análise de experiências com aplicações internas na empresa que contribuem para o desenvolvimento de uma solução de DT;
- Revisão de requisitos críticos para uma solução de DT, incluindo interfaces, serviços e integrações;
- Proposição de uma arquitetura que permita a construção e evolução de uma plataforma de DT.

3.2. Modelo para Detecção de Anomalias

Conforme Chapman *et al.* (2000), as fases de um projeto de mineração de dados segundo o modelo CRISP-DM são apresentadas na Figura 3.1 e descritas brevemente a seguir.

- **Entendimento do Processo** (*Business Understanding*): busca-se a compreensão adequada do problema que necessita ser resolvido;
- **Compreensão dos Dados** (*Data Understanding*): inicia com a coleta inicial dos dados e análises para inspecionar, organizar e descrever os dados disponíveis, é importante avaliar quais dados podem ser relevantes para decifrar o problema ou formar hipóteses sobre informações escondidas;
- **Preparação dos Dados** (*Data Preparation*): considera todas as etapas de construção do conjunto de dados a partir dos dados iniciais, podendo ser necessário executar repetidamente até que se obtenha um conjunto de dados pronto para modelagem;
- **Modelagem** (*Modeling*): seleção das técnicas de modelagem, incluindo parametrização, calibração e aplicação ao conjunto de dados, sendo que pode ser necessário retornar à preparação dos dados para aplicar correções;
- **Avaliação** (*Evaluation*): avalia os modelos construídos na fase anterior conforme métricas estabelecidas e os passos para sua construção são revisados para certificar de que os modelos atingem o objetivo inicial;
- **Publicação** (*Deployment*): aplica o(s) modelo(s) desenvolvido(s) no dia a dia da organização de uma forma que os usuários possam utilizar as informações geradas.

3.2.1. Entendimento do Processo

Conforme apresentado na Seção 1.3, as correias são equipamentos responsáveis pelo transporte de minério de ferros por longas distâncias, desde a mina até os processos de tratamento e também entre os diversos processos em uma usina. Portanto, pode-se considerar que tais equipamentos são importantes para a operação de uma usina de beneficiamento e a sua utilização deve ser maximizada e, por dedução, as paradas por falhas devem ser minimizadas.

Por exemplo, uma correia transportadora com capacidade de 16.000 toneladas por hora operando com utilização de 72% em um período de 30 dias transporta 345.600 toneladas de minério de ferro. Um aumento de 1% em sua utilização significa transportar mais 4.800 toneladas de minério no mesmo período, o equivalente a US\$ 321.360.

Então, foram analisados os dados das séries históricas dos sensores de uma correia transportadora selecionada dentro uma usina de beneficiamento de minério de ferro para construção

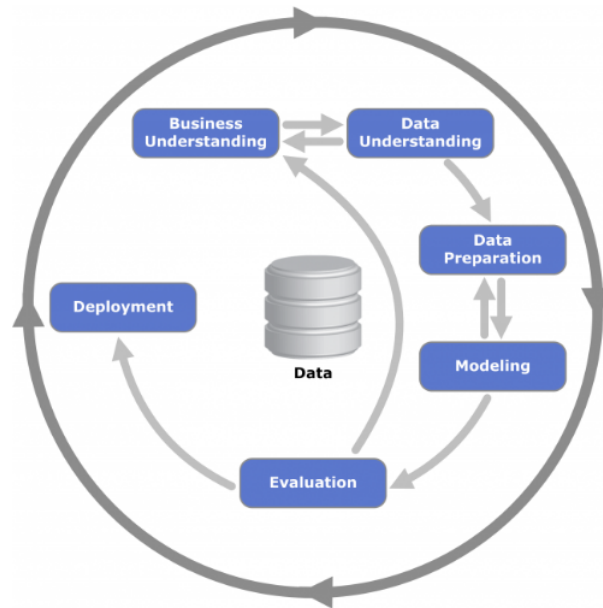


Figura 3.1: Fases do modelo de referência CRISP-DM. Fonte: (CHAPMAN *et al.*, 2000)

de modelos preditores de falhas, com *Machine Learning*, para alertar os operadores e supervisores de processo quanto à operação do equipamento e auxiliar na revisão dos *set points* operacionais e programação de manutenção.

A correia transportadora selecionada é parte do processo de transporte do minério de alta granulometria dos britadores semi-móveis nas frentes de lavra para o primeiro processo de tratamento na usina de beneficiamento. O equipamento possui de 7 sistemas de acionamento elétrico para tracionar a correia e transportar minério de ferro à taxa nominal de 19.200 toneladas por hora. Considerado crítico para a usina, o equipamento possui operação e manutenção complexas e, portanto, apresenta maior número de paradas não programadas, sendo um bom candidato para análise.

3.2.2. Compreensão dos Dados

A partir de entrevistas com analistas operacionais e engenheiros de automação obteve-se maior conhecimento sobre a operação e manutenção do equipamento. Foram identificadas 192 variáveis associadas ao TCLD no sistema de informações de processos (PIMS) da usina. As variáveis foram amostradas a cada minuto, em um período de nove meses (Abril/2018 a Dezembro/2018) para capturar dinâmicas operacionais do equipamento e obter um resultado mais realista. Dentre as variáveis identificadas, encontram-se:

- Sensores de corrente, tensão, potência, velocidade, torque, número de partidas e horímetro dos sistemas de acionamento e do sistema como um todo;
- Palavras de estado de cada sistema de acionamento e do equipamento como um todo;
- Estados calculados de cada sistema de acionamento e do equipamento como um todo;

- Nível dos chutes na entrada e saída da correia;
- Limite de segurança;
- Sensor de falha no acionamento do freio;
- Sensores de comunicação e *heart beat* com o sistema de controle;
- Sensores de frenagem.

O estado da correia transportadora, a variável alvo, está representado na variável S11D_2011KS_TR07_STS_C que é calculado a partir da palavra de estado do equipamento (registrada na variável S11D_2011KS_TR07_STS_AUX). A correia transportadora pode assumir os estados FUNCIONANDO, FUNCIONANDO COM ALARME, DEFEITO e PARADO COM ALARME.

A Figura 3.2 apresenta o diagrama de Pareto dos estados da correia transportadora no período considerado para análise. Dado que a amostragem é de um minuto, pode-se considerar que a correia funcionou por 75,2% e ficou em defeito por 12,0% do tempo. A operação da correia funcionando com alarme na maior parte do tempo é esperada porque que a usina de beneficiamento está em processo de *ramp up* da produção.

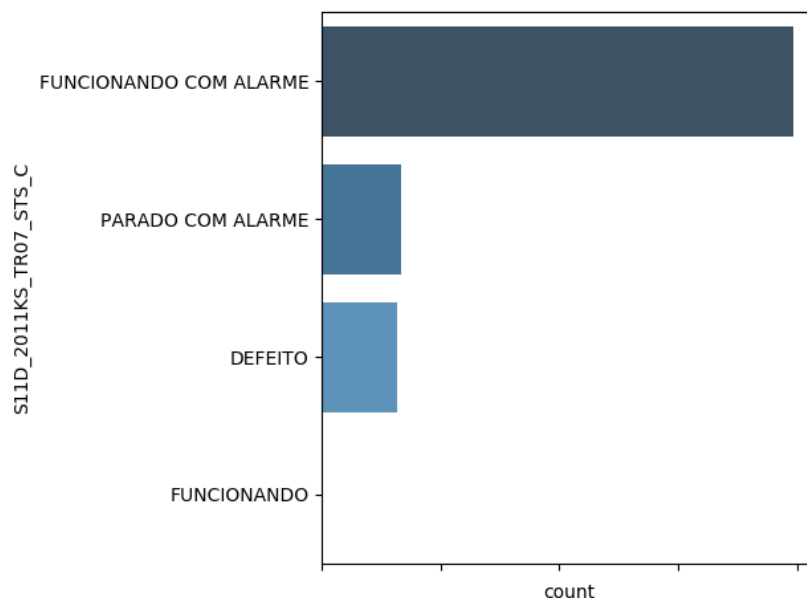


Figura 3.2: Distribuição dos estados da correia transportadora. Fonte: Autor.

3.2.3. Preparação dos Dados

Em uma aplicação detecção de anomalias o modelo é treinado para identificar um padrão nas variáveis e classificá-los conforme a ocorrência de falhas no *dataset*. Equipamentos industriais possuem muitos sensores, sendo que alguns podem não ser significativos ou con-

ter informação suficiente sobre a condição do equipamento, portanto, é necessário analisar as variáveis para aumentar a acurácia do classificador (WIDODO e YANG, 2007).

A análise das variáveis mostrou que existem informações duplicadas ou redundantes no banco de dados do PIMS e que, portanto, podem ser removidas sem prejuízo para o processo de modelagem. Abaixo alguns exemplos das variáveis do PIMS removidas nesta fase:

- S11D_2011KS_TR_07M1_ET e S11D_2011KS_TR07M1_IT captam a corrente do sistema de acionamento 1;
- S11D_2011KS_TR07M6_ST, S11D_2011KS_TR07M06_ActualSpeed e S11D_2011KS_TR07M06_CM_ActualSpeed captam a velocidade do sistema de acionamento 6;
- S11D_2011KS_TR07M1_STS_C é o estado do sistema de acionamento 1 calculado a partir da palavra de estado S11D_2011KS_TR07M1_STS_AUX.

Sobre o último exemplo listado acima, é importante descrever que pode-se considerar a palavra de estado (e todos seus bits) redundantes com o estado calculado porque a informação do estado é suficiente para o propósito dos modelos que se propõe construir: antecipar a ocorrência de falhas. Em um trabalho futuro em direção ao *Digital Twin*, sugere-se considerar as palavras de estado para prever qual a falha que ocorrerá no equipamento e/ou ainda indicar ações que o operador deve tomar para evitar a falha (neste caso, seria necessário incluir as variáveis de comando do operador que não estão disponíveis nos sistemas da usina).

Portanto, devido ao tamanho do *dataset*, foi necessário selecionar as variáveis mais significativas para treinamento do modelo, conforme recomendações apresentadas em Dash e Liu (1997), Guyon e Elisseeff (2003) e Saeys *et al.* (2007). Dessa forma, o modelo SVM foi treinado com os mesmos parâmetros, para comparação de resultados, com todas as variáveis, uma seleção de 25 variáveis utilizando o método estatístico Chi-Quadrado (HUAN LIU e SETIONO, 2002) e redução de dimensões baseada em análise de componentes principais (PCA) como demonstrado por Sun *et al.* (2007) e Yang e Shahabi (2004). As variáveis foram normalizadas a partir de sua média e desvio padrão com objetivo de otimizar a execução do algoritmo de treinamento.

O *dataset* para análise do modelo de aprendizado de máquina foi definido com 93 variáveis e 385.869 amostras, ou seja, 35.885.817 dados, divididos em 70% para treino do modelo e 30% para testes. É importante explorar relações entre as variáveis para direcionar a seleção de uma função *kernel* adequada sendo que, neste caso, identificou-se uma não-linearidade, o que determina, inicialmente, a função *radial basis function* com complexidade computacional $O(N^2n)$, onde N é o número de amostras e n é o número de variáveis.

3.3. Experiências com Aplicações Industriais

Na Vale encontram-se diversas experiências com aplicações industriais de tecnologias recentes gerando dados e informações relevantes e que precisam ser integradas ao DT para aumentar a acurácia e precisão dos modelos dos equipamentos e processos.

3.3.1. Internet das Coisas Industriais

No que tange correias transportadoras, a utilização de aplicações Internet das Coisas Industriais na indústria de mineração é estudada no âmbito da manutenção com a finalidade de aumentar a utilização dos equipamentos. Portanto, os dados gerados são de relativa importância para um algoritmo de detecção de anomalias e consequentemente para o funcionamento de um DT. Duas aplicações destacam-se entre as iniciativas avaliadas da empresa: roletes inteligentes (ou *smart idler/roller*); e sensores de vibração tri-axial com sensores de temperatura.

Lodewijks *et al.* (2007) propõem o conceito de um *smart idler*, uma solução de IoT com eletrônica embarcada para leitura e transmissão da temperatura dos roletes, no entanto o protótipo descrito possui um custo elevado como substituto para os roletes tradicionais. Pang e Lodewijks (2011) apresentam uma evolução deste conceito para aumentar a confiabilidade, reduzir a necessidade de *gateways* e que não depende de bateria. A instalação de *smart idlers* foi avaliada pela empresa e desconsiderada devido ao custo de instalação e manutenção considerando o número de roletes existentes em uma planta.

Li *et al.* (2013) apresentam uma solução de acelerômetros com sensores de temperatura para analisar a vibração de correias transportadoras e identificar os modos de falha do equipamento que apresenta menor custo que um *smart idler*. Partindo deste conceito, a companhia desenvolve uma solução, junto a uma empresa nacional, de um dispositivo com sensores de vibração tri-axial e temperatura com custo unitário cerca de 90% menor do que os fornecedores típicos para a indústria.

O dispositivo possui duas versões (baixa e alta frequência) para aplicações diferentes e com custo diferenciado, as leituras são transmitidas através de *bluetooth low energy* (BLE) de forma manual para um *smartphone* ou *tablet* utilizando um aplicativo desenvolvido pela empresa ou para um *gateway* que pode ser posicionado até 40 m do dispositivo. Nas duas situações os dados são transmitidos para a nuvem do fornecedor através de GPRS, onde uma aplicação analisa o espectro de frequência para identificar os diferentes modos de falha conforme uma assinatura padrão definida pelos especialistas do equipamento industrial. O resultado da análise é enviado para um *data lake* de manutenção existente na nuvem privada da companhia.

O desenvolvimento inicial da solução de IIoT não considerou a integração com outros sistemas da companhia, mas o *gateway* desenvolvido possui interface para redes cabeadas ou LoRa para permitir a integração dos dados com o PIMS para padronizar as integrações no ambiente corporativo e disponibilizar os dados para análises com as informações de operação e

manutenção dos equipamentos. A conectividade entre o *gateway* e o PIMS será implementada sobre uma infraestrutura de rede segmentada das redes de controle/supervisão e corporativa, com medidas de segurança da informação para mitigar o risco de invasão à infraestrutura crítica e conferir maior segurança operacional.

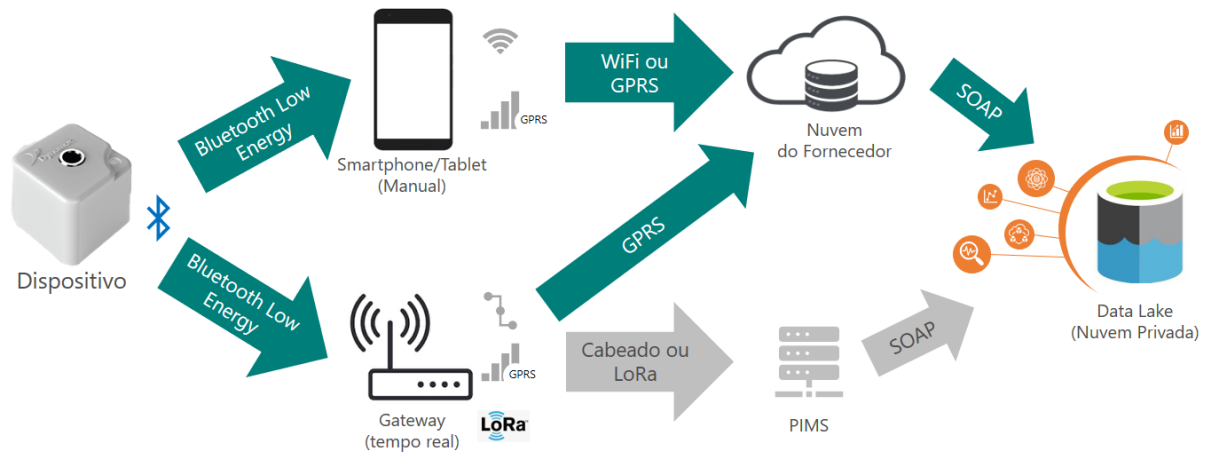


Figura 3.3: Fluxo de dados da solução para leitura de vibração de equipamentos. Fonte: Documentação Interna da Vale

Moura *et al.* (2017) propuseram o *framework* apresentado na Figura 3.4 para implementação de soluções de IIoT e que norteia os desenvolvimento de aplicações da empresa. O *framework* não determina que todas as camadas sejam preenchidas e serve como guia para suportar a definição, padronização e aplicação de tecnologias em cada caso de uso. Na aplicação de análise de vibração, os sensores (*Things*) fornecem dados para o *gateway* e que os conecta (*Data Ingest*) à nuvem da empresa onde estão implementadas as funções de *Analytics Repository*, *Advanced Analytics* e *Analytics Presentation*. Com a inclusão do PIMS na solução, pretende-se atender aos requisitos de *Information Hub* e *Analytics Repository* utilizando funcionalidades consolidadas nos ambientes industrial e corporativo, conforme apresentado na Subseção 2.2.1.

Nascimento *et al.* (2017) propõem uma arquitetura de sistema integrado de inspeção de correias transportadoras com o uso de VANT e processamento de imagem, portanto uma proposta de arquitetura para DT deve considerar a integração dos dados gerados por este sistema.

A integração dos dados das soluções de IIoT com o PIMS foi padronizado na companhia para endereçar os temas a seguir:

1. Contextualização das informações conforme o modelo de ativos e hierarquia de locais da companhia para apresentação aos usuários finais e integração com sistemas;
2. Monitoramento remoto e em tempo real da operação dos equipamentos em questão;
3. Atualização dos indicadores de desempenho do equipamento existente no PIMS;
4. Análise histórica da correlação entre os novos dados e dados provenientes dos sistemas de controle e supervisão;

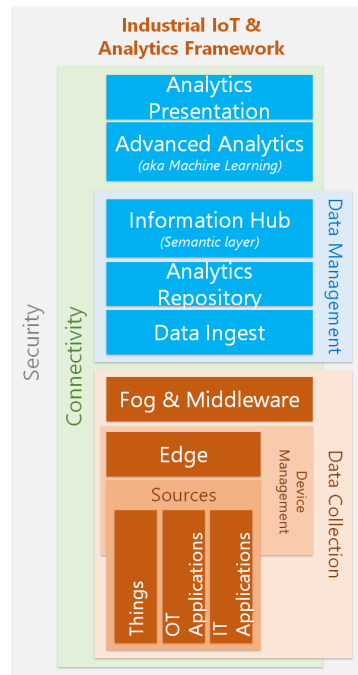


Figura 3.4: *Framework de IIoT Analytics*. Fonte: adaptado de Moura *et al.* (2017)

5. Utilização da infraestrutura de integração existente com sistemas e aplicações corporativas como CMMS, MES, *data lake* e *Advanced Analytics*.

3.3.2. Integração dos Dados de Processos e Equipamentos

Todas as plantas da Vale possuem um banco de dados do PIMS segundo o modelo de arquitetura apresentado na Figura 2.7. Estes sistemas são muito utilizados pelas equipes de operação e manutenção das plantas, em suas atividades cotidianas, para garantir a integração das ações e melhorar a tomada de decisões através de monitoramento de indicadores chave de processo calculados em tempo real.

Para integração ao barramento de integração, os PIMS disponibilizam diferentes formas de consultas aos dados (históricos e tempo real) em interfaces *web services* e API REST. Em se tratando do software Aspen InfoPlus.21, fornecido pela Aspen Technologies, Inc., o barramento de integração disponibiliza uma interface pública síncrona (*request-response*) em *web service* onde os sistemas integrados enviam uma solicitação de execução de uma *stored procedure* desenvolvida sob medida no PIMS que provê as informações necessárias. Este método permite maior grau de especialização ao custo de maior complexidade da integração no lado do PIMS, mas permite que sejam executados lógicas, funções e cálculos disponíveis somente nos sub-sistemas do banco de dados histórico.

Os sistemas MES, CMMS, LIMS e APM integram-se ao PIMS por meio do barramento de integração (Figura 3.5), sendo que todos utilizam principalmente a interface pública síncrona (cada qual com sua *stored procedure* especializada). As mensagens seguem a linguagem XML sendo que há um dicionário de dados para cada *stored procedure*. O barramento de integração

direciona as mensagens para os servidores corretos do PIMS, uma vez que existem mais de 30 sistemas distintos na empresa, e faz a orquestração das mensagens para traduzir do formato XML entregue pelo PIMS para o formato esperado pelo sistema solicitante. A volumetria de dados destas integrações é pequena e o menor período de consulta encontrado é de 5 minutos.

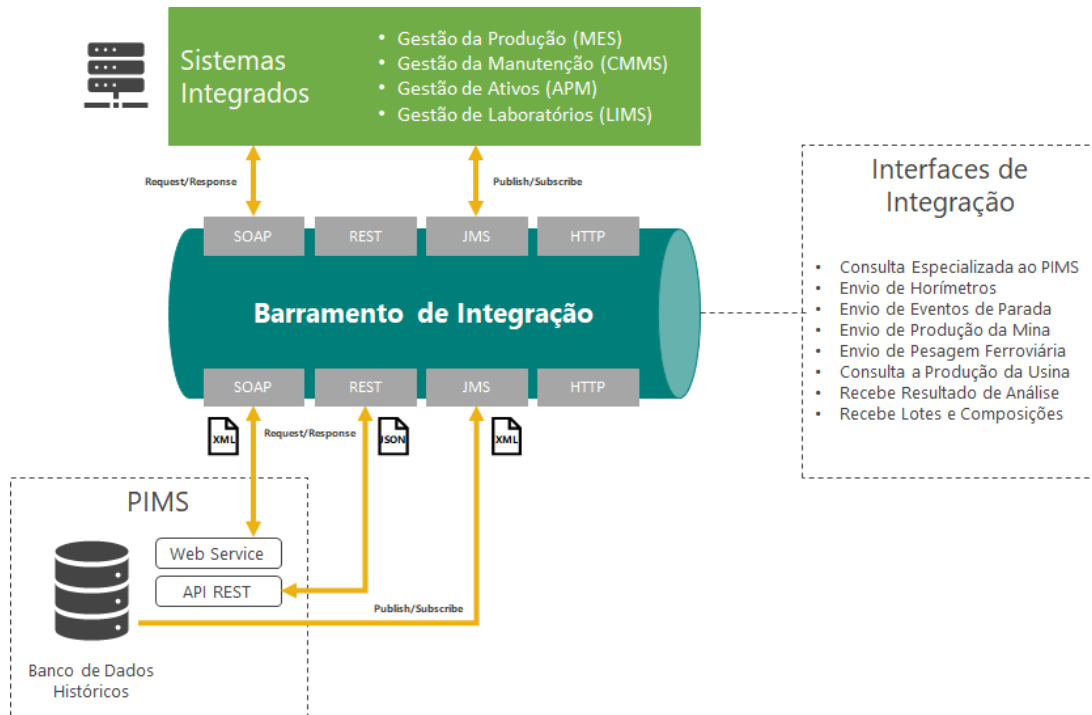


Figura 3.5: Integrações do PIMS que provêm dados de processos para sistemas corporativos da Vale. Fonte: Autor

Nas integrações de maior volumetria de dados, o barramento de integração implementa o JMS para fornecer dados em modo *publish-subscribe*. No PIMS são desenvolvidos algoritmos com *triggers* periódicos de 1 a 30 minutos, conforme a demanda pelas informações, que executam consultas próprias na base de dados histórica pelas informações, formata o corpo XML da mensagem conforme dicionário da integração e publica nas filas de mensagens.

As aplicações de BDA que utilizam dados de processos e equipamentos consultam um *data lake* que pode ser alimentado pelo PIMS de duas formas distintas: *backfilling* ou *streaming*. A Figura 3.6 apresenta a arquitetura do *backfilling* de dados do PIMS para o *data lake*, iniciado por uma solução de ETL sob demanda (quando o período de consulta é grande) ou uma vez ao dia (dependendo do volume de dados). A consulta ao banco de dados históricos é executada utilizando o padrão Open Database Connectivity (ODBC) em busca de dados agregados ou interpolados (conforme demanda para a análise); o resultado da consulta apresenta os dados em um formato adequado para envio ao *data lake* e remove a necessidade para execução de transformação nos dados. A solução de *advanced analytics*, hospedada em infraestrutura na nuvem, consulta os dados de processo e equipamentos no *data lake* e escreve os resultados utilizando o padrão ODBC.

O *streaming* dos dados de processo e equipamentos é possível somente a partir de um

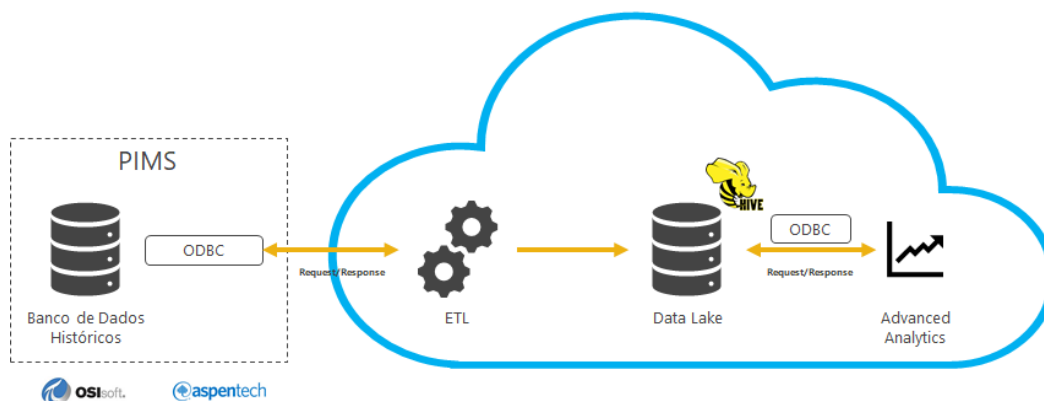


Figura 3.6: Integração do PIMS para *backfilling* do *data lake* e *advanced analytics*. Fonte: adaptado de documentação da Vale

software de PIMS utilizado pela Vale, porque apenas um fabricante da plataforma fornece uma aplicação nativa com tal capacidade. Foram realizados testes de uma solução desenvolvida sob medida para o software que não possui aplicação nativa para *streaming*, tal solução não foi implementada porque comprometeu o desempenho de serviços críticos do *pims* e impediu o funcionamento normal do sistema.

As tarefas de *streaming* de dados são configuradas em uma aplicação web. Nela os usuários podem visualizar os ativos (processos ou equipamentos) configurados no banco de dados do PIMS, selecionar atributos para a consulta, o método (interpolados ou agregados) e período de consulta. A partir da consulta registrada configura-se o envio das mensagens para um *endpoint* do Azure IoT Hub, este, por sua vez, foi configurado para escrever os dados recebidos na mensagem no *data Lake*. (OSISOFT, 2018)

A arquitetura implementada para a integração de *streaming* de dados do PIMS para o *data lake* está representada na Figura 3.7. A aplicação nativa permite ainda a configuração de vários nós executantes de tarefas para paralelizar o processamento como forma de escalar conforme a demanda por envio de dados em tempo real para o *data lake* na nuvem.

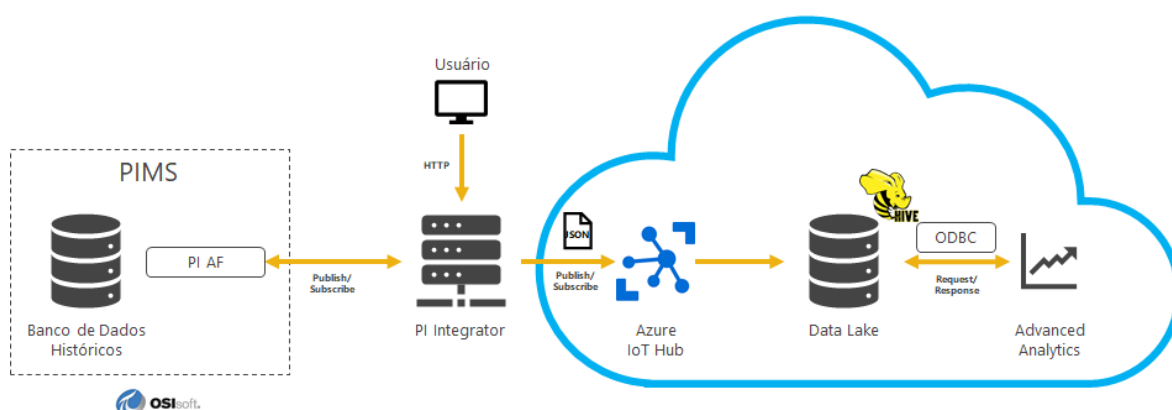


Figura 3.7: Integração do PIMS para *streaming* de dados para o *data lake*. Fonte: adaptado de documentação da Vale

4. Resultados e Discussão

As seções deste capítulo discorrem sobre os resultados encontrados com o desenvolvimento do modelo SVM e com a integração do PIMS com o *Data Lake* para então, a partir da análise desses resultados, desenvolver uma proposta de arquitetura para implementação de uma plataforma de DT, integrada aos sistemas de chão de fábrica e corporativos, baseada em computação na nuvem.

4.1. Avaliação do Modelo SVM

Para treinamento do modelo SVM, considerou-se como saída a variável de estado geral do equipamento com atraso de sete dias para avaliar a capacidade do modelo de prever um defeito com dada antecedência.

A Tabela 4.1 apresenta a acurácia obtida a partir da execução do mesmo algoritmo de treinamento com e sem balanceamento das classes e com diferentes números de variáveis de entrada (o treino do modelo balanceado com 93 variáveis não foi executado devido ao longo tempo esperado para sua execução). A variação no tempo de execução do treinamento do modelo deveu-se essencialmente ao número de amostras e balanceamento das classes, evidenciado mais a frente nesta seção.

Tabela 4.1: Acurácia atingida pelo modelo SVM com os dados originais. Fonte: Autor

Variáveis	Tempo	Acurácia	Precisão	Recall	F1
93	6,24 h	31,6%	0,154	0,684	0,251
25 (χ^2)	1,92 h	33,1%	0,162	0,719	0,265
5 (PCA)	2,20 h	40,8%	0,165	0,622	0,260

A matriz de confusão, representada na Figura 4.1, complementa a análise do resultado dos modelos treinados porque exemplifica o impacto do desbalanço de classes nas predições executadas considerando o mesmo peso para o erro de classificação das duas classes no problema, conforme descrito por He e Garcia (2009). O modelo treinado com o *dataset* com 93 variáveis registrou um número maior de falso negativos, ou seja, previu que o transportador de correia apresentaria defeito quando na verdade funcionou normalmente. Já o modelo treinado com o *dataset* com 5 componentes principais obteve uma acurácia maior. As duas situações não são aceitáveis para um modelo de detecção de anomalias neste cenário, portanto, foi necessário uma avaliação junto a especialistas para considerar uma adequação do conjunto de dados.

O *dataset* original foi, então, revisado para contemplar as transições de estado do transportador de correia, considerando os mesmos sete dias de antecedência para previsão, e endereçar o problema de desbalanço entre as classes. Neste *dataset*, foram incluídas variáveis computando minutos desde a mudança de estado, mínimo, máximo e média das variáveis analógicas e removidas as variáveis com valores instantâneos das variáveis analógicas.

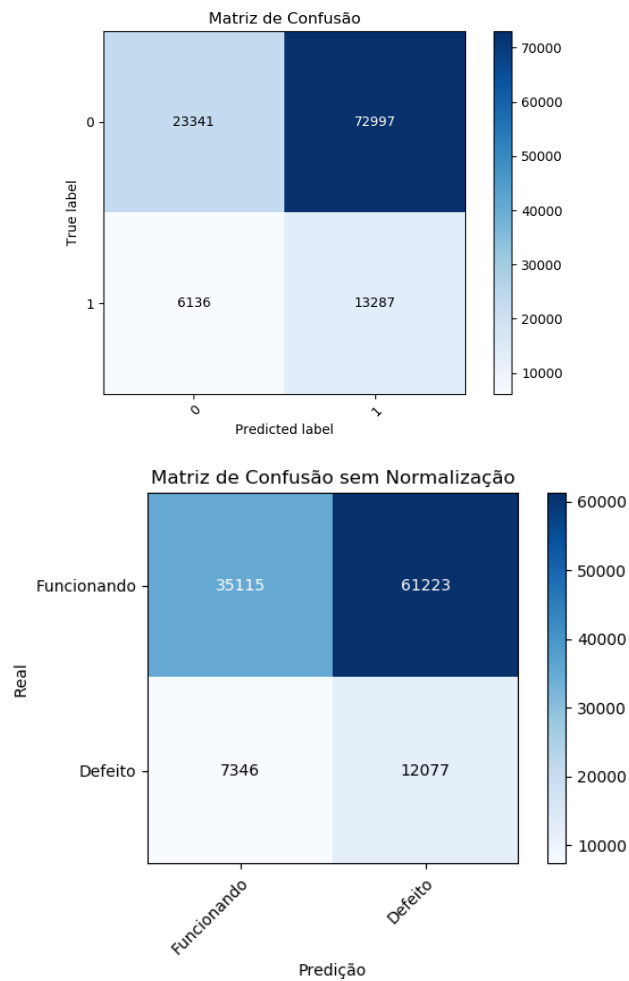


Figura 4.1: Matrizes de confusão dos modelos treinados com os dados originais de 93 variáveis (acima) e PCA (abaixo). Fonte: Autor

Nota-se, ao analisar a Tabela 4.2, que o tempo de execução reduziu consideravelmente devido ao menor número de amostras nos *datasets* utilizados. Houve também uma melhora na acurácia do modelo, sendo que o melhor resultado foi atingido pelo *dataset* com 25 variáveis selecionadas com base no teste χ^2 . Os índices apresentados para precisão, *recall* e F1 medem, respectivamente, classificações de defeito corretas, classificações de defeito corretas no total de defeitos e efetividade das classificações (HE e GARCIA, 2009). Portanto, conclui-se o modelo treinado com o *dataset* com 25 variáveis classificou 95,2% dos defeitos com sucesso, mas não é exato porque classificou 39,4% dos defeitos quando o transportador de correias funcionou.

Este cenário pode introduzir algum ganho para a empresa na operação e manutenção do transportador de correias e, no contexto do DT, precisa ser aperfeiçoado para aumentar a representatividade do ente físico e reduzir o número de detecções de defeito incorretas (evitando paradas desnecessárias do equipamento para manutenção). Em um trabalho futuro, propõe-se buscar uma nova abordagem na preparação do *dataset* ou utilização de novos modelos neste sentido.

Pensando neste processo de desenvolvimento de modelos de aprendizado de máquina,

Tabela 4.2: Acurácia atingida pelo modelo SVM com *datasets* de transição do estado. Fonte: Autor

Variáveis	Tempo	Acurácia	Precisão	Recall	F1
93	0,910 s	56,7%	0,564	0,601	0,582
25 (χ^2)	0,172 s	66,5%	0,606	0,952	0,740
5 (PCA)	0,185 s	65,1%	0,599	0,915	0,724

procurou-se uma aplicação dentro do ambiente Microsoft Azure contratado pela empresa, que ofereça funcionalidades que atendam aos requisitos identificados. Como parte da plataforma integrada do DTE, considerou-se o serviço Machine Learning Studio, pois este permite aos usuários executar todas as etapas do processo de desenvolvimento de modelos com aprendizado de máquina, e o serviço Azure Machine Learning Service porque permite a integração com o *data lake* de forma simplificada.

Como proposta para trabalho futuro, é importante estudar uma evolução do serviço de detecção de anomalias considerando a gestão de desempenho dos modelos de aprendizado de máquina em operação dentro de cada DTI e a criação de uma governança para a biblioteca de modelos disponíveis para cada tipo de equipamento ou processo no DTE.

4.2. Integração do PIMS com *Data Lake*

A Tabela 4.3 apresenta o resultado encontrado para duas métricas definidas para análise do resultado: o tempo (em segundos) desde a consulta à escrita dos dados no *data lake* e o número de linhas retornadas pela consulta. No cenário de testes foram configuradas quatro consultas, com períodos de 30 e 60 segundos pelos métodos de interpolação e média, de 10 equipamentos cada um com 109 atributos, totalizando 1.090 variáveis. Percebeu-se um desempenho aceitável neste cenário de teste em todas as consultas configuradas na aplicação. identificou-se que o desempenho da solução é dependente da latência, especialmente entre o servidor da aplicação nativa e o servidor do PIMS, e possui alto consumo de processamento e memória, o que é resolvido pela configuração de mais nós para execução de tarefas.

Tabela 4.3: Métricas para o *streaming* de dados. Fonte: documentação de projeto Vale

Método de Consulta	Interpolação		Média	
Período [s]	30	60	30	60
Duração [s]	0,06	0,09	0,05	0,70
Linhas Retornadas	29.430	58.860	29.430	58.860

Frente aos resultados atingidos, propões-se, então, a implementação da integração do PIMS diretamente com a Base de Conhecimentos conforme proposto na Subseção 3.3.2. A utilização da infraestrutura e serviços do ESB foi desconsiderada porque a integração direta

apresentou, além do bom resultado acima, menor complexidade e custo de propriedade na avaliação realizada, e a implementação com o ESB representaria mais um passo em um fluxo de dados considerado crítico para a operação em tempo real das DTIs.

4.3. Arquitetura Integrada de *Digital Twin Environment*

A experiência com o desenvolvimento do modelo de SVM, sensores de vibração e integrações do PIMS ressaltou a necessidade por uma plataforma dedicada para DT. O transportador de correia analisado possui muitos dados, consequentemente, o volume de dados gerado em tempo real é grande, o que demanda uma integração de alto desempenho entre algumas fontes dos dados (especialmente o PIMS e dispositivos IIoT) e a base de conhecimento do DT. Pelo mesmo motivo, a plataforma precisa prover escalabilidade e alta capacidade de processamento paralelo para treinar e executar os diversos modelos de detecção de anomalias e disponibilizar os resultados na base de conhecimento.

Adicionalmente, a plataforma de DT precisa integrar-se a diversos sistemas já existentes nas plantas que suportam os processos de planejamento e controle da produção (MES) e os processos de manutenção dos equipamentos (CMMS). Portanto, a plataforma proposta deve considerar, também, a integração das informações geradas por uma DTI de um equipamento a esses sistemas, além de receber dados para melhorar o resultado dos algoritmos de análise e simulação.

A arquitetura, conforme demonstrada na Figura 4.2, propõe a criação do DTE em duas camadas (Serviços e Integração) sobre serviços existentes em uma plataforma de computação na nuvem para implementar cada um dos serviços identificados como essenciais para uma DTI. As seções a seguir descrevem os serviços do DTE.

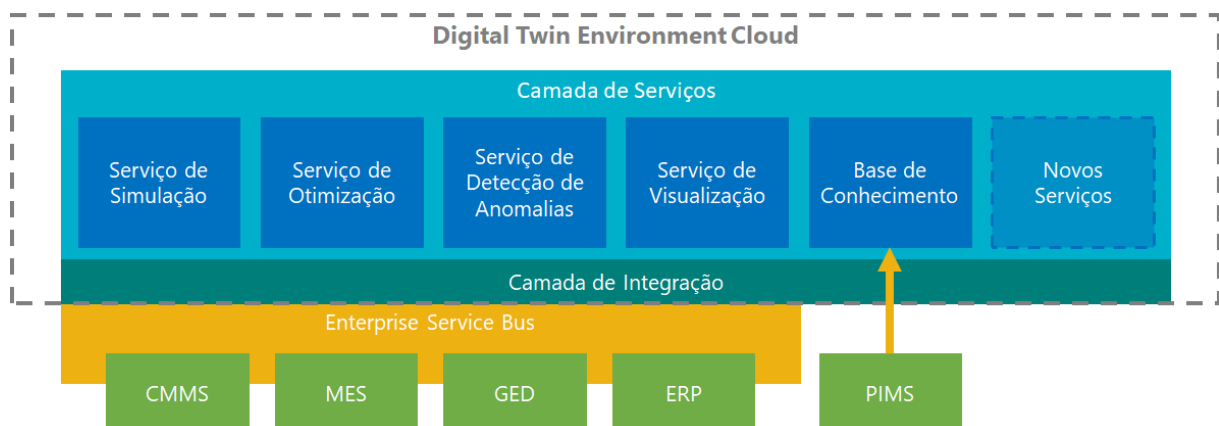


Figura 4.2: Arquitetura em alto nível para implementação da solução de DT. Fonte: Autor

4.3.1. Base de Conhecimento

A Base de Conhecimento é o local onde são armazenadas as informações necessárias para o completo funcionamento dos serviços do DTE e das DTIs e, dado o volume de dados que se espera para uma aplicação de DT, faz sentido considerar a implantação da base de conhecimento em uma infraestrutura de *data lake* para prover a escalabilidade necessária em termos de capacidade de armazenamento e processamento dos dados. Estes dois requisitos são importantes porque espera-se um crescimento rápido dos dados depender de: total de DTIs na plataforma, quantidade de modelos de detecção de anomalias por DTI, número de variáveis de cada modelo de detecção de anomalias, número de variáveis do modelo fenomenológico e resolução dos dados para alimentar os modelos.

A seguir é apresentada uma lista não exaustiva dos conjuntos de dados e informações que a base de conhecimento deve conter para atender aos requisitos da plataforma de DT.

Dados e Informações do DTE

- **Parâmetros de configuração:** conjunto das fontes de dados disponíveis e configurações específicas de cada serviço;
- **Tipos de equipamentos ou processos:** lista de equipamentos ou processos configurados no DTE para criação de DTIs;
- **Modelos de detecção de anomalias:** lista de modelos ou algoritmos de detecção de anomalias, por tipo de equipamento ou processo, aprovados para serem utilizados em uma DTI com os seus respectivos parâmetros obrigatórios;
- **Modelos fenomenológicos para os equipamentos e processos:** lista de modelos implementados, por tipo de equipamento ou processo, aprovados para serem utilizados em uma DTI com os seus respectivos parâmetros obrigatórios;

Dados e Informações de uma DTI

Cada instância tem um conjunto de informações específicas que variam conforme o ente físico evolui em seu ciclo de vida. Elas também geram ou coletam um volume muito grande de dados, o que determina uma seção específica no esquema para armazenar este conjunto de dados.

- **Parâmetros de construção do equipamento ou processo:** dados do ente físico para fins de simulação e visualização do equipamento ou processo. Considerando o transportador de correio como exemplo, as informações “como-construído” seriam: link para arquivo de dados de geometria, tipo da correia, material da correia, número de acionamentos elétricos, tensão/corrente/potência projetada dos acionamentos, quantidade de roletes, posição do esticador, tensão projetada, capacidade nominal, velocidade nominal, etc;

- **Parâmetros dos modelos de detecção de anomalias:** dados que permitam a configuração e execução do algoritmo de detecção de anomalias, por exemplo: modelo treinado, variáveis de entrada e variáveis de saída;
- **Variáveis reais:** dados das variáveis reais recebidas dos sistemas em operação na planta e corporativos, seguindo o exemplo do transportador de correia: tensão/corrente/potência dos acionamentos, estado do equipamento, modo de operação, comandos dos operadores, carga, velocidade, entre outras;
- **Variáveis simuladas:** valores simulados pelo modelo fenomenológico do equipamento ou processo;

Esquema

A criação uma nova DTI na base de conhecimento gera pelo menos quatro arquivos no *data lake* para armazenar os dados e informações seguindo a hierarquia e esquema de dados apresentados na Figura 4.3. O esquema de dados nos arquivos é definido conforme equipamentos, processos e modelos selecionados para a DTI, sendo mandatório informar o *timestamp* para garantir histórico de todas as configurações e variáveis. Portanto, adota-se um padrão de *Schema On Load* para a estrutura e semântica dos dados e informações no *data lake*.

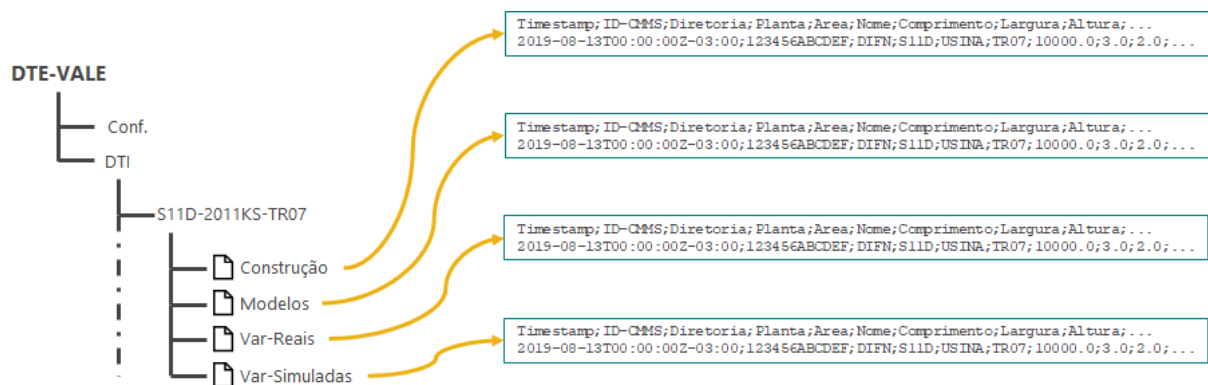


Figura 4.3: Hierarquia dos arquivos e esquema de dados proposto para o *data lake*. Fonte: Autor

Considera-se que não haverá edição dos dados históricos recebidos dos sistemas de controle da planta e sistemas corporativos, apenas inclusão de novos dados aos arquivos existentes.

4.3.2. Camada de Integração

Na camada de integração do DTE estão os serviços responsáveis por garantir o recebimento dos dados de sistemas da companhia, enviar os dados para o *data lake* e fornecer dados para outros serviços. Nesta camada são definidas três integrações, descritas a seguir.

Integração de Dados em Tempo Real

A infraestrutura dos sistemas que fornecem os dados dos processos e equipamentos das plantas deve disponibilizar serviços padronizados e simplificados a fim de se atingir baixa latência e permitir a operacionalização da plataforma do DTE em tempo real. Este é hoje um grande desafio para empresas com muitas plantas porque esses sistemas tipicamente são complexos, utilizam tecnologia proprietária e há heterogeneidade entre as plantas.

A partir desta constatação, é razoável considerar que esta integração precisa ser implementada diretamente entre as fontes de dados em tempo real e a camada de integração do DTE. A experiência apresentada na seção Subseção 3.3.2 mostra que as ferramentas nativas são capazes de atender ao requisito de latência, no entanto, não existe evidência de que a implementação do ESB atual da empresa não atenderia o mesmo requisito (objeto de trabalho futuro).

No contexto deste trabalho, considerou-se o PIMS como a fonte dos dados dos equipamentos ou processos para os serviços de simulação, detecção de anomalias e simulação. Para que os dados do PIMS sejam transferidos via *streaming* para a base de conhecimento do DTE, é necessário configurar duas aplicações diferentes: *PI Integrator for Business Analytics* (aplicação nativa do PIMS) e PI AF (cadastro de ativos e processos do PIMS), considerando que o Azure IoT Hub está previamente configurado para receber as mensagens no formato JSON, segundo um esquema pré-definido, e enviá-las para o *data lake* (na estrutura de pasta adequada).

Na aplicação nativa do PIMS deve ser configurada uma *view* que executa a consulta por todos os ativos baseados no modelo de elemento padrão, formata cada mensagem conforme um esquema JSON pré-definido e depois envia a mensagem para o *endpoint* do Azure IoT Hub.

O PI AF possui um modelo padrão para os ativos que fornecerão dados para uma DTI que deve ser utilizado para a criação de todos os ativos que enviarão dados para uma DTI. Existe uma hierarquia no PI AF à qual devem ser associados os ativos que fornecem dados para uma DTI. Portanto, após confirmação da criação da DTI, deve-se executar as ações a seguir para configurar o envio dos dados em tempo real utilizando a aplicação nativa do PIMS.

1. Criar um elemento na hierarquia do PI AF, a partir do modelo padrão, tendo como nome o identificador da DTI;
2. Adicionar ao novo elemento os atributos com os dados do PIMS;
3. Acessar a aplicação nativa do PIMS para confirmar que o novo elemento é retornado na *view* da integração;

Em um trabalho futuro, deve-se estudar um método para automatizar a configuração da aplicação nativa do PIMS para iniciar o envio dos dados de processos e equipamentos para o *endpoint* do Azure IoT Hub na criação e atualização de uma instância de DTI na plataforma.

Integração Interna

Este serviço tem como principal função integrar os dados das DTIs armazenados na Base de Conhecimento com os outros serviços da plataforma do DTE e na proposta de arquitetura ele deve disponibilizar interfaces *request/response* (síncronas) e filas de mensagens (*publish/subscribe*) conforme a natureza da integração com os serviços do DTE.

Por exemplo, o serviço de visualização, ao instanciar uma DTI, solicita os dados de construção de um equipamento através de uma interface *request/response* disponibilizada pelo serviço de Integração Interna para criar o modelo virtual na sua interface de visualização. Com o modelo virtual construído, o serviço de visualização solicita à Integração Interna que inicie o envio de dados (simulados ou reais) em sua fila de mensagens para apresentação.

Da mesma forma, sempre que uma DTI for criada na Base de Conhecimento serão criadas uma ou mais instâncias de modelos nos serviços de simulação, otimização e detecção de anomalias que receberão os dados através de filas de mensagens dedicadas para cada serviço.

Outra importante função da Integração Interna é garantir que os modelos em execução nos serviços da plataforma do DTE são atualizados sempre que houver alguma modificação, por exemplo: quando se inclui uma nova variável em um modelo de detecção de anomalias ou quando ocorrer manutenção no equipamento com mudança nos parâmetros de construção.

Integração de Configuração

A Integração de Configuração tem como objetivo criar novas DTIs na Base de Conhecimento e manter a sincronia entre os equipamentos e processos físicos com as DTIs em operação. As funções desta integração iniciam no ESB, hospedado *on premises* no ambiente de TI da companhia, que recebe os arquivos de configuração de uma DTI em uma mensagem XML postada em uma interface de integração corporativa. Os arquivos de criação são gerados pelos sistemas de CMMS e/ou gestão de documentos de engenharia segundo um esquema padronizado, proposto para utilizar o padrão *AutomationML* porque foi desenvolvido para armazenar dados de construção de equipamentos e interfaces com sistemas de controle. A Figura 4.4 apresenta um exemplo esquema *AutomationML* para um arquivo a ser utilizado na criação de uma DTI.

A integração do ESB escreve o arquivo para criação da DTI em um local do *data lake* da Base de Conhecimento, e provoca a execução de uma rotina na plataforma da nuvem para processar o arquivo de criação e criar os arquivos da DTI conforme a estrutura apresentada na Figura 4.3 ou atualizar os arquivos de configuração. No caso da atualização de uma DTI, esta rotina informa o serviço de Integração Interna qual instância foi atualizada para que o serviço execute a sincronia das configuração com todos os modelos virtuais em execução nos outros serviços da plataforma do DTE.

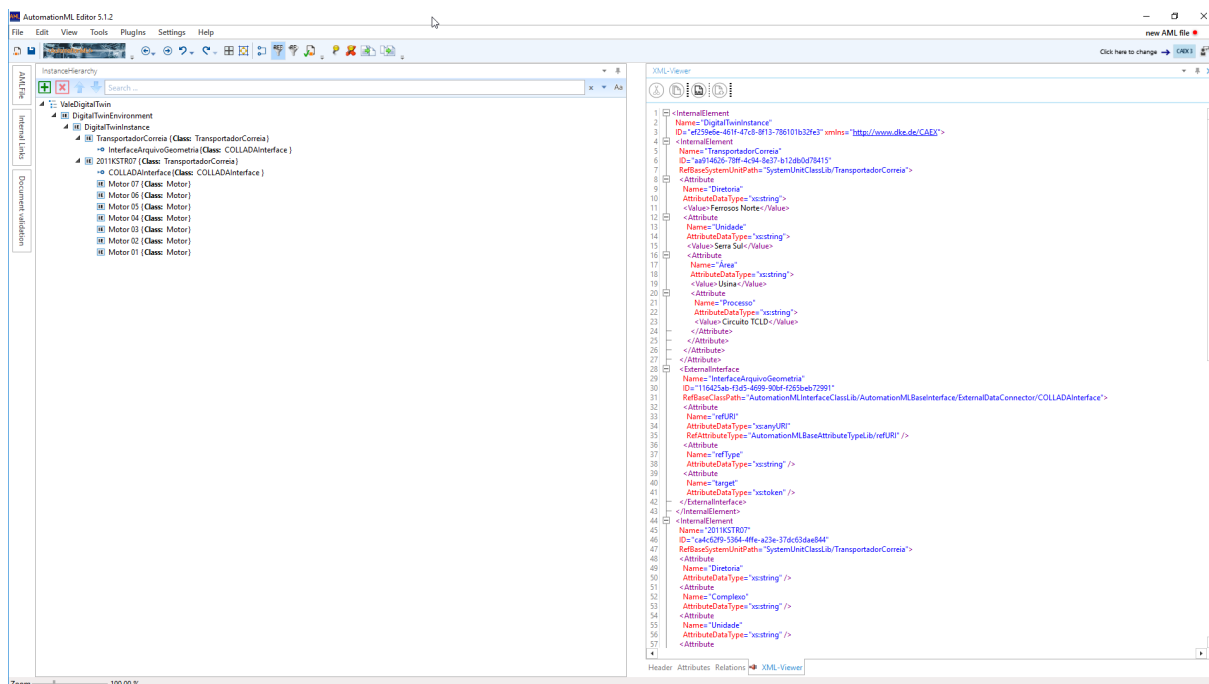


Figura 4.4: Esquema de mensagem de integração proposto baseado no padrão *AutomationML* para criação de uma DTI. Fonte: Autor

4.3.3. Serviço de Detecção de Anomalias

O propósito do Serviço de Detecção de Anomalias é executar um ou mais modelos de análise avançada de dados, isto é, aprendizado de máquina, sobre os dados recebidos em tempo real dos sistema de controle e supervisão da planta para prever uma provável parada do equipamento ou processo e alertar os usuários a fim de aumentar a disponibilidade do equipamento para produção (aumento da produtividade) ou melhorar o planejamento de manutenção (redução de custos).

Para atender ao propósito acima, o serviço deve prover um ambiente para um usuário capacitado desenvolver, treinar e publicar os algoritmos dos modelos de aprendizado de máquina para os diferentes tipos de DTI cadastrados na plataforma do DTE. No decorrer do projeto o modelo de detecção de anomalias foi desenvolvido em uma plataforma local utilizando bibliotecas existentes para a linguagem Python, mas para o ambiente da companhia propõe-se a utilização da plataforma Azure da Microsoft uma vez que a empresa já possui uma assinatura global com acesso ao Azure Machine Learning.

O ambiente do Azure Machine Learning permite a execução de todos os processos apresentados na Seção 3.2 por meio de uma interface visual, como apresentado na Figura 4.5, e possibilita o uso de bibliotecas Python com uma diversa gama de modelos de aprendizado de máquina (*scikit-learn*, *PyTorch*) e *deep learning* (TensorFlow). (MICROSOFT, 2019a,b)

Uma funcionalidade importante deste ambiente são interfaces REST e um SDK que permitam a automação do processo de implantação de modelos a partir de arquivos de configuração (MICROSOFT, 2019a). Dessa forma, o serviço de Integração Interna pode criar instâncias dos

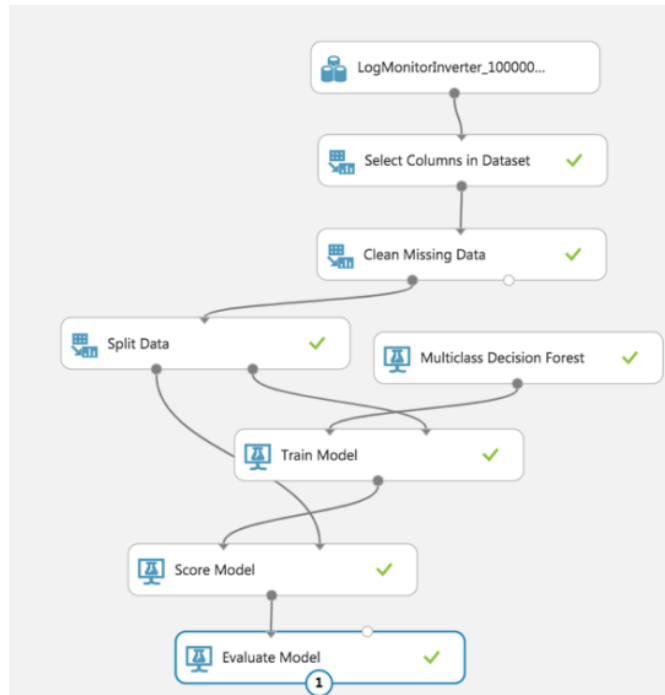


Figura 4.5: Esquema típico de um modelo de classificação no Azure Machine Learning Studio.
Fonte: (PAOLANTI *et al.*, 2018)

modelos do DTE sem a necessidade de intervenção humana. O processo automatizado também pode executar um treinamento do modelo com dados disponíveis na Base de Conhecimento, caso necessário, antes de executar a implantação do modelo.

As interfaces REST e o SDK também são utilizados para execução do modelo de uma DTI, o serviço de Integração Interna busca os dados de entrada na Base de Conhecimento, formata em uma mensagem pré-definida em um padrão JSON, executa uma chamada a um *web service* disponibilizado na criação do modelo e recebe como resposta o resultado do modelo para armazenar na Base de Conhecimento.

5. Conclusão

Uma aplicação real para uma DTI exige modelos com alta acurácia para endereçar os desafios e requisitos apresentados neste trabalho. Ainda não se conhecem plataformas de DTE que atendam a todos os requisitos da indústria de mineração, especialmente na capacidade de escalabilidade da solução no que diz respeito aos diversos tipos de equipamento e processos produtivos.

Ficou evidente que o resultado do modelo de aprendizado de máquina, neste estudo de caso, depende mais da etapa de análise exploratória e preparação dos dados do que da escolha do melhor modelo para detecção de anomalias. Apesar deste processo demandar interação humana, entende-se ser necessário o desenvolvimento de uma plataforma de DTE com autonomia a fim de reduzir a intervenção humana, sem eliminá-la por completo, para permitir a construção das DTIs de equipamentos e processos em maior escala e reduzir a demanda de manutenção dos modelos virtuais por parte de especialistas.

Do ponto vista da operação e manutenção do transportador de correia, há ausência de dados importantes para a detecção mais precisa de anomalias, por exemplo sensores de vibração dos acionamentos. Os roletes também não possuem instrumentação para medições de deslocamento dos eixos, vibração e temperatura, dados que agregariam ao modelo para melhora dos resultados. Deve-se considerar também cada bit das palavras de estado para desenvolver a capacidade de prever qual falha ocorrerá no equipamento e/ou ainda indicar ações que o operador deve tomar para evitá-la.

Finalmente, tem-se a composição de diferentes soluções, tecnologias e aplicações disponíveis na plataforma de computação na nuvem existente que permitem o desenvolvimento de um piloto para a plataforma integrada do DTE, abrindo caminho para se conhecer totalmente a implementação da arquitetura integrada apresentada neste trabalho e evoluir o desenho proposto para alcançar uma integração simples baseada em serviços escaláveis na nuvem, onde será possível criar DTIs para os diversos processos e equipamentos de uma planta de mineração.

6. Trabalhos Futuros

As seções neste capítulo descrevem trabalhos futuros identificados no decorrer do projeto e que possuem relevância para o desenvolvimento da tecnologia de DT e evolução da arquitetura integrada proposta neste trabalho.

Mesmo não sendo parte do objeto de estudo neste trabalho, os serviços apresentados neste capítulo são considerados essenciais para uma plataforma de DTE no qual as instâncias implementadas para os equipamentos e processos representam fielmente as suas respectivas entidades físicas. Portanto, faz sentido descrever as funções previstas para os serviços em questão no contexto da arquitetura de plataforma proposta.

De qualquer forma, estes serviços devem ser objeto de um projeto de pesquisa detalhado em busca de contribuição para o conhecimento do tema e definição quanto à implementação seguindo uma arquitetura como a proposta neste trabalho.

6.1. Serviço de Visualização

Uma plataforma de DTE demanda uma interface para visualização dos modelos virtuais de cada DTI em operação. A visualização do modelo virtual de uma DTI poderia ser implementada em uma plataforma de software hospedada no ambiente corporativo da empresa. No entanto, entende-se por melhor a implementação deste serviço, considerando a arquitetura apresentada, no ambiente de computação na nuvem da plataforma do DTE porque apresentaria, a princípio, menor complexidade de integração e manutenção para a plataforma em geral.

6.2. Serviço de Simulação

A simulação de equipamentos e processos é essencial para o DTE dado que a operação de uma DTI deve ser independente da operação da entidade física. A execução dos modelos fenomenológicos ou empíricos para cada uma das DTIs cadastradas na plataforma do DTE poderia ser implementada como um serviço baseado em software especialista dedicado a tal função, que forneça também uma interface para engenharia de simulação e biblioteca de modelos pré-definidos para processos e equipamentos conhecidos.

Considera-se importante realizar estudos para evoluir a arquitetura de conectividade entre os sistemas de controle e supervisão das plantas com o serviço de simulação do DTE em direção à criação de um ambiente virtual de treinamento e capacitação de pessoal operacional. No mesmo sentido, será possível evoluir a plataforma para permitir a operação autônoma dos equipamentos ou processos conforme resultados dos serviços de simulação e/ou otimização de uma DTI.

6.3. Serviço de Otimização

A otimização do equipamento ou processo é uma função importante na plataforma de DTE porque executa algoritmos que prescrevem uma ação para melhorar a operação da planta considerando um objetivo da empresa (produtividade, margem operacional, manutenção, etc.).

Este serviço poderia ser implementado como parte do serviço de detecção de anomalias, pois sua característica principal são algoritmos de análise avançada de dados; ou como parte do serviço de simulação, dependendo da solução de software selecionada e/ou evolução da arquitetura apresentada.

6.4. Outros Trabalhos

- Desenvolver um protótipo individual para cada um dos serviços propostos no desenho de arquitetura do DTE na nuvem apresentados neste trabalho;
- Detalhar soluções de mercado frente aos protótipos desenvolvidos e aos requisitos de arquitetura apresentados neste trabalho para analisar um cenário *make-or-buy* para a solução de DTE a ser adotada pela empresa;
- Sugere-se avaliar outras formas de preparação dos dados (árvore de decisões como proposto por Saimurugan *et al.* (2011)) e o desenvolvimento de outros algoritmos de detecção de anomalias com *deep learning* (*Recurrent Neural Networks* e *Long Short-Term Memory*) como referência para construção de DTI com maior fidelidade.

Referências Bibliográficas

- ABB. “Digital Twin Application - What is new”. 2019. Disponível em: <<https://new.abb.com/control-systems/features/digital-twin-applications>>.
- ADDO-TENKORANG, R., HELO, P. T. “Big data applications in operations/supply-chain management: A literature review”, *Computers and Industrial Engineering*, v. 101, pp. 528–543, 11 2016. ISSN: 03608352. doi: 10.1016/j.cie.2016.09.023.
- ASPEN TECHNOLOGY. “Aspen Cim-IO for OPC DA User’s Guide”. 2014.
- ASPEN TECHNOLOGY. “Aspen Cim-IO User’s Guide”. 2016.
- ASPEN TECHNOLOGY. “Aspen InfoPlus.21 Database User’s Guide”. 2017.
- BECERRA, K., PALACIOS, G. “Converting raw data into information-the CEMEX PIMS approach”. Em: *IEEE-IAS/PCA Cement Industry Technical Conference.*, pp. 333–344, Dallas, Texas USA, 2003. IEEE. ISBN: 1079-9931 VO. doi: 10.1109/CITCON.2003.1204734.
- CERRONE, A., HOCHHALTER, J., HEBER, G., et al.. “On the effects of modeling as-manufactured geometry: Toward digital twin”, *International Journal of Aerospace Engineering*, v. 2014, pp. 1–10, 2014. ISSN: 16875974. doi: 10.1155/2014/439278.
- CHAPMAN, P., CLINTON, J., KERBER, R., et al.. “CRISP-DM 1.0 Step-by-step data mining guide”, 2000.
- CHEN, K.-Y., CHEN, L.-S., CHEN, M.-C., et al.. “Using SVM based method for equipment fault detection in a thermal power plant”, *Computers in Industry*, v. 62, n. 1, pp. 42–50, 1 2011. ISSN: 0166-3615. doi: 10.1016/J.COMPIND.2010.05.013.
- DASH, M., LIU, H. *Feature Selection for Classification*. Relatório técnico, 1997.
- DELLOITE. *Tracking the trends 2017: The top 10 trends mining companies will face in the coming year*. Relatório técnico, 2017. Disponível em: <<https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Energy-and-Resources/gx-er-tracking-the-trends-2017.pdf>>.

- DUBEY, R., GUNASEKARAN, A., CHILDE, S. J., et al.. “The impact of big data on world-class sustainable manufacturing”, *International Journal of Advanced Manufacturing Technology*, v. 84, n. 1-4, pp. 631–645, 2016. ISSN: 14333015. doi: 10.1007/s00170-015-7674-1.
- ENDREI, M., ANG, J., ARSANJANI, A., et al.. *Patterns: Service-Oriented Architecture and Web Services*. IBM Corporation, 2004. ISBN: 9780738453170.
- ERL, T. *SOA Principles of Service Design*. Prentice Hall, 2008. ISBN: 9780132344821.
- FANG, H. “Managing data lakes in big data era: What’s a data lake and why has it become popular in data management ecosystem”, *2015 IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems, IEEE-CYBER 2015*, pp. 820–824, 2015. doi: 10.1109/CYBER.2015.7288049.
- FISCHER, J., OBST, B., LEE, B. “Integrating material flow simulation tools in a service-oriented industrial context”. Em: *15th International Conference on Industrial Informatics*, pp. 1135–1140, Emden, Alemanha, 2017. IEEE. ISBN: 9781538608371. doi: 10.1109/INDIN.2017.8104933.
- FRUCHTBAUM, J. *Bulk Materials Handling Handbook*, v. 11. Springer Science & Business Media, 2013. ISBN: 1475746954.
- GARTNER. “Gartner Top 10 Strategic Technology Trends for 2018 - Smarter With Gartner”. 2017. Disponível em: <<https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2018/>>.
- GARTNER. “5 Trends Emerge in the Gartner Hype Cycle for Emerging Technologies, 2018”. 2018a. Disponível em: <<https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/>>.
- GARTNER. “Gartner Survey Reveals Nearly Half of Organizations Implementing IoT Are Using or Plan to Use Digital Twin Initiatives in 2018”. 2018b. Disponível em: <<https://www.gartner.com/newsroom/id/3868363>>.
- GE DIGITAL. *The Digital Twin: Compressing time-to-value for digital industrial companies*. Relatório técnico, 2018. Disponível em: <https://www.ge.com/digital/sites/default/files/The-Digital-Twin_Compressing-Time-to-Value-for-Digital-Industrial-Companies.pdf>.
- GE DIGITAL. “Digital Twin”. 2019. Disponível em: <<https://www.ge.com/digital/applications/digital-twin>>.

- GELAIS, M. A., KLEINSORGE, M. D. P., MARCONDES, M. D. A., et al.. “Uma análise comparativa entre o cálculo estático e dinâmico de transportadores de correia”, *Revista Ciência e Tecnologia*, v. 19, n. 35, pp. 1–11, 12 2016.
- GRIEVES, M., VICKERS, J. “Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems”. Em: *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*, Springer International Publishing, pp. 85–113, Cham, 2016. ISBN: 9783319387567. doi: 10.1007/978-3-319-38756-7{_}4.
- GUYON, I., ELISSEEFF, A. “An Introduction to Variable and Feature Selection”, *Journal of Machine Learning Research*, v. 3, pp. 1157–1182, 2003.
- HE, H., GARCIA, E. “Learning from imbalanced data”, *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, v. 21, n. 9, pp. 1263–1284, 2009. doi: 10.1109/TKDE.2008.239. Disponível em: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5128907>.
- HELO, P., SUORSA, M., HAO, Y., et al.. “Toward a cloud-based manufacturing execution system for distributed manufacturing”, *Computers in Industry*, v. 65, n. 4, pp. 646–656, 2014. ISSN: 01663615. doi: 10.1016/j.compind.2014.01.015.
- HUAN LIU, SETIONO, R. “Chi2: Feature Selection and Discretization of Numeric Attributes”, *Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence*, pp. 388–391, 2002. doi: 10.1109/tai.1995.479783.
- KEEN, M., ACHARYA, A., BISHOP, S., et al.. *Patterns: Implementing an SOA using an enterprise service bus*. IBM Corporation, 2004. ISBN: 0738490008.
- KLETTI, J. *Manufacturing execution systems - MES*. Berlin, Heidelberg, Springer Berlin Heidelberg, 2007. ISBN: 9783540497431. doi: 10.1007/978-3-540-49744-8.
- KOLZ, R. J. “Real-Time Plant / ERP Integration”, *IEEE International Conference on Systems, Man and Cybernetics*, v. 3, pp. 2092–2097, 2000. doi: 10.1109/ICSMC.2000.886424.
- KOVALENKO, O., GRANGEL-GONZÁLEZ, I., SABOU, M., et al.. “AutomationML Ontology: Modeling Cyber-Physical Systems for Industry 4.0”, *Semantic Web Journal–Interoperability, Usability, Applicability*, v. 1, pp. 1–5, 2018.
- LI, W., WANG, Z., ZHU, Z., et al.. “Design of online monitoring and fault diagnosis system for belt conveyors based on wavelet packet decomposition and support vector machine”, *Advances in Mechanical Engineering*, v. 2013, 2013. ISSN: 16878132. doi: 10.1155/2013/797183.

LODEWIJKS, G., DUINKERKEN, M., LOPEZ DE LA CRUZ, A., et al.. “The application of RFID technology in belt conveyor systems”, *International Materials Handling Conference BeltCon 14*, pp. 1–17, 2007.

LODEWIJKS, G. “Strategies for automated maintenance of belt conveyor systems”, *Bulk Solids Handling*, v. 24, n. 1, pp. 16–24, 2004. ISSN: 01739980.

LÜDER, A., SCHMIDT, N. *AutomationML in a Nutshell*. Relatório Técnico November, 2015. Disponível em: <https://www.automationml.org/o.red/uploads/dateien/1447420977-AutomationMLinaNutshell_151104.pdf>.

LUO, W., HU, T., ZHANG, C., et al.. “Digital twin for CNC machine tool: modeling and using strategy”, *Journal of Ambient Intelligence and Humanized Computing*, v. 10, n. 3, pp. 1129–1140, 3 2019. ISSN: 18685145. doi: 10.1007/s12652-018-0946-5.

MANUFACTURING ENTERPRISE SOLUTIONS ASSOCIATION. “MESA International”. 2018. Disponível em: <<http://www.mesa.org/>>.

MARCONDES, M. D. A. *Desenvolvimento de uma metodologia para cálculo e padronização de componentes estruturais para transportadores de correia*. Tese de Doutorado, Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

MATHIS, C. “Data Lakes”, *Datenbank-Spektrum*, v. 17, n. 3, pp. 289–293, 2017. ISSN: 1618-2162. doi: 10.1007/s13222-017-0272-7.

MICROSOFT. “Azure Data Lake Storage Gen2”. 2018. Disponível em: <<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction>>.

MICROSOFT. “Azure Machine Learning Service Documentation”. 2019a. Disponível em: <<https://docs.microsoft.com/en-us/azure/machine-learning/service/>>.

MICROSOFT. “Machine Learning Studio Documentation”. 2019b. Disponível em: <<https://docs.microsoft.com/en-us/azure/machine-learning/studio/>>.

MINING.COM. “BeltGenius ERIC: Digital twin increases the performance of belt conveyor systems”. 2018. Disponível em: <<https://www.mining.com/web/beltgenius-eric-digital-twin-increases-performance-belt-conveyor-systems/>>.

MORARIU, O., MORARIU, C., BORANGIU, T., et al.. “Manufacturing Systems at Scale with Big Data Streaming and Online Machine Learning”, *Studies in Computational Intelligence*, v. 762, pp. 253–264, 2018. ISSN: 1860949X. doi: 10.1007/978-3-319-73751-5{-}19.

- MOURA, R. L. D., CEOTTO, L. D. L. F., GONZALEZ, A. “Industrial IoT and Advanced Analytics Framework: An Approach for the Mining Industry”. Em: *Proceedings of the 2017 International Conference on Computational Science and Computational Intelligence*, v. 1, pp. 1308–1314, Las Vegas, NV, USA, 2017. IEEE. ISBN: 9781538626528. doi: 10.1109/CSCI.2017.228.
- NASCIMENTO, R. *Inspecção de Transportadores de Correia: Arquitetura Integrada para uma Plataforma de Inspecção com uso de VANTs*. Dissertação (Mestrado em Instrumentação, Controle e Automação de Processos de Mineração) - Escola de Minas, Universidade Federal de Ouro Preto, Ouro Preto, 2018.
- NASCIMENTO, R., CARVALHO, R., DELABRIDA, S., et al.. “An Integrated Inspection System for Belt Conveyor Rollers - Advancing in an Enterprise Architecture”, *Proceedings of the 19th International Conference on Enterprise Information Systems*, v. 2, pp. 190–200, 2017. doi: 10.5220/0006369101900200.
- OSISOFT. “PI Integrator for Business Analytics 2018 User Guide”. 2018.
- PANG, Y., LODEWIJKS, G. “The application of RFID technology in large-scale dry bulk material transport system monitoring”, *EESMS 2011 - 2011 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems, Proceedings*, pp. 5–9, 2011. doi: 10.1109/EESMS.2011.6067043.
- PAOLANTI, M., ROMEO, L., FELICETTI, A., et al.. “Machine Learning approach for Predictive Maintenance in Industry 4.0”. Em: *2018 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, pp. 1–6, Oulu, Finlândia, 2018. IEEE. ISBN: 9781538646434. doi: 10.1109/MESA.2018.8449150.
- PRINSLOO, W. “A Digital Twin for the Bulk Materials Industry”. 2019. Disponível em: <<https://rocky.esss.co/blog/digital-twin-for-the-bulk-materials-industry/>>.
- RAMAKRISHNAN, R., SRIDHARAN, B., DOUCEUR, J. R., et al.. “Azure Data Lake Store: A Hyperscale Distributed File Service for Big Data Analytics”. Em: *Proceedings of the 2017 ACM International Conference on Management of Data*, pp. 51–53, Chicago, Illinois, USA, 2017. ACM. ISBN: 9781450341974. doi: 10.1007/978-1-4842-2143-3{-}9.
- RASCHKA, S., MIRJALILI, V. *Python Machine Learning - Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow*. Segunda ed ed. Birmingham, Reino Unido, Packt Publishing, 2017. ISBN: 978-1-78712-593-3.

- RÍOS, J., HERNÁNDEZ, J. C., OLIVA, M., et al.. “Product avatar as digital counterpart of a physical individual product: Literature review and implications in an aircraft”, *Advances in Transdisciplinary Engineering*, v. 2, pp. 657–666, 2015. ISSN: 2352-7528. doi: 10.3233/978-1-61499-544-9-657.
- ROSEN, R., VON WICHERT, G., LO, G., et al.. “About the importance of autonomy and digital twins for the future of manufacturing”. Em: *IFAC-PapersOnLine*, v. 28, pp. 567–572, 2015. ISBN: 2405-8963. doi: 10.1016/j.ifacol.2015.06.141.
- SAEYS, Y., INZA, I., LARRANAGA, P. “A review of feature selection techniques in bioinformatics”, *Bioinformatics*, v. 23, n. 19, pp. 2507–2517, 10 2007. doi: 10.1093/bioinformatics/btm344.
- SAIMURUGAN, M., RAMACHANDRAN, K., SUGUMARAN, V., et al.. “Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine”, *Expert Systems with Applications*, v. 38, n. 4, pp. 3819–3826, 4 2011. ISSN: 0957-4174. doi: 10.1016/J.ESWA.2010.09.042.
- SCHROEDER, G., STEINMETZ, C., PEREIRA, C. E., et al.. “Visualising the digital twin using web services and augmented reality”. Em: *IEEE International Conference on Industrial Informatics (INDIN)*, pp. 522–527, Emden, Alemanha, 2017. IEEE. ISBN: 9781509028702. doi: 10.1109/INDIN.2016.7819217.
- SCHROEDER, G. N., STEINMETZ, C., PEREIRA, C. E., et al.. “Digital Twin Data Modeling with AutomationML and a Communication Methodology for Data Exchange”, *IFAC-PapersOnLine*, v. 49, n. 30, pp. 12–17, 2016. ISSN: 24058963. doi: 10.1016/j.ifacol.2016.11.115.
- SHAFTO, M., CONROY, M., DOYLE, R., et al.. *Modeling, Simulation, information Technology & Processing Roadmap*. Relatório técnico, National Aeronautics and Space Administration, Washington, EUA, 2012.
- SIEMENS. “Digital Twin”. 2019. Disponível em: <<https://www.plm.automation.siemens.com/global/en/our-story/glossary/digital-twin/24465>>.
- SUN, W., CHEN, J., LI, J. “Decision tree and PCA-based fault diagnosis of rotating machinery”, *Mechanical Systems and Signal Processing*, v. 21, n. 3, pp. 1300–1317, 4 2007. ISSN: 0888-3270. doi: 10.1016/J.YMSSP.2006.06.010.
- TAO, F., ZHANG, M. “Digital Twin Shop-Floor: A New Shop-Floor Paradigm Towards Smart Manufacturing”, *IEEE Access*, v. 5, pp. 20418–20427, 2017. ISSN: 21693536. doi: 10.1109/ACCESS.2017.2756069.

- TAO, F., CHENG, J., QI, Q., et al.. “Digital twin-driven product design, manufacturing and service with big data”, *International Journal of Advanced Manufacturing Technology*, v. 94, n. 9-12, pp. 3563–3576, 2018. ISSN: 14333015. doi: 10.1007/s00170-017-0233-1.
- THE APACHE SOFTWARE FOUNDATION. “Apache Hadoop Documentation”. 2018a. Disponível em: <<http://hadoop.apache.org/docs/current/>>.
- THE APACHE SOFTWARE FOUNDATION. “Apache Kafka Documentation”. 2018b. Disponível em: <<https://kafka.apache.org/documentation/>>.
- THE APACHE SOFTWARE FOUNDATION. “Apache Spark Documentation”. 2018c. Disponível em: <<https://spark.apache.org/docs/latest/>>.
- THE INTERNATIONAL SOCIETY OF AUTOMATION. “ISA95, Enterprise-Control System Integration”. 2018. Disponível em: <<https://www.isa.org/isa95/>>.
- UHLEMANN, T. H., SCHOCK, C., LEHMANN, C., et al.. “The Digital Twin: Demonstrating the Potential of Real Time Data Acquisition in Production Systems”, *Procedia Manufacturing*, v. 9, pp. 113–120, 2017. ISSN: 23519789. doi: 10.1016/j.promfg.2017.04.043.
- VOITH. “BeltGenius”. 2019. Disponível em: <<http://voith.com/corp-en/products-services/automation-digital-solutions/beltgenius.html>>.
- WALKER, C., ALREHAMY, H. “Personal Data Lake with Data Gravity Pull”. Em: *IEEE 5th International Conference on Big Data and Cloud Computing, BDCloud 2015*, pp. 160–167. IEEE, 2015. ISBN: 9781467371834. doi: 10.1109/BDCloud.2015.62.
- WIDODO, A., YANG, B.-S. “Mechanical Systems and Signal Processing Support vector machine in machine condition monitoring and fault diagnosis”, *Mechanical Systems and Signal Processing*, v. 21, pp. 2560–2574, 2007. doi: 10.1016/j.ymssp.2006.12.007.
- WORLD ECONOMIC FORUM. *Digital Transformation Initiative: Mining and Metals Industry*. Relatório técnico, World Economic Forum, 2017. Disponível em: <<http://reports.weforum.org/digital-transformation/mining-and-metals/>>.
- XU, Y., SUN, Y., LIU, X., et al.. “A Digital-Twin-Assisted Fault Diagnosis Using Deep Transfer Learning”, *IEEE Access*, v. 7, pp. 19990–19999, 2019. ISSN: 21693536. doi: 10.1109/ACCESS.2018.2890566.

YANG, K., SHAHABI, C. “A PCA-based similarity measure for multivariate time series”. Em: *Proceedings of the 2nd ACM international workshop on Multimedia databases - MMDB '04*, p. 65, New York, New York, USA, 2004. ACM Press. ISBN: 1581139756. doi: 10.1145/1032604.1032616.

ZIEGLER, M. “Digital twin based method to monitor and optimize belt conveyor maintenance and operation”. Em: *Proceedings of the 2019 Coal Operators Conference*, n. February, pp. 125–132, Wollongong, Australia, 2019. University of Wollongong.

Apêndice A: Lista dos sensores da correia transportadora identificados no PIMS

S11D_2011KS.DEFO.TR07M10	S11D_2011KS.TR07M1.WT	S11D_2011KS.TR07.NP
S11D_2011KS.DEFO.TR07M9	S11D_2011KS.TR07M2.ET	S11D_2011KS.TR07.ST
S11D_2011KS.EQPE.TR07M10	S11D_2011KS.TR07M2.HORI	S11D_2011KS.TR07.STS
S11D_2011KS.EQPE.TR07M9	S11D_2011KS.TR07M2.IT	S11D_2011KS.TR07.STS.AUX
S11D_2011KS.POAB.TR07M10	S11D_2011KS.TR07M2.JT	S11D_2011KS.TR07.STS.C
S11D_2011KS.POAB.TR07M9	S11D_2011KS.TR07M2.NP	S11D_2011KS.TR07.WT
S11D_2011KS.TR01M07.HORI	S11D_2011KS.TR07M2.ST	S11D_2011KS.TR_07M10.LimSegVelEmg
S11D_2011KS.TR01M07.NOINT	S11D_2011KS.TR07M2.STS.AUX	S11D_2011KS.TR_07M1.ET
S11D_2011KS.TR01M07.NP	S11D_2011KS.TR07M2.STS.C	S11D_2011KS.TR_07M1.IT
S11D_2011KS.TR01M07.RFS	S11D_2011KS.TR07M2.WT	S11D_2011KS.TR_07M1.JT
S11D_2011KS.TR01M07.RUN	S11D_2011KS.TR07M3.ET	S11D_2011KS.TR_07M1.ST
S11D_2011KS.TR01M07.STS.AUX	S11D_2011KS.TR07M3.HORI	S11D_2011KS.TR_07M1.WT
S11D_2011KS.TR01M07.STS.C	S11D_2011KS.TR07M3.IT	S11D_2011KS.TR_07M2.ET
S11D_2011KS.TR01M07.TRIP	S11D_2011KS.TR07M3.JT	S11D_2011KS.TR_07M2.IT
S11D_2011KS.TR07M01.ActualSpeed	S11D_2011KS.TR07M3.NP	S11D_2011KS.TR_07M2.JT
S11D_2011KS.TR07M01.CM.ActualSpeed	S11D_2011KS.TR07M3.ST	S11D_2011KS.TR_07M2.ST
S11D_2011KS.TR07M01.NOINT	S11D_2011KS.TR07M3.STS.AUX	S11D_2011KS.TR_07M2.WT
S11D_2011KS.TR07M01.RFS	S11D_2011KS.TR07M3.STS.C	S11D_2011KS.TR_07M3.ET
S11D_2011KS.TR07M01.RUN	S11D_2011KS.TR07M3.WT	S11D_2011KS.TR_07M3.IT
S11D_2011KS.TR07M01.TRIP	S11D_2011KS.TR07M4.ET	S11D_2011KS.TR_07M3.JT
S11D_2011KS.TR07M02.ActualSpeed	S11D_2011KS.TR07M4.HORI	S11D_2011KS.TR_07M3.ST
S11D_2011KS.TR07M02.CM.ActualSpeed	S11D_2011KS.TR07M4.IT	S11D_2011KS.TR_07M3.WT
S11D_2011KS.TR07M02.NOINT	S11D_2011KS.TR07M4.JT	S11D_2011KS.TR_07M4.ET
S11D_2011KS.TR07M02.RFS	S11D_2011KS.TR07M4.NP	S11D_2011KS.TR_07M4.IT
S11D_2011KS.TR07M02.RUN	S11D_2011KS.TR07M4.ST	S11D_2011KS.TR_07M4.JT
S11D_2011KS.TR07M02.TRIP	S11D_2011KS.TR07M4.STS.AUX	S11D_2011KS.TR_07M4.ST
S11D_2011KS.TR07M03.ActualSpeed	S11D_2011KS.TR07M4.STS.C	S11D_2011KS.TR_07M4.WT
S11D_2011KS.TR07M03.CM.ActualSpeed	S11D_2011KS.TR07M4.WT	S11D_2011KS.TR_07M5.ET
S11D_2011KS.TR07M03.NOINT	S11D_2011KS.TR07M5.ET	S11D_2011KS.TR_07M5.IT
S11D_2011KS.TR07M03.RFS	S11D_2011KS.TR07M5.HORI	S11D_2011KS.TR_07M5.JT
S11D_2011KS.TR07M03.RUN	S11D_2011KS.TR07M5.IT	S11D_2011KS.TR_07M5.ST
S11D_2011KS.TR07M03.TRIP	S11D_2011KS.TR07M5.JT	S11D_2011KS.TR_07M5.WT
S11D_2011KS.TR07M04.ActualSpeed	S11D_2011KS.TR07M5.NP	S11D_2011KS.TR_07M6.ET
S11D_2011KS.TR07M04.CM.ActualSpeed	S11D_2011KS.TR07M5.ST	S11D_2011KS.TR_07M6.IT
S11D_2011KS.TR07M04.NOINT	S11D_2011KS.TR07M5.STS.AUX	S11D_2011KS.TR_07M6.JT
S11D_2011KS.TR07M04.RFS	S11D_2011KS.TR07M5.STS.C	S11D_2011KS.TR_07M6.ST
S11D_2011KS.TR07M04.RUN	S11D_2011KS.TR07M5.WT	S11D_2011KS.TR_07M6.WT
S11D_2011KS.TR07M04.TRIP	S11D_2011KS.TR07M6.ET	S11D_2011KS.TR_07M7.ET
S11D_2011KS.TR07M05.ActualSpeed	S11D_2011KS.TR07M6.HORI	S11D_2011KS.TR_07M7.IT
S11D_2011KS.TR07M05.CM.ActualSpeed	S11D_2011KS.TR07M6.IT	S11D_2011KS.TR_07M7.JT
S11D_2011KS.TR07M05.NOINT	S11D_2011KS.TR07M6.JT	S11D_2011KS.TR_07M7.ST
S11D_2011KS.TR07M05.RFS	S11D_2011KS.TR07M6.NP	S11D_2011KS.TR_07M7.WT
S11D_2011KS.TR07M05.RUN	S11D_2011KS.TR07M6.ST	S11D_2011KS.TR_07M9.LimSegVelEmg

S11D_2011KS.TR07M05.TRIP	S11D_2011KS.TR07M6.STS_AUX	S11D_2011KS_TR_07_BrakeFault
S11D_2011KS.TR07M06.ActualSpeed	S11D_2011KS.TR07M6.STS_C	S11D_2011KS_TR_07_ET
S11D_2011KS.TR07M06.CM_ActualSpeed	S11D_2011KS.TR07M6.WT	S11D_2011KS_TR_07_Fault
S11D_2011KS.TR07M06.NOINT	S11D_2011KS.TR07M7_ET	S11D_2011KS_TR_07_HeadMSR
S11D_2011KS.TR07M06.RFS	S11D_2011KS.TR07M7_HORI	S11D_2011KS_TR_07_HeartBeat
S11D_2011KS.TR07M06.RUN	S11D_2011KS.TR07M7_JT	S11D_2011KS_TR_07_JT
S11D_2011KS.TR07M06.TRIP	S11D_2011KS.TR07M7_JT	S11D_2011KS_TR_07_InspMode
S11D_2011KS.TR07M07.ActualSpeed	S11D_2011KS.TR07M7_NP	S11D_2011KS_TR_07_JT
S11D_2011KS.TR07M07.CM_ActualSpeed	S11D_2011KS.TR07M7_ST	S11D_2011KS_TR_07_PBComm
S11D_2011KS.TR07M07.NOINT	S11D_2011KS.TR07M7.STS_AUX	S11D_2011KS_TR_07_Ready
S11D_2011KS.TR07M07.RFS	S11D_2011KS.TR07M7.STS_C	S11D_2011KS_TR_07_ST
S11D_2011KS.TR07M07.RUN	S11D_2011KS.TR07M7_WT	S11D_2011KS_TR_07_StartFault
S11D_2011KS.TR07M07.TRIP	S11D_2011KS.TR07_ActualSpeed	S11D_2011KS_TR_07_TailMSR
S11D_2011KS.TR07M1_ET	S11D_2011KS.TR07_ET	S11D_2011KS_TR_07_WT
S11D_2011KS.TR07M1_HORI	S11D_2011KS.TR07_HORI	S11D_2011KS_TR_07_Warning
S11D_2011KS.TR07M1_JT	S11D_2011KS.TR07_JT	S11D_DEFO_DP_2011KS_TR07M10
S11D_2011KS.TR07M1_JT	S11D_2011KS.TR07_JT	S11D_DEFO_DP_2011KS_TR07M9
S11D_2011KS.TR07M1_NP	S11D_2011KS.TR07_LIT0810	S11D_EQPE_DP_2011KS_TR07M10
S11D_2011KS.TR07M1_ST	S11D_2011KS.TR07_LIT0810/0811_01	S11D_EQPE_DP_2011KS_TR07M9
S11D_2011KS.TR07M1.STS_AUX	S11D_2011KS.TR07_LIT0810/0811_02	S11D_POAB_DP_2011KS_TR07M10
S11D_2011KS.TR07M1.STS_C	S11D_2011KS.TR07_LIT0811	S11D_POAB_DP_2011KS_TR07M9