

UMA ABORDAGEM ORIENTADA A DADOS PARA ANÁLISE DE EVENTOS E ALARMES NA INDÚSTRIA 4.0

AGUINALDO BEZERRA*, IVANOVITCH SILVA†, GUSTAVO LEITÃO†, LUIZ AFFONSO GUEDES‡

**Programa de Pós-graduação em Engenharia Elétrica e de Computação
Universidade Federal do Rio Grande do Norte
Natal, RN, Brasil*

†*Instituto Metr pole Digital
Universidade Federal do Rio Grande do Norte
Natal, RN, Brasil*

‡*Departamento de Engenharia de Computa o e Automa o
Universidade Federal do Rio Grande do Norte
Natal, RN, Brasil*

Emails: aguinaldo@ufrn.edu.br, ivan@imd.ufrn.br, gustavo.leitao@imd.ufrn.br,
affonso@dca.ufrn.br

Abstract— Alarm and event logs are a latent and massive source of knowledge commonly underutilized in industry. The advances in Data Science and Big Data practices and the current high competitiveness and efficiency landscape, boosted by Industry 4.0, demand new approaches when dealing with plant-generated data. This work proposes the use of Exploratory Data Analysis (EDA) as a preliminary data-driven analysis approach, fully able to extract valuable insights from data.

Keywords— Big Data, Data Science, Alarm Management, Exploratory Data Analysis.

Resumo— Os registros de alarmes e eventos constituem uma latente e massiva fonte de conhecimento comumente subaproveitada na ind stria. Os avan os das pr ticas em Ci ncia de Dados e *Big Data* e o atual cen rio de alta competitividade e efici ncia, impulsionado pela Ind stria 4.0, demandam novas abordagens no trato dos dados oriundos das plantas. Este trabalho prop e a utiliza o da An lise Explorat ria de Dados (EAD) como abordagem preliminar de an lise, orientada somente aos dados e com alto potencial na extra o de valiosos *insights* a partir dos dados.

Palavras-chave— *Big Data*, Ci ncia de Dados, Gerenciamento de Alarmes, An lise Explorat ria de Dados.

1 Introdu o

A ind stria definitivamente j  entrou na era da Ci ncia de Dados e do *Big Data*. Com as inova es e avan os tecnol gicos nesses dom nios, a possibilidade de transformar em informa o  til a grande e crua massa de dados produzidas pelo monitoramento de opera o de plantas industriais, de modo a orientar operadores e gerentes na tomada de decis es, no planejamento de a es e na melhoria cont nua das opera es, v m constituindo uma inexor vel tend ncia no setor industrial (Allen-Bradley, 2017).

No contexto da Ind stria 4.0, onde os sistemas ciber-f sicos e a an lise inteligente fundem-se em uma nova forma de pensar o gerenciamento da produ o e transforma o da f brica (Lee et al., 2014), as tecnologias associadas com *Big Data* t m papel decisivo. Contudo, resta muito a ser alcan ado na moldagem do j  bem estabelecido paradigma do *Big Data*   realidade e   necessidade industrial. Essa moldagem refere-se, elementarmente,  s defini es do que fazer com os dados industriais e de como fazer, visto que a seara industrial difere em v rios aspectos dos outros contextos onde as tecnologias de *Big Data* v m sendo tipicamente e consolidadamente utili-

zadas (Lee et al., 2014; Basanta-Val, 2017), como m dias sociais, *marketing*, com rcio e governo eletr nicos, sa de, economia, entre outros.

Nesse espectro, diversos trabalhos recentes v m abordando a problem tica do "Big Data industrial", com cerne nos v rios problemas e desafios a serem superados com a definitiva absor o do paradigma e nas propostas de arquiteturas adequadas   ind stria (Lee et al., 2014; Obitko et al., 2013; Basanta-Val, 2017; Lee et al., 2015), o que sugere que a tem tica encontra-se ainda em fase de desenvolvimento e consolida o.

Diante da relev ncia do tema, o presente trabalho concentra-se na apresenta o de uma abordagem orientada a dados para an lise de registros de dados de eventos e alarmes referentes a opera o de plantas industriais, com foco na An lise Explorat ria de Dados (AED) e as diferentes formas de visualizar seus resultados.

O restante do trabalho est  estruturado da seguinte forma. A Se o 2 realiza uma contextualiza o do paradigma do *Big Data* no setor industrial. A Se o 3 apresenta conceitos inerentes    rea de gerenciamento de alarmes e explicita a import ncia da an lise de dados de eventos e alarmes no contexto de supervis o de processos industriais. A Se o 4 congrega algumas abordagens t picas de

AED aplicadas ao conjunto de dados analisado e a Seção 5 justifica a necessidade de uma infraestrutura de *Big Data* apropriada à realidade industrial para que as análises propostas sejam viabilizadas. A Seção 6 encerra o trabalho com as conclusões e indicação de trabalhos futuros.

2 A Ciência de Dados e *Big Data* na Indústria

Ciência de Dados, no sentido mais amplo, refere-se ao conjunto interdisciplinar de tecnologias, processos e sistemas utilizados para extrair dos dados conhecimento e introspecções definitivas no embasamento do processo decisório sob várias fontes de incerteza (Academies, 2017). Dando suporte à evolução das práticas em Ciências de Dados, emerge o conceito de *Big Data*, que se refere, fundamentalmente, ao domínio científico da computação aplicada a grandes, complexos e diversos conjuntos de dados (Bryant et al., 2008).

Embora usualmente ambos os conceitos estejam fundidos, sinônimos ou complementares na literatura, a Ciência de Dados reúne as amplas áreas da matemática, estatística, engenharia e ciência da computação no emprego de técnicas e processos de análise numérica e estatística, mineração de dados, aprendizado de máquina e visualização, enquanto que o conceito de *Big Data* firma-se nos meios de viabilizar esse processamento na atual realidade de “dilúvio de dados”. Nesse cenário, as abordagens tradicionais mostram-se insatisfatórias ou inadequadas para tratar com problemas *Big Data*, pois necessita de uma infraestrutura computacional rápida e moderna de captura, armazenamento, gerenciamento, consulta, análise, compartilhamento e visualização de dados, forjada fundamentalmente em computação distribuída, paralela e de alto desempenho. É evidente, portanto, a relação de complementaridade entre os conceitos, dos quais os dados são protagonistas e fonte de onde todo valor pode ser extraído.

É importante ressaltar que essa orientação a dados vem aperfeiçoando o processo decisório e tem sido propiciada por tecnologias pioneiramente introduzidas por companhias de Tecnologia da Informação (Obitko et al., 2013), sendo notório movimento de convergência em torno da ideia do *Big Data* por parte de vários outros domínios.

Na indústria, esse movimento não é diferente. Com o estabelecimento dos conceitos de Indústria 4.0 como novo paradigma, a Ciência de Dados e o *Big Data* vêm constituindo o caminho natural para o aumento do desempenho e, consequentemente, da competitividade do setor. Nesse contexto, Reis and Gins (2017) apresentam uma abrangente retrospectiva dos paradigmas e tendências que nortearam e norteiam a indústria, com enfoque no gerenciamento de processos industriais, ao longo de quase um século de evolução, culminando na era atual do *Big Data* e sua inserção

no processo de contínua evolução dos sistemas e processos industriais.

O *Big Data* Industrial é conceito já bastante difundido na literatura e herda, de acordo com Basanta-Val (2017), as características definidoras do conceito original de *Big Data* de propósito geral, tais como volume, variedade, velocidade, variabilidade e veracidade (5 Vs), bem como estende esse conceito adicionando novos Vs: visibilidade, que se refere à descoberta de introspecções inesperadas sobre dados já processados e valor, que concerne à nova cadeia de valor agregado advindo da análise dos dados. Enfatiza-se ainda que, na indústria, os dados são mais estruturados, correlatos e prontos para análise, já que são gerados automaticamente por equipamentos em ambientes e processos controlados.

3 Análise de Dados de Alarmes e Eventos

Quando se trata da enorme massa de dados produzidos atualmente nas plantas industriais, logo vem à mente a imagem de um grande e diverso repositório de séries temporais das variáveis de processo provenientes de equipamentos de campo (sensores e atuadores). Entretanto, também advém das plantas outra grande massa de dados: os registros (*oulogs*) de eventos e alarmes. Esses dados, também intimamente ligados aos processos da planta, compõem o registro dos episódios relevantes da operação da planta industrial, sendo que, geralmente, apenas parte deles são levados ao conhecimento dos operadores por meio de Interfaces Homem-Máquina (IHM).

No contexto de um processo industrial, um evento, sem prejuízo do sentido amplo da palavra, consiste em qualquer ocorrência relevante no âmbito operacional do sistema. Os eventos verbalizam as condições gerais de operação da planta e, geralmente, não requerem qualquer tipo de ação (de conhecimento ou interventiva) imediata ou simplesmente não requerem qualquer tipo de ação do operador. Já os alarmes, de acordo com a Norma ISA-18.2 (ANSI/ISA, 2016), “são meios audíveis e/ou visíveis de indicar ao operador malfuncionamento de equipamento, desvios no processo ou condição anormal, exigindo uma resposta do operador”. Os alarmes constituem, portanto, subconjunto dos eventos, apontam os eventuais problemas e demandam um conjunto de ações preventivas ou corretivas por parte dos operadores.

Apesar de ser matéria já bem estabelecida no ramo da automação, o gerenciamento de alarmes, de acordo com O’Brien and Woll (2010), é um dos mais subvalorizados e subutilizados aspectos nos processos de automação e muitas plantas ainda utilizam uma filosofia de gerenciamento de alarmes datada de sua construção, tornando o processo gradativamente menos eficaz. A ineficácia dos sistemas de alarmes em detectar situ-

ações anormais, é inclusive, apontada como uma das causas de diversos acidentes industriais recentes (EEMUA, 2007; Habibi and Hollifield, 2006).

Com a melhoria tecnológica dos sistemas de automação, a quantidade de dados e informações obtidas a partir dos sistemas industriais tem crescido exponencialmente (Habibi and Hollifield, 2006). Em determinadas situações, os operadores são submetidos a uma avalanche de eventos oriundos dos sistemas de gerenciamento de alarmes e, por isto, não é incomum a tomada de decisões operacionais incorretas (da Silva et al., 2016). Assim, essa grande massa de informações e dados, se não apresentada adequadamente aos operadores, pode tornar a atividade de gerenciamento de alarmes e eventos bastante dificultada.

Além disso, nos últimos anos, a área de gerenciamento de alarmes vem recebendo uma grande atenção relativa à busca pelos melhores parâmetros e configurações dos sistemas, de forma a tornar a operação do processo mais eficiente (Wang et al., 2016). No entanto, os eventos, de modo geral, não tem recebido o mesmo tratamento. Comumente, os eventos são considerados apenas diante da necessidade do aprofundamento da análise de algum incidente ou desvio operacional detectado, embora constituam uma importante fonte de dados que pode ser proativamente aproveitada.

Contribuindo para a modernização da área, a Norma ANSI/ISA-18.2 (ANSI/ISA, 2016) padroniza uma série de aspectos do gerenciamento de alarmes e eventos, dirimindo antigos problemas em termos de definições, classificações, requerimentos, ciclos de vida, atividades e processos de trabalho dos operadores, bem como divergências de regulações existentes mas não tão específicas à área. Em especial, a ANSI/ISA-18.2 prega uma mudança na filosofia de gerenciamento de alarmes, relacionando a área muito mais à evolução dos processos de trabalho do que a questões apenas de *hardware* ou *software*. Para isso, recomenda a definição vários aspectos importantes como os papéis, responsabilidades, política de priorização, estratégia de monitoramento, gerenciamento de mudanças, etc. A norma sugere, ainda, a utilização de métodos avançados de análise de dados nas etapas de monitoramento, avaliação, auditoria e *benchmark* do ciclo de vida dos sistemas de gerenciamento de alarmes.

Diante desse contexto, fica claro que o exame dos eventos e alarmes de uma planta não deve ficar restrita ao que é apresentado nas IHMs dos sistemas de gerenciamento ou à inspeção visual desse grande volume de registros por parte dos especialistas. É oportuno, portanto, que esses registros sejam melhor aproveitados sob a égide das atuais práticas em Ciências de Dados. Entre estas práticas está a AED, metodologia de essencial importância no estágios iniciais de análise em busca de informações importantes.

4 Análise Exploratória de Dados de Eventos e Alarmes

Os registros de alarmes e eventos compõem uma enorme e não processada massa de dados, que pode encobrir informações valiosas, possivelmente não contempladas nos projetos das IHMs de sistemas de gerenciamento de alarmes ou simplesmente passadas despercebidas pelo crivo dos operadores. O formato tabular dos registros é, por sua vez, pouco adequado à cognição humana, o que dificulta a análise visual. Para que essa potencial fonte de informação não seja negligenciada, propõe-se a aplicação de técnicas de AED como abordagem inicial e mínima para a análise desses dados.

A AED é uma abordagem de análise estritamente orientada a dados. Utiliza um conjunto de técnicas destinadas à investigação de dados na busca por informações e relacionamentos interessantes, a partir de uma abordagem ativamente incisiva, com ênfase real na descoberta do inesperado (Jones, 1987). A AED isola padrões e características dos dados e os revela vigorosamente ao analista (Jones, 1987), sem que haja a necessidade de conhecimento prévio ou de que existam hipóteses pré-especificadas, ou seja, sem exigir do analista questionamentos antecipados e bem definidos a serem direcionados aos dados (Mao, 2003).

Um componente-chave no AED é a utilização de diferentes técnicas gráficas para a conveniente apresentação de seus resultados, integrando a aguçada percepção humana com o poder e versatilidade computacional atualmente disponíveis. Na prática, o AED mescla as técnicas quantitativas da análise clássica com técnicas e ferramentas da análise gráfica (Croarkin and Tobias, 2012), com intuito de fornecer uma apresentação apropriada do conhecimento obtido aos especialistas.

A AED é uma abordagem com ampla variedade de técnicas e metodologias, e, nesse trabalho, apenas algumas etapas serão explicitadas com foco na demonstração visual de seus resultados.

4.1 Dados Utilizados

A Tabela 1 representa uma amostra de um *dataframe* analisado, construído a partir dos registros de alarmes e eventos relativos à infraestrutura de uma planta petroquímica. Algumas colunas foram omitidas em virtude da insuficiência de espaço. O *dataframe* consta de 1.020.765 entradas correspondentes aos eventos e alarmes registrados no intervalo de cerca de 3 dias. Nesse caso, cada entrada corresponde a uma observação das múltiplas variáveis categóricas (colunas) que definem os eventos. Já a Tabela 2 apresenta um sumário do conjunto de dados analisados, apontando, para cada variável, a contagem de ocorrências e singularidades, assim como as ocorrências mais comuns

Date_Time	Event_Type	Category	Node	Module	State	Event_Level
2018-02-22 01:41:50	EVENT	PROCESS	CTRL_URC_009	TI-3283737_S	ACTIVE	4-INFO
2018-02-22 12:04:54	ALARM	PROCESS	CTRL_URC_005	DI-1230103D	ACT/UNACK	15-CRITICAL
2018-02-21 08:14:04	ALARM	SYSTEM	CTRL_URC_004	VE-C-123003-01B	ACT/UNACK	15-CRITICAL
2018-02-23 09:27:48	ALARM	INSTRUMENT	CTRL_URC_009	TI-3283735_S	ACT/UNACK	07-ADVISORY
2018-02-22 12:04:22	EVENT	PROCESS	CTRL_URC_604	HIC-42834001SP	INACT/ACK	NaN
2018-02-23 01:06:52	EVENT	PROCESS	CTRL_URC_604	HIC-42834001SP	ACT/ACK	NaN
2018-02-23 08:23:00	EVENT	PROCESS	CTRL_URC_009	PDI-3283725_S	ACTIVE	4-INFO
2018-02-23 08:45:08	EVENT	PROCESS	CTRL_URC_009	XI-3283726_S	ACTIVE	4-INFO
2018-02-22 01:56:24	ALARM	INSTRUMENT	CTRL_URC_009	TI-3283722_S	ACT/UNACK	07-ADVISORY
2018-02-23 02:13:12	EVENT	PROCESS	CTRL_URC_614	SMB424102B-35-02	ACTIVE	4-INFO

Tabela 1: Exemplo de *dataframe* composto pelos registros de alarmes e eventos.

feature	count	unique	top	freq
Event_Type	1020765	5	EVENT	598503
Category	949319	6	PROCESS	643300
Area	1020765	75	U1230_HVAC	278450
Node	1020765	46	CTRL_URC_009	283645
Module	1014842	1280	VE-C-123003-01B	72679
Module_Desc	1014665	695	C-123003 GAS DET	189004
Attribute	620352	115	MODULE_ALM	189709
State	1013222	16	ACTIVE	399063
Event_Level	823989	9	4-INFO	400152
Desc1	1020429	1126	MODBAD	189709
Desc2	948636	10326	I/O Input Failure	273377
IsArchived	1020765	2		842692

Tabela 2: Sumário do *dataframe*.

nas variáveis e suas frequências.

4.2 Análise da Qualidade dos Dados

Uma importante etapa no EDA é a análise da qualidade dos dados, que contempla a investigação acerca de eventuais dados faltantes no conjunto em análise. Com a contabilização e ponderação dessas ausências nos registros, cria-se uma consciência situacional acerca dos dados faltantes, que pode ser aliada na busca por problemas de configuração nos dispositivos, falha humana ou problemas de comunicação. Essa consciência também orienta a estratégia de tratamento dessas ausências (desconsideração ou imputação) nas etapas de análise seguintes.

A Figura 1 exibe um sumário da análise dos dados ausentes através de uma matriz de nulidade, de onde é possível perceber a dispersão dos dados faltantes em cada coluna, e de um gráfico de barras que demonstra a contagem e percentual das ausências em cada variável. A partir do gráfico, percebe-se que os atributos *Attribute* e *Event_Level* possuem mais dados faltantes. Na análise, os dados ausentes foram apenas desconsiderados e a coluna de pior qualidade, *Attribute*, não foi utilizada.

4.3 Análise Quantitativa

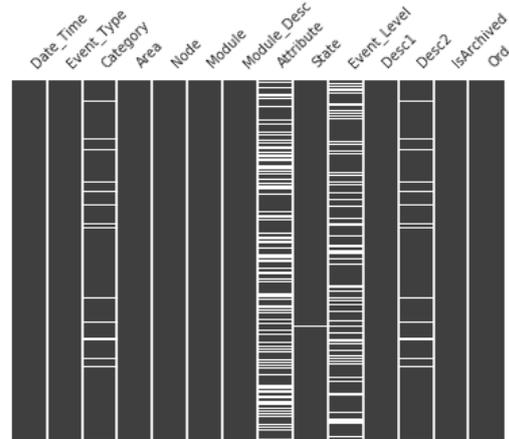
É notável que o conjunto de dados mostrado na Tabela 1 é essencialmente categórico. Uma abordagem natural nesses casos é realizar uma apuração quantitativa dos elementos dessas categorias, demonstrando os resultados em um formato de visualização conveniente.

Como exemplo dessa abordagem, a Figura 2 fornece, por meio de gráficos de barra (em diferentes formatos), uma contabilização das entradas do *dataframe*, evidenciando (a) as relações quantitativas entre os tipos dos eventos (coluna

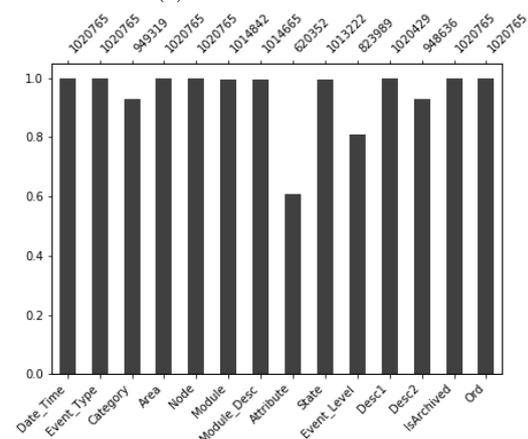
Event_Type) e as categorias dos eventos (coluna *Category*) e (b) a contabilização dos níveis dos eventos (coluna *Event_Level*) das entradas em associação ao turno do dia, coluna suplementar criada a partir da marcação de tempo disponível nos registros. A partir desses gráficos é possível constatar, por exemplo, que a maioria dos registros de eventos são da categoria “PROCESS” ou que a maioria das entradas de nível crítico ocorrem pelo período da manhã.

4.4 Análise Categórica

As variáveis categóricas apontam os vários grupos a que um indivíduo (entrada) do conjunto de dados em análise pode pertencer e, por isso, podem mascarar informações interessantes. Uma

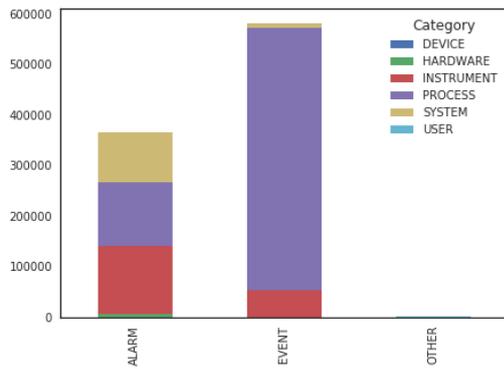


(a) Matriz de nulidade.

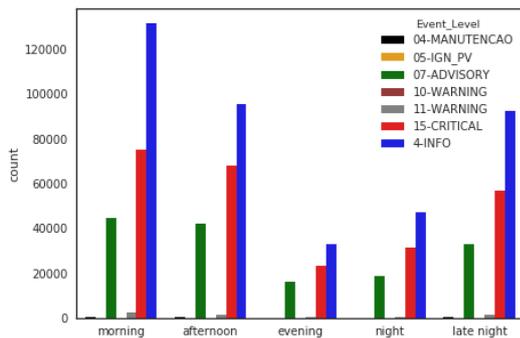


(b) Contabilização da nulidade por coluna.

Figura 1: Visão geral dos dados ausentes.



(a) Tipos de entradas e categorias.



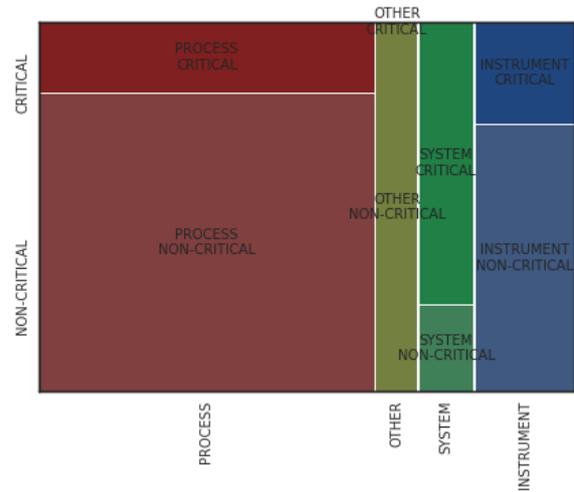
(b) Nível dos eventos com relação aos turnos do dia.

Figura 2: Análise quantitativa expressa em gráficos de barra.

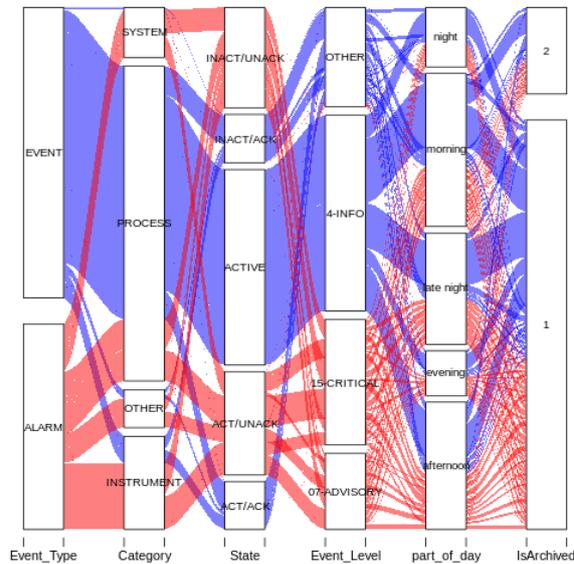
análise sob dados categóricos tem enfoque em realçar os eventuais relacionamentos entre as categorias através de diagramas.

O **diagrama mosaico** (SAS, 2013), utilizado para evidenciar graficamente a relação proporcional entre as colunas *Category* e *Event_Level*, é mostrado na Figura 3 (a). Neste caso, procedeu-se o agrupamento de algumas categorias dessas colunas em favor de uma melhor visualização. Nesse diagrama, a título de exemplo, fica evidente que cerca de 75% das entradas da categoria “SYSTEM” tem nível crítico e a maioria das entradas da categoria “PROCESS” tem nível não-crítico.

Outra forma conveniente de visualização de dados categóricos é o **diagrama aluvial** (Michal Bojanowski and Edwards, 2016). Este diagrama propicia a visualização de relações proporcionais entre múltiplas variáveis categóricas ao mesmo tempo, fornecendo uma ideia de fluxo e dispersão qualitativa entre as entradas e suas diversas características. A Figura 3 (b) mostra o diagrama aluvial construído para 5 colunas mais representativas do conjunto de dados originais adicionadas da coluna de suplementar de turno do dia, onde as faixas em azul destacam os eventos enquanto que as vermelhas, os alarmes. A partir deste diagrama, é fácil reparar, por exemplo, a relação entre as entradas do tipo “ALARM”, a categoria “PROCESS”, o estado “ACT/UNACK” e o nível



(a) Diagrama do tipo mosaico.



(b) Diagrama do tipo aluvial.

Figura 3: Resultados de análise categórica expressa em diagramas mosaico e aluvial.

“15-CRITICAL”, ou ainda, perceber que a maioria das entradas do tipo “EVENT” são da categoria “PROCESS”, possuem estado “ACTIVE”, nível “INFO” e ocorrem entre os turnos da madrugada e da manhã.

Como mostrado na Tabela 2, algumas colunas possuem elevada unicidade, o que dificulta o processo de visualização em gráficos ou diagramas, assim como prejudica a confrontação com as demais. Contudo, para alguns desses casos, pode ser desejável conhecer as frequências das ocorrências em uma coluna, o que pode ser alcançado pela **nuvem de palavras**. Esta técnica de mineração de texto procura evidenciar a frequência das palavras (sentenças) em um texto, organizando hierarquicamente em um plano, por meio do tamanho e cor da fonte, as sentenças mais frequentes.

A técnica foi utilizada para evidenciar as ocorrências mais frequentes na coluna *Desc2*, para entradas de **alarmes** com nível (*Event_Level*) **crítico**. A coluna *Desc2* refere-se a descrição textual

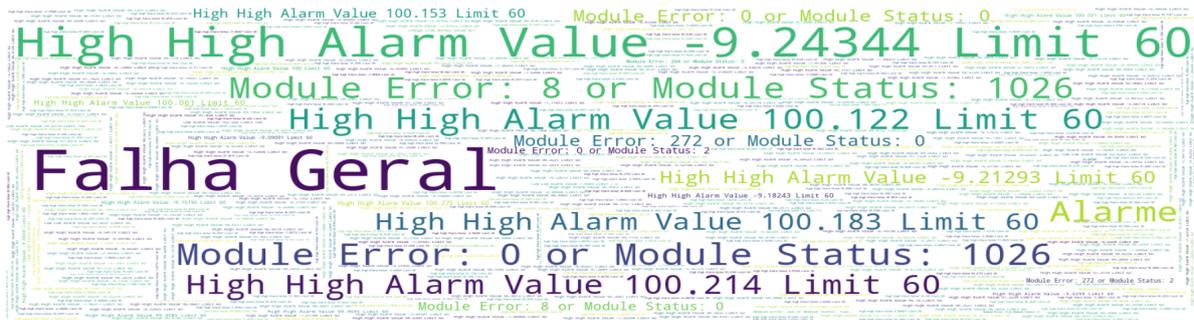


Figura 4: Nuvem de palavras referente à coluna *Desc2* para alarmes críticos.

mais específica das entradas, assim, inspecionando a Figura 4, é possível identificar o teor dos alarmes críticos mais comuns na planta.

5 Infraestrutura de *Big Data* Industrial

The EDA excerpt presented in this paper is A análise apresentada foi realizada em um cenário laboratorial computacionalmente limitado e *offline*, utilizando dados cedidos de apenas uma única unidade, no total de dez, de um processo industrial. Em toda a planta, aproximadamente 3,3 milhões de registros são produzidos diariamente, perfazendo aproximadamente 650 *MegaBytes* em dados brutos por dia de operação.

Para que a AED e outras técnicas de análise mais sofisticadas sejam factíveis e viáveis num cenário real de produção, com uma quantidade muito grande de dados, é necessário o estabelecimento de uma infraestrutura de *hardware* e *software* de suporte capaz de atender a grande demanda computacional exigida pela nova abordagem orientada a dados.

Como discutido em (Obitko et al., 2013), a infraestrutura computacional industrial corrente, no tocante ao gerenciamento e processamento dos dados da planta, concentra-se, sobretudo, na coleta, seleção e armazenamento de dados sob cadências adequadas, preservando as séries históricas da planta em um repositório de acesso sob demanda. Logo, quaisquer processamento adicional, análises ou consultas mais avançadas estão além da capacidade da infraestrutura computacional típica. Faz-se necessário, portanto, um ajustamento da infraestrutura computacional industrial em direção aos preceitos do *Big Data*, que requer, entre outras características, um elevado poder computacional, a habilidade de processamento paralelo *offline* e *online* de grandes quantidades de dados, o armazenamento massivo de dados em um sistema de arquivos distribuído com acesso seguro, uma infraestrutura de comunicação inteligente, rápida e robusta, assim como uma sistematização de análise lastreada no conhecimento específico da planta. (Obitko et al., 2013; Basanta-Val, 2017; Lee et al., 2015).

6 Conclusões e Trabalhos Futuros

O artigo empenhou-se em realizar uma contextualização em torno dos conceitos de Ciência de Dados e *Big Data* no cenário industrial, introduzindo a possibilidade do uso extensivo do AED, componente fundamental em uma empreitada de análise de dados, na inspeção de registros de alarmes e eventos industriais.

No ciclo clássico de análise científica de dados, após a preparação e formatação dos dados, em geral, o AED é adotado como uma primeira frente de análise de dados, imprimindo uma abordagem mais panorâmica sobre os dados e indicando relacionamento mais gerais. Sem fugir à prática, o AED foi utilizado como uma abordagem inicial, porém versátil e promissora, que aliou métodos quantitativos, qualitativos e visuais para obtenção, sem qualquer suposição prévia, de informações de elevado valor sobre os dados analisados. Os gráficos e diagramas apresentados, os quais representam apenas amostras da potencialidades do AED, exprimem uma soma de informações até então ocultadas, de notável importância para os especialistas da área.

O escopo do AED é amplo e contempla também análises estatísticas, que possibilitarão, por exemplo, o descobrimento de correlações, relações de causa-efeito e dependência entre os registros.

Após a etapa de AED, almeja-se adotar estratégias de análise de dados mais específicas, que em geral, utilizam técnicas de aprendizagem de máquina no agrupamento dos dados de alarmes e eventos e em processos de diagnóstico, prognóstico e predições de situações anormais de operação de processos industriais.

Referências

- Academies, N. (2017). *Strengthening Data Science Methods for Department of Defense Personnel and Readiness Missions*, The National Academies Press, Washington, DC.
- Allen-Bradley (2017). Abb ability manufacturing operations management - process and production intelligence, *Technical report*.

- ANSI/ISA (2016). Management of alarm systems for the process industries, *ISA ANSI/ISA-18.2-2016*, ISA.
- Basanta-Val, P. (2017). An efficient industrial big-data engine, *IEEE Transactions on Industrial Informatics* **PP**(99): 1–1.
- Bryant, R., Katz, R. H. and Lazowska, E. D. (2008). Big-data computing: creating revolutionary breakthroughs in commerce, science and society.
- Croarkin, C. and Tobias, P. (2012). Nist/sematech e-handbook of statistical methods.
- da Silva, M. J., Pereira, C. E. and Götz, M. (2016). A dynamic approach for industrial alarm systems, *2016 International Conference on Computer, Information and Telecommunication Systems (CITS)*, pp. 1–5.
- EEMUA (2007). *Alarm Systems: A Guide to Design, Management and Procurement*, EEMUA Publication Series, E E M U A (Engineering Equipment & Materials Users Association).
- Habibi, E. and Hollifield, B. (2006). Alarm Systems Greatly Affect Offshore Facilities Amid High Oil Prices, *World Oil Magazine* **227**(9).
- Jones, L. (1987). *The Collected Works of John W. Tukey: Philosophy and Principles of Data Analysis 1965-1986*, number v. 4, Taylor & Francis.
- Lee, J., Bagheri, B. and Kao, H.-A. (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems, *Manufacturing Letters* **3**: 18–23.
- Lee, J., Kao, H.-A. and Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment, *Procedia Cirp* **16**: 3–8.
- Mao, Y. (2003). *Data Visualization in Exploratory Data Analysis: an Overview of Methods and Technologies*, Master's thesis, The University of Texas at Arlington, USA.
- Michal Bojanowski, R. and Edwards (2016). *Package 'alluvial'*.
- Obitko, M., Jirkovský, V. and Bezdíček, J. (2013). Big data challenges in industrial automation, in V. Mařík, J. L. M. Lastra and P. Skobelev (eds), *Industrial Applications of Holonic and Multi-Agent Systems*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 305–316.
- O'Brien, L. and Woll, D. (2010). Alarm management strategies, *Technical report*, ARC Strategies.
- Reis, M. S. and Gins, G. (2017). Industrial process monitoring in the big data/industry 4.0 era: from detection, to diagnosis, to prognosis, *Processes* **5**(3).
- SAS, I. (2013). *JMP 11 Basic Analysis*, SAS Institute.
- Wang, J., Yang, F., Chen, T. and Shah, S. L. (2016). An overview of industrial alarm systems: Main causes for alarm overloading, research status, and open problems, *IEEE Transactions on Automation Science and Engineering* **13**(2): 1045–1061.