

**UNIVERSIDADE DE CAXIAS DO SUL
CURSO DE ENGENHARIA DE PRODUÇÃO**

CHARLES DA LUZ POLA

**APLICAÇÃO DE PROCESSO DE CLASSIFICAÇÃO E TÉCNICA DE BAYES NA
BASE DE DADOS DE ACIDENTES OCUPACIONAIS DE UMA EMPRESA
METALÚRGICA**

CAXIAS DO SUL

2018

CHARLES DA LUZ POLA

**APLICAÇÃO DE PROCESSO DE CLASSIFICAÇÃO E TÉCNICA DE BAYES NA
BASE DE DADOS DE ACIDENTES OCUPACIONAIS DE UMA EMPRESA
METALÚRGICA**

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia de Produção da Universidade de Caxias do Sul, como requisito parcial à obtenção do grau de Bacharel em Engenharia de Produção.

Orientador Prof. Dr. Enor Jose Tonolli Junior

CAXIAS DO SUL

2018

À minha mãe, Ana, por seu apoio e amor incondicional.

*“Não importa quantos recursos você possui,
se não souber como utilizá-los, eles nunca
serão o suficiente.”*

Stephen Kyeyune

RESUMO

O barateamento, miniaturização e o aumento de capacidade de processamento levaram à disseminação de equipamentos, dispositivos e processos capazes de produzir e armazenar volumes maiores de dados. No mundo empresarial e na sociedade, os modelos de negócios estão aumentando progressivamente a utilização de conhecimentos analíticos em seus processos. De mesma forma, a massa crescente de informações possibilita a humanidade novas oportunidades de se estudar ou endereçar problemas antes muito complexos, como a previsão de acontecimentos baseada em informações disponíveis. Diante desta realidade o trabalho demonstra a aplicação do método de processo padrão inter-indústrias para mineração de dados (CRISP-DM), utilizando a técnica de Bayes por meio do algoritmo NaiveBayes no *software* R para a criação de um modelo preditivo de acidentes em uma empresa metalúrgica de Caxias do Sul. O projeto utiliza conceitos de ciência de dados em análises implícitas e técnicas de classificação, passando-se pela produção, armazenamento, transformação e análise de registros. Os resultados obtidos foram avaliados por meio de uma matriz de confusão e juntamente com o algoritmo criado e a caracterização do perfil de acidentados da empresa por meio de critérios e suas probabilidades, foram entregues para gestão da companhia. A técnica se mostrou válida e aberta a aplicações futuras na previsão de comportamentos baseada em registros históricos da companhia.

Palavras-Chave: Ciência de Dados. Classificação. Bayes. NaiveBayes. Predição. Acidentes.

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 – Revoluções industriais | 10 |
| Figura 2 – Ciclo de vida do dado..... | 16 |
| Figura 3 – Construção de um <i>data warehouse</i> | 20 |
| Figura 4 – <i>Map e Reduce</i> | 21 |
| Figura 5 – Atributos, instância e classes..... | 23 |
| Figura 6 – Modelo K-means | 26 |
| Figura 7 – Modelo DBSCAN | 26 |
| Figura 8 – Modelo <i>dashboard</i> | 29 |
| Figura 9 – Pirâmide de orientação decisória | 30 |
| Figura 10 – Gráfico de atendimentos no centro de saúde..... | 31 |
| Figura 11 – Gráfico de acidentes sem perda de tempo..... | 32 |
| Figura 12 – Gráfico de acidentes com perda de tempo | 32 |
| Figura 13 – Fluxograma de proposta de trabalho | 33 |
| Figura 14 – Resposta <i>software R</i> | 45 |
| Figura 15 – <i>Cluster</i> considerado..... | 45 |
| Figura 16 – <i>A-priori probabilities</i> NaiveBayes | 46 |
| Figura 17 – <i>Conditional probabilities</i> NaiveBayes | 47 |
| Figura 18 – Retorno matriz de confusão | 48 |
| Figura 19 – Gráfico probabilidades condicionais: chegada | 48 |
| Figura 20 – Gráfico probabilidades condicionais: refeição..... | 49 |
| Figura 21 – Gráfico probabilidades condicionais: refeição extra..... | 49 |
| Figura 22 – Gráfico probabilidades condicionais: atendimentos centro de saúde | 50 |
| Figura 23 – Gráfico probabilidades condicionais: tempo de empresa..... | 50 |
| Figura 24 – Gráfico probabilidades condicionais: idade | 51 |

LISTA DE QUADROS

| | |
|--|----|
| Quadro 1 – <i>Drivers</i> da quarta revolução industrial | 14 |
| Quadro 2 – Armazenamento de dados..... | 19 |
| Quadro 3 – Exemplo de matriz de confusão..... | 24 |
| Quadro 4 – Processos, técnicas e algoritmos..... | 24 |
| Quadro 5 – Tipos de elementos gráficos adequados | 28 |
| Quadro 6 – Reunião de formulação de critérios | 39 |
| Quadro 7 – Critérios | 39 |
| Quadro 8 – Volume de dados | 42 |
| Quadro 9 – Registros mensais | 42 |
| Quadro 10 – Variações critérios (atributos) | 43 |
| Quadro 11 – Reunião de apresentação modelo | 52 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|----------|---|
| ACPT | Acidentes Com Perda de Tempo |
| ASPT | Acidentes Sem Perda de Tempo |
| AT | Atendimentos Ocupacionais |
| CRISP-DM | <i>Cross Industry Standard Process for Data Mining</i> / Processo Padrão Inter-Indústrias para Mineração de Dados |
| CSV | <i>Comma-Separated Values</i> / Valores Separados por Vírgula |
| DW | <i>Data Warehouse</i> / Armazém de Dados |
| ERP | <i>Enterprise Resource Planning</i> / Planejamento de Recurso Corporativo |
| ETL | <i>Extract, Transform and Load</i> / Extração, Transformação e Carga |
| GPS | <i>Global Positioning System</i> / Sistem de Posicionamento Global |
| HP | <i>Hewlett-Packard</i> |
| IoT | <i>Internet of Things</i> / Internet das Coisas |
| NoSQL | <i>Not Only SQL</i> / Não Somente SQL |
| OLAP | <i>Online Analytical Processing</i> / Processamento Analítico em Tempo Real |
| OLTP | <i>Online Transaction Processing</i> / Processamento de Transações em Tempo Real |
| RAM | <i>Random Access Memory</i> / Memória de Acesso Aleatório |
| RNAs | Redes Neurais Artificiais |
| SEMMA | <i>Sample, Explore, Modify, Model and Assess</i> / Amostriar, Explorar, Modificar, Modelar e Avaliar) |
| SQL | <i>Structured Query Language</i> / Linguagem de Consulta Estruturada |
| SSM | Segurança, Saúde e Meio Ambiente |
| TI | Tecnologia da Informação |
| UTI | Unidade de Tratamento Intensivo |
| Weka | <i>Waikato Environment for Knowledge Analysis</i> / Ambiente para Análise de Conhecimento <i>Waikato</i> |

SUMÁRIO

| | | |
|-------|---|-----------|
| 1 | INTRODUÇÃO | 10 |
| 1.1 | JUSTIFICATIVA | 11 |
| 1.2 | OBJETIVOS | 12 |
| 1.2.1 | Objetivo geral | 12 |
| 1.2.2 | Objetivos específicos | 12 |
| 1.3 | ABORDAGEM E DELIMITAÇÃO DO TRABALHO | 13 |
| 2 | FUNDAMENTAÇÃO TEÓRICA..... | 14 |
| 2.1 | INDÚSTRIA 4.0..... | 14 |
| 2.2 | CICLO DE VIDA DO DADO | 16 |
| 2.2.1 | Produção | 17 |
| 2.2.2 | Armazenamento | 18 |
| 2.2.3 | Análise | 21 |
| 2.3 | VISUALIZAÇÃO DE DADOS | 27 |
| 3 | PROPOSTA DE TRABALHO | 30 |
| 3.1 | CENÁRIO ATUAL | 30 |
| 3.2 | PROPOSTA DE TRABALHO | 33 |
| 3.2.1 | Entendimento do negócio | 34 |
| 3.2.2 | Entendimento dos dados | 34 |
| 3.2.3 | Preparação dos dados | 35 |
| 3.2.4 | Modelagem | 35 |
| 3.2.5 | Avaliação | 35 |
| 3.2.6 | Implementação | 36 |
| 4 | IMPLEMENTAÇÃO..... | 37 |
| 4.1 | ENTENDIMENTO DO NEGÓCIO | 37 |
| 4.2 | ENTENDIMENTO DOS DADOS | 38 |
| 4.3 | PREPARAÇÃO DOS DADOS | 40 |
| 4.3.1 | Registros de acidentes em <i>Microsoft Excel</i> | 41 |
| 4.3.2 | Registros de atendimentos ao centro de saúde | 41 |
| 4.3.3 | Registros de acessos no <i>Ponto Soft</i> | 41 |

| | | |
|--------------|--|-----------|
| 4.4 | MODELAGEM | 43 |
| 4.4.1 | Classificação: NaiveBayes | 43 |
| 4.4.2 | Seleção de atributos | 44 |
| 4.4.3 | Modelagem NaiveBayes..... | 46 |
| 4.5 | AVALIAÇÃO..... | 47 |
| 4.6 | APRESENTAÇÃO..... | 51 |
| 5 | CONCLUSÃO | 53 |
| | REFERÊNCIAS | 55 |
| | APÊNDICE A – REGISTROS ACPT RANON IMPLEMENTOS CAXIAS DO SUL | 57 |
| | APÊNDICE B – REGISTRO ATENDIMENTOS CENTRO DE SAÚDE | 59 |
| | ANEXO A – ALERTA DE SEGURANÇA | 60 |

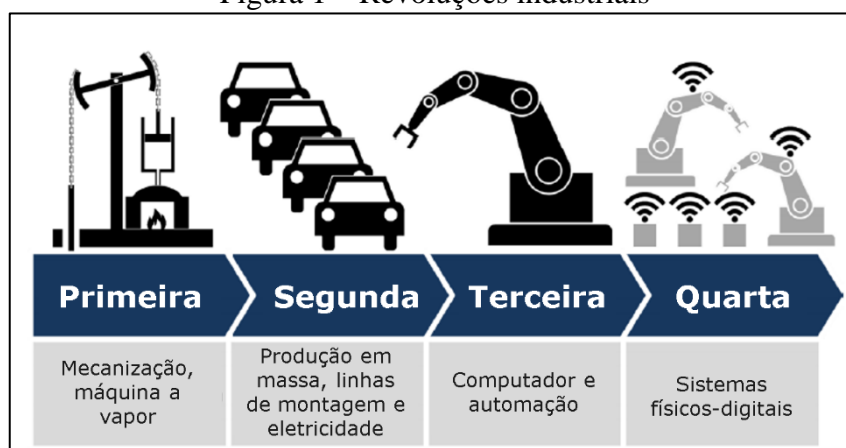
1 INTRODUÇÃO

Dos diversos e fascinantes desafios que a sociedade enfrenta atualmente, o mais intenso e importante de se entender é a revolução tecnológica. As novas possibilidades e inovações estão mudando fundamentalmente o modo como o ser humano vive, trabalha e se relaciona (SCHWAB, 2017).

No ambiente industrial, as empresas estão vivenciando uma união entre tecnologias digitais e o ambiente tradicional. Segundo Stãncioiu (2017), existe uma divisão de opiniões entre o uso do termo revolução ou evolução para apresentar os acontecimentos evidenciados. Nos Estados Unidos da América o movimento é chamado de *Smart Manufacturing*, na China se discute o *Made in China 2025* e no Japão tem-se o *Innovation 25*.

Historicamente, a primeira revolução industrial aconteceu entre os anos de 1760 e 1840, sendo caracterizada pela invenção da máquina a vapor e da possibilidade de mecanização da produção. No final do século XIX teve início a segunda revolução, esta foi reconhecida pela produção em massa, linhas de montagem e o advento da eletricidade. Já em 1960 a terceira revolução industrial foi caracterizada pela concepção do computador e da automação. Atualmente vivencia-se a quarta revolução, diferenciada pela onipresença de informações por meio de dispositivos móveis, por sensores e processadores, pela inteligência artificial e *machine learning* (SCHWAB, 2017). A Figura 1 apresenta de modo ilustrativo a evolução da manufatura ao longo do tempo.

Figura 1 – Revoluções industriais



Fonte: Adaptado de Stãncioiu (2017).

Segundo Amaral (2016), um dos protagonistas desta nova revolução é o dado. O barateamento, miniaturização e o aumento de capacidade de processamento levaram à disseminação de equipamentos, dispositivos e processos capazes de produzir e armazenar

volumes cada vez maiores de dados. Hoje, dados são produzidos massivamente em redes sociais, sensores, câmeras, leitores, identificadores, satélites, computadores, celulares inteligentes, *tablets*, carros, entre outros.

Para exemplificar o fenômeno *Big Data*, Amaral (2016) elencou o volume de algumas informações movimentadas na rede mundial de computadores: quase três milhões de e-mails enviados por segundo, cem horas de vídeos são carregados no *YouTube* por minuto, quinhentos milhões de *tweets* por dia e cem terabytes de dados postados diariamente no *Facebook*. Um *smartphone* possui maior capacidade que o melhor computador de 1985, seis bilhões de pessoas utilizam celulares e mais de 1,7 bilhões estão usando redes sociais. Totalizando mais de 2,3 trilhões de gigabytes de dados criados em um único dia.

Portanto, o fenômeno de dados é evidente. Nessa nova realidade, cabe às companhias observarem as possibilidades e utilizarem a vantagem analítica destes dados para incrementar sua produtividade. Esse trabalho aborda esta temática em um ambiente caracterizado pela oportunidade principal de transição entre o modelo de ações corretivas para a predição de comportamento, gerando conhecimento e atendendo as exigências do negócio.

1.1 JUSTIFICATIVA

A massa de dados em crescente volume, padronização de apresentação e organização, derivada das tecnologias envolvidas na quarta revolução industrial, traz para a humanidade novas oportunidades de se estudar ou endereçar problemas antes muito complexos (FREDERICK, 2016). Na indústria, por exemplo, pode-se utilizar o fenômeno *Big Data* para responder questões relacionadas aos custos de produção, tempo de processamento, estratégias de negócios, otimização de ofertas, tomada de decisões, identificação de falhas e defeitos; na educação é possível identificar o progresso dos estudantes, bem como implementar sistemas efetivos de avaliação e suporte; no governo, para gerenciamento de recursos, tráfego de veículos ou na prevenção de crimes; na saúde, possibilita o gerenciamento do histórico de saúde dos pacientes, planos de prevenção a doenças e prescrição de medicamentos (MUTULA, 2016).

Negligenciar ou não aproveitar tamanha oportunidade de gerar e obter informação não pode ser considerada uma opção, pois em uma época em que empresas oferecem produtos similares e utilizam tecnologias comparáveis, as bases anteriores de competição não estão mais disponíveis. Para se manter competitiva, a empresa deve ter máxima eficiência e eficácia em seus processos, tomando decisões de negócio baseadas em inteligência analítica e informações disponíveis (DAVENPORT; HARRIS, 2007).

Segundo Amaral (2016), o fenômeno dos dados traz boas e más notícias para o mundo empresarial, como positivo tem-se as diversas oportunidades e possibilidades que surgirão, como por exemplo a previsão de acontecimentos futuros baseados em informações históricas. E como negativo, companhias que não utilizarem *Big Data* irão desaparecer, engolidas pelas concorrentes que serão mais eficientes, apresentando menores custos, produtos de qualidade superior e clientes satisfeitos.

Entendendo esta realidade, o trabalho buscou aplicar conceitos de ciência de dados na observação da capacidade informacional do conjunto de dados gerados em uma empresa metalúrgica de Caxias do Sul, sendo evidente a necessidade e oportunidade de se absorver o grande volume de informações de seus processos, gerando conhecimento competitivo para a gestão da companhia.

1.2 OBJETIVOS

Visto a corrente mundial em análise de dados e as possibilidades em se aproveitar toda informação gerada em seus processos para melhorar a segurança ocupacional de seus funcionários, a empresa entende como necessário o avanço de suas tecnologias analíticas e a utilização de modelos preditivos, ultrapassando-se a barreira da correção e atuando na prevenção de acidentes. Desta forma, deve-se relacionar todos dados disponíveis com o intuito de se extrair conhecimento, gerando retorno financeiro e organizacional para a companhia.

1.2.1 Objetivo geral

O objetivo do trabalho é a aplicação de um algoritmo de predição de ocorrências de acidentes por meio da mineração de dados.

1.2.2 Objetivos específicos

Do objetivo geral derivaram-se os seguintes objetivos específicos:

- a) criar um *cluster* de causas de acidentes de trabalho;
- b) buscar histórico de informações nos bancos de dados relacionados às possíveis causas informadas;
- c) analisar o volume de informações utilizando o processo de classificação e a técnicas de Bayes;

- d) avaliar o algoritmo de classificação resultante por meio de matriz de confusão, identificando sua precisão e taxa de erro;
- e) validar junto a gestão, a técnica desenvolvida.

1.3 ABORDAGEM E DELIMITAÇÃO DO TRABALHO

O presente trabalho possui uma abordagem quantitativa descritiva realizada por meio de procedimento *ex-post-facto*. Segundo Gil (2008, p. 54) “pode-se definir pesquisa *ex-post-facto* como uma investigação sistemática e empírica na qual o pesquisador não tem controle direto sobre as variáveis independentes, porque já ocorreram suas manifestações ou porque são intrinsecamente não manipuláveis”. Pesquisas descritivas “têm como objetivo primordial a descrição das características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis” (GIL, 2008, p. 28).

O estudo limitou-se na análise do banco de dados da Randon S/A Implementos e Participações e não contemplou informações externas ao ambiente corporativo, como redes sociais, dispositivos móveis ou histórico de doenças clínicas dos funcionários.

De mesma forma, não observou a etapa de implementação do algoritmo dentro da companhia, visto o impacto financeiro e de recursos humanos para centralização de bancos de dados e criação de relações entre *softwares* distintos. O trabalho, teve como entrega o algoritmo de predição de acidentes validado, bem como sua precisão e taxa de erro avaliadas no estudo.

O trabalho foi desenvolvido na célula de sistemas de manufatura da engenharia de processos, com o auxílio da área de saúde, segurança e meio ambiente da companhia.

2 FUNDAMENTAÇÃO TEÓRICA

A revisão bibliográfica proposta no presente trabalho buscou descrever modelos, técnicas e conhecimentos em mineração de dados para dar suporte à criação do algoritmo de predição de ocorrências de acidentes na empresa Randon S/A Implementos e Participações.

2.1 INDÚSTRIA 4.0

Diferentemente das evoluções de manufatura verificadas anteriormente, a quarta revolução industrial se baseia em rupturas exponenciais, modelos dirigidos por avanços tecnológicos que permitem novos e melhores desenvolvimentos digitais. Além da velocidade, o envolvimento e a integração de diversas tecnologias em áreas distintas clarificam o movimento. Esta plataforma apresenta tamanho impacto que está presente em todo ciclo de vida do produto, alterando a forma de se produzir, consumir, transportar e entregar soluções (SCHWAB, 2017).

Esta revolução não está limitada à manufatura, mas se manifesta em todos aspectos da sociedade, influenciando todos os campos da vida humana. Os desenvolvimentos digitais, físicos e biológicos caracterizam as três principais tecnologias que direcionam o movimento, englobando a *Internet of Things* (IoT), inteligência artificial e *machine learning*, *big data* e *cloud computing*, carros autônomos, impressoras 3D, engenharia genética e neurotecnologias, como apresentado no Quadro 1 (LI; HOU; WU, 2017).

Quadro 1 – *Drivers* da quarta revolução industrial

| Tecnologias <i>drivers</i> | Campos |
|----------------------------|--|
| Digital | A Internet das Coisas (IoT) Inteligência artificial e <i>machine learning</i> <i>Big data</i> e <i>cloud computing</i> Plataformas digitais |
| Físico | Carros autônomos Impressora 3D |
| Biológico | Engenharia genética Neurotecnologias |

Fonte: Adaptado de Li, Hou e Wu (2017).

Segundo Schwab (2017) essa velocidade de inovação e difusão é mais rápida do que nunca, novos negócios que revolucionaram antigos paradigmas, como *Airbnb* e *Uber*, eram relativamente desconhecidos a poucos anos atrás. O primeiro iPhone, por exemplo, foi lançado

em 2007 e no final de 2015 mais de 2 bilhões de *smartphones* já eram utilizados. Um simples aparelho, como um *tablet*, possui o poder de processamento de 5000 computadores de trinta anos atrás, enquanto 1 gigabyte de armazenamento custa menos de 0,03 dólares, comparado com mais de 10.000 dólares há duas décadas.

As novas possibilidades e plataformas digitais, que são de fácil acesso via telefones inteligentes, conectam pessoas, recursos e dados, criando novas maneiras de se consumir bens e serviços. Como exemplo, o aplicativo *Uber*, sendo a maior empresa de transporte do mundo sem possuir nenhum veículo, o *Facebook* como a maior plataforma de propaganda sem criar nenhum conteúdo, o *Alibaba*, maior empresa de vendas sem possuir estoque e o *Airbnb* a maior empresa de acomodação sem possuir nenhum imóvel (LI; HOU; WU, 2017).

Para a manufatura, conforme esclarece Stãncioiu (2017), esta grande revolução industrial depende do avanço de pequenas tecnologias em diversos campos, como:

- a) aplicação de tecnologias de comunicação e informação para digitalizar e integrar sistemas de projeto, desenvolvimento, manufatura e uso de produtos;
- b) novos sistemas para modelamento, simulação e virtualização da manufatura;
- c) desenvolvimento de sistemas físicos-virtuais para monitorar e controlar processos reais;
- d) a evolução de impressoras 3D e manufatura aditiva para simplificar a fabricação;
- e) suporte de decisões de operadores utilizando realidade aumentada.

Uma das principais pontes entre o mundo real e virtual proposto pela Indústria 4.0 é a IoT, sendo esta a aplicação de sensores e inúmeros outros meios de conectar o mundo físico às redes virtuais. Este volume de dispositivos conectados à internet vem sendo incrementado drasticamente ao longo dos anos, gerando uma quantidade sem precedentes de dados que possibilitam e alteram radicalmente o modelo de monitoramento e de otimização de recursos e atividades em todos os níveis, gerando impacto nos processos de todas indústrias e negócios (SCHWAB, 2017).

No mundo empresarial e na sociedade, os modelos de negócio estão aumentando progressivamente sua interligação aos dados, porém, estes existem em quantidades nunca antes imaginadas e nos mais diversos formatos. Uma variedade de padrões, tecnologias, modelos, algoritmos, conceitos e equipamentos são necessários para produzi-los, armazená-los, tratá-los, analisá-los, visualizá-los e mantê-los íntegros e seguros. Chega-se então a uma nova ciência, a ciência de dados (*data science*), que tem por objetivo estudar estes elementos que demonstram promissor protagonismo em uma nova era da história humana (AMARAL, 2016).

Nos próximos itens, serão descritos os conceitos relacionados aos dados, como seu ciclo de vida, produção, armazenamento, análise e visualização.

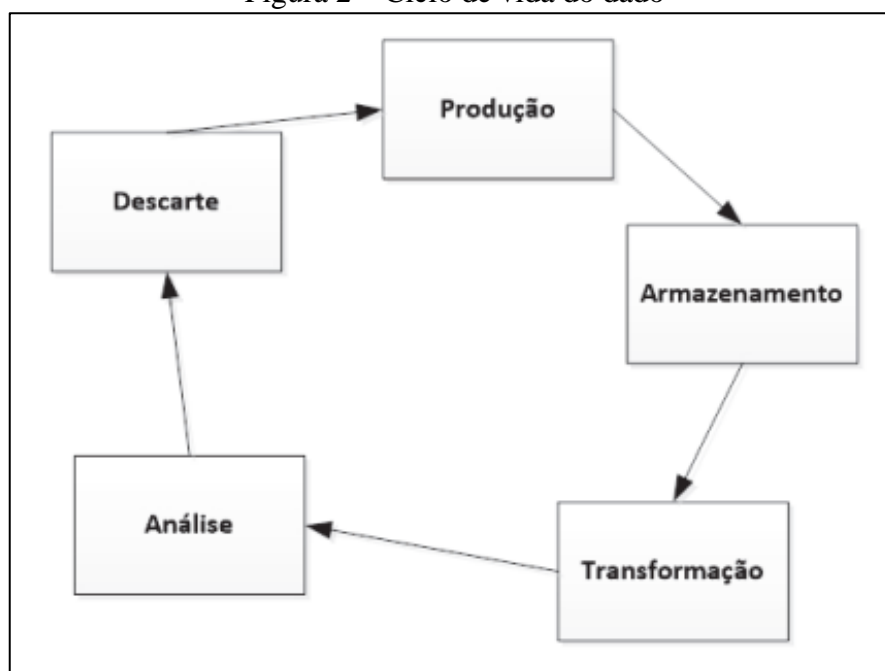
2.2 CICLO DE VIDA DO DADO

Antes mesmo de se entender a ciência de dados ou o *big data*, é necessário compreender suas matérias-primas: o dado, a informação e o conhecimento. Para Amaral (2016, p. 3) os “dados são fatos coletados e normalmente armazenados. Informação é o dado analisado e com algum significado. O conhecimento é a informação interpretada, entendida e aplicada para um fim”.

Os dados podem ser observados em três formatos diferentes: sendo o primeiro o dado não eletrônico, normalmente impresso em papel. O dado analógico, transmitido por ondas e passível de interferências eletromagnéticas e o dado digital, transmitido por pacote de bits, ou seja, aqueles armazenados na forma de zeros e uns, independentemente de sua estrutura (AMARAL, 2016).

Ainda segundo Amaral (2016), a ciência de dados é normalmente associada de forma equivocada apenas aos processos de análise de dados. Nesta visão inadequada, a *data science* passa a ser confundida com a estatística, porém, a análise é apenas uma etapa desta ciência, que pode ser definida pelos processos, modelos e tecnologias que estudam os dados durante todo o seu ciclo de vida: da produção ao descarte, conforme evidenciado na Figura 2.

Figura 2 – Ciclo de vida do dado



Fonte: Amaral (2016, p. 6).

Atualmente o termo *big data* vem sendo discutido em função dos desafios e vantagens de se coletar e processar o grande volume de dados gerados pelas tecnologias utilizadas. No entanto, o primeiro problema em se definir o termo vem de seu próprio nome, a expressão “*big*” pode denotar uma preocupação única com o volume de dados gerados, mas em sua conceituação a variedade de informações é outro problema relevante, diferentes fontes gerando diferentes protocolos de estruturação devem ser observados e incorporados pelo modelo de análise (FERNÁNDEZ et al., 2014).

Além do mais, a eficiência e velocidade de sistemas computacionais denotam preocupação com a latência na busca e processamento dos dados, pois alguns conhecimentos descobertos podem perder sua utilidade pelo fator tempo (RABELO; CAMPOS, 2014).

Por fim, segundo Fernández (2014) o último aspecto que deve ser avaliado é referente à veracidade das informações, manter a integridade dos dados evitando ruídos e interferências garante confiança da informação para tomada de decisões.

Estes 4Vs (volume, variedade, velocidade e veracidade) definem *big data*, é possível apontar demais definições englobando valor, como dados com significado para os negócios, que contribuam com valor agregado e visualização, observando o modelo de ilustrações de informações geradas (FERNÁNDEZ et al., 2014).

Para Amaral (2016), estes conceitos não fazem menção às causas e consequências, para ele *big data* é um fenômeno e não uma tecnologia. Um movimento em que dados são produzidos em vários formatos e armazenados por uma grande quantidade de equipamentos, gerados pela disseminação de dispositivos e processos capazes de produzir e armazenar informações.

Big data não se trata apenas de processos que geram grandes volumes de dados, mas sim do envolvimento e uso de diversos conceitos e tecnologias, tenho seu impacto social, cultural e empresarial capaz de alterar o mundo como é conhecido atualmente (AMARAL, 2016).

2.2.1 Produção

Dados podem ser comprados, produzidos ou simplesmente coletados. Os dados são adquiridos por meio de empresas especializadas em vende-los, as *data brokers*, como Acxiom, CoreLogic, eBureau, IDAnalytics, Intelius e PeekYou. Já os produzidos são gerados por sistemas transacionais, enquanto os coletados podem provir de variados meios, sistemas, pesquisas, dados históricos, arquivos ou até mesmo um *data warehouse* (AMARAL, 2016).

Nos anos de 1950 os dados passaram a ser recolhidos por meio de sistemas legados em formato estruturado de dados relacionais, inicialmente se resumia a processadores que tornaram eletrônicos os processos que antes eram mecânicos. Em 1990, outras aplicações, como o *Enterprise Resource Planning* (ERP) ganharam força, gerando elevado volume de informação. Já na chegada dos anos 2000, foi observado o advento da internet e do comércio eletrônico, tendo como resultado o registro instantâneo de acontecimentos. Nos últimos anos foi evidenciado um salto na produção de dados, justificado por meio da massificação de dispositivos móveis, computação em nuvem, sensores e conexão de diversos equipamentos utilizados no cotidiano (SILVEIRA; MARCOLIN; FREITAS, 2015).

Amaral (2016) exemplifica a elevada utilização de sensores por meio simples *smartphone*, o qual apresenta câmeras, telas sensíveis ao toque, acelerômetros, *Global Positioning System* (GPS), giroscópio e magnetômetro. Tendo como tendência mundial o crescimento no uso de sensores em celulares e outros tipos de dispositivos.

Como fonte de dados pode-se ainda citar os serviços web, mídias sociais, serviços de dados abertos como governos, arquivos de dados em bibliotecas, repositórios, cartões de crédito, televisões, computadores, infraestrutura de cidades, sensores prediais, trens, ônibus, aviões, pontes, indústrias, entre outros. Estima-se que apenas 0.5% dos dados disponíveis são analisados, muito em função de limitações do modelo tradicional de computação, armazenamento e ferramentas de captura (MUTULA, 2016).

2.2.2 Armazenamento

Os dados produzidos são normalmente armazenados, garantindo sua recuperação, duplicação ou análise no futuro. O armazenando deve atender algumas premissas, como: segurança da informação, integridade, minimização de redundância, concorrência e otimização de espaço. Estas informações podem ser mantidas em dispositivos não voláteis, que mantêm sua integridade na ausência de energia elétrica, como discos rígidos ou memórias *flash*, ou equipamentos voláteis que dependem da eletricidade, como o caso da *Random Access Memory* (memória RAM) de um computador. Durante seu ciclo de vida, o dado é constantemente replicado entre diferentes tipos de dispositivos com o intuito de atender variados propósitos, podendo sofrer ou não transformação de sua estrutura (AMARAL, 2016).

Historicamente, a estruturação de armazenamento de dados evoluiu conforme a demanda de processamento, complexidade e volume de informações, exigindo gerenciados

capazes de incluir, alterar e indexar conteúdo cada vez melhores. O Quadro 2 traz um resumo dos modelos utilizados ao longo das últimas décadas.

Quadro 2 – Armazenamento de dados

| Modelo | Década |
|---------------------|--------|
| Hierárquico e Rede | 1960 |
| Relacional | 1970 |
| Orientado a Objetos | 1980 |
| NoSQL | 2000 |

Fonte: Amaral (2016).

Segundo Amaral (2016), na década de 1960 foram desenvolvidos modelos com estrutura de navegação, onde dados eram armazenados em registros vinculados e para sua recuperação era necessária a navegação entre todos os vínculos criados, esta estrutura foi chamado de hierárquica e rede, conhecida também como pré-relacional. Buscando reduzir a redundância de dados e manter a integridade entre as transações, Edgar Frank Codd criou o modelo relacional em 1970, esta estrutura é baseada em álgebra relacional, normalizando bancos de dados através de tabelas, tendo uma linguagem de consulta estruturada chamada de *Structured Query Language* (SQL).

O modelo SQL tornou-se bastante popular, sendo implementado em milhões de sistemas dos mais diversos tipos de negócios, rodando em praticamente todas plataformas computacionais e sendo ainda utilizado nos dias de hoje.

Com o advento da internet e o fenômeno do *big data*, novos requisitos para o armazenamento de dados surgiram, como escalabilidade e redundância, bem como a capacidade de processar informações não estruturadas. Estas especificações não são eficientemente atendidas pelo modelo relacional, surgindo assim o banco de dados orientado a objetos e o *Not Only SQL* (NoSQL). Isto não significa que o SQL está defasado, estas são soluções diferentes para problemas diferentes (AMARAL, 2016).

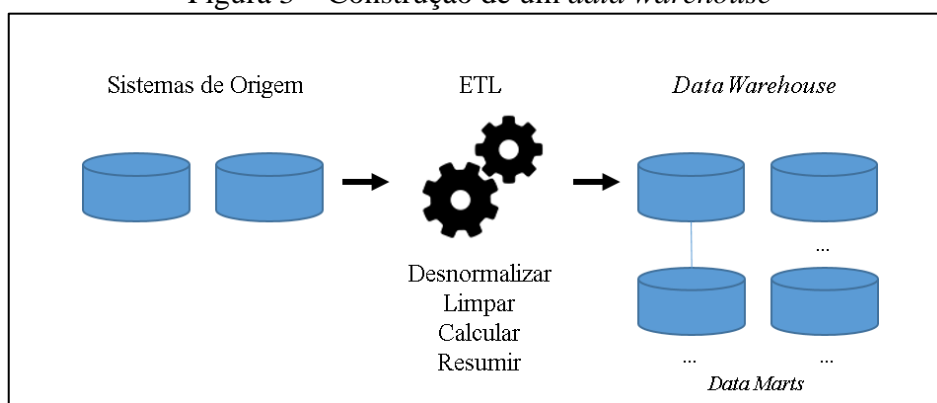
Para Amaral (2016, p. 30), o modelo orientado a objetos busca “abstrair entidades reais do mundo em objetos descritos por atributos e procedimentos executados como métodos”, enquanto o NoSQL são bancos de dados que não estão apenas baseados no modelo relacional, contemplando gestão de documentos, análise de séries temporais, agregando grandes volumes de dados, textos, imagens e apresentando uma estrutura com dados não normalizados, poucas restrições de integridade e controle mínimo de transações, gerando um banco de dados mais escalável, flexível e com menor custo.

Neste momento é importante observar que o modelo para armazenar dados transacionais, chamado de processamento de transações em tempo real, ou *Online Transaction Processing* (OLTP), busca otimizar o processamento e manter as operações por meio de modelos relacionais de transações repetitivas, em grandes quantidades e manipulações simples. Para a análise de dados, estas estruturas não apresentam melhor formato, pois sobre o estudo de um evento qualquer, será necessário a transformação da informação em um formato distinto de acordo com a especificidade da análise a ser executada (TURBAN et al., 2009).

Portanto, surge a necessidade de separar o OLTP do processamento analítico em tempo real, o *Online Analytical Processing* (OLAP), caracterizado pelo exame de muitos itens de dados em relacionamentos complexos e identificado regularmente pela presença de um *data warehouse*. Para Turban et al. (2009), um *Data Warehouse* (DW) é um conjunto de dados produzidos para oferecer suporte à tomada de decisões, sendo também um repositório de dados atuais e históricos, estruturados de modo a estarem disponíveis em um formato pronto para as atividades de processamento analítico.

As operações de transformação entre os sistemas de OLTP e OLAP são realizadas pelos processos de extração, transformação e carga, chamadas de *Extract, Transform and Load* (ETL) e são componentes integrais de qualquer projeto centrado em dados, realizando as atividades de leitura dos dados de um ou mais bancos, conversão da informação extraída de acordo com a necessidades e carregamento do histórico no DW (TURBAN el al., 2009). Para ilustrar estas relações, a Figura 3 pode ser observada.

Figura 3 – Construção de um *data warehouse*



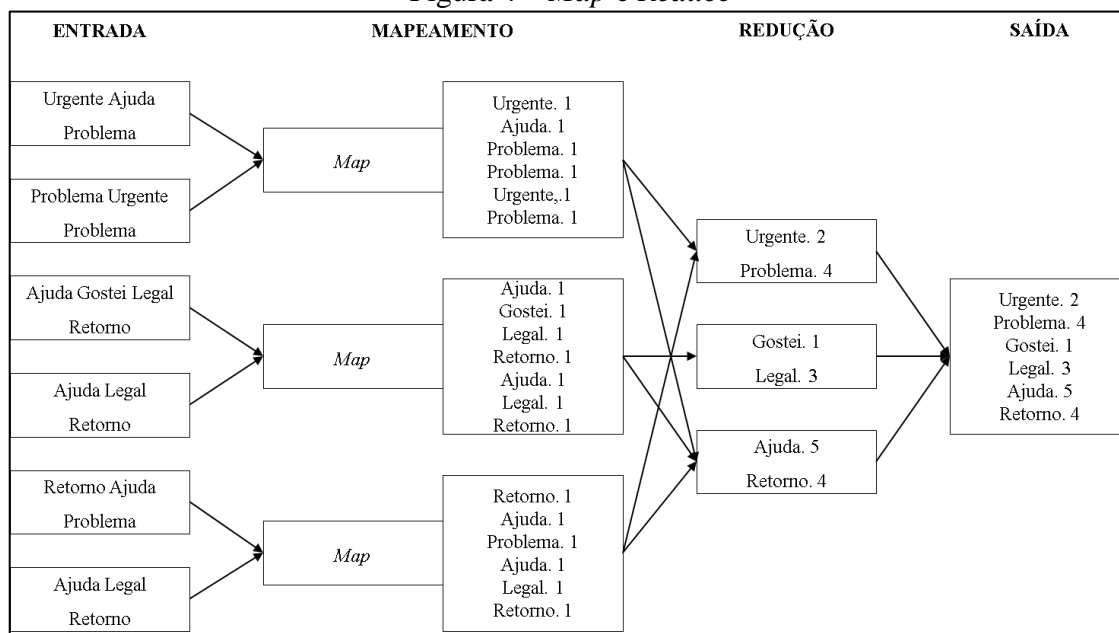
Fonte: Adaptado de Amaral (2016).

Com o novo desafio de se armazenar e processar um grande volume de dados de forma simplificada, novas tecnologias são demandadas, segundo Victorino et al. (2017), entre os modelos mais promissores encontra-se o *middleware* Hadoop. Sistema que destaca-se por implementar o conceito apresentado pela empresa Google, denominado MapReduce,

abordagem que visa dividir unidades de trabalho e processá-las em paralelo em uma rede de computadores chamados de nós. Os dados se encontram distribuídos pelos nós e o processo é dividido em dois estágios, sendo o primeiro de mapeamento e o segundo de redução.

No mapeamento busca-se encontrar os dados nos nós de acordo com a função escrita e a redução visa receber os resultados e consolidar as informações. É importante ressaltar que o dado é processado no nó específico, o tráfego se dá apenas nos resultados, evitando utilização desnecessária de armazenamento e recurso computacional (AMARAL, 2016). A Figura 4 representa o processo clássico de aplicação do MapReduce, com objetivo de contar palavras, textos são distribuídos em vários nós, funções de mapeamento contam estas palavras e funções de redução consolidam os dados recebidos pelos nós.

Figura 4 – Map e Reduce



Fonte: Adaptado de Amaral (2016).

Para Amaral (2016), o MapReduce não é uma substituição ao *data warehouse* tradicional, mas sim uma complementação, ele entende que estas aplicações serão fontes de dados para os *datacenters*, onde somente a informação de maior valor, estruturada e tratada será carregada.

2.2.3 Análise

Analisar dados é aplicar algum tipo de transformação em busca de conhecimento. Dentro dos modelos de análise, pode-se observar dois principais: a explícita, onde a informação já está disponível explicitamente nos dados e apenas uma operação de baixa complexidade é

necessária; e a análise implícita, onde o conhecimento não é claro, mesmo com filtros e cálculos a informação só será produzida se houver o uso de funções mais sofisticadas (AMARAL, 2016).

Segundo Amaral (2016), é importante classificar a análise para buscar a técnica apropriada, de acordo com o seu objetivo. Dentre as ferramentas da análise explícita, pode-se citar brevemente as técnicas quantitativas, utilizando média, mediana, amplitude, desvio padrão, diagramas de dispersão, diagramas de caixa, histogramas, nuvem de palavras e caras de Chernoff.

O presente trabalho apresenta um enfoque em análises implícitas em função da conceituação de *big data*, e por tal motivo não discorrerá detalhadamente sobre modelos explícitos.

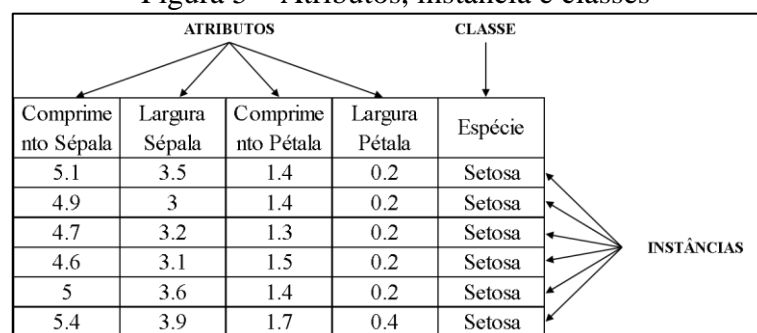
O primeiro conceito a ser entendido em análises implícitas é referente ao aprendizado de máquina computacional (*machine learning*) e a mineração de dados (*data mining*), embora ambos contemplem aplicações técnicas computacionais que buscam encontrar padrões ocultos, o aprendizado de máquina trata de algoritmos que buscam reconhecer padrões em dados e a mineração é a aplicação destes algoritmos em grandes conjuntos em busca de informação e conhecimento. Suas aplicações são praticamente ilimitadas: onde houver dados, pode haver um processo de mineração. O *big data* destacou esta conceituação, pois não é válido uma elevada quantidade de registros se não for possível aproveitá-los para produzir conhecimento (AMARAL, 2016).

Existem variados produtos no mercado que mineram dados. Desde privados como Microsoft e Oracle, até produtos *open source*, como R e *Waikato Environment for Knowledge Analysis* (Weka). Sendo o R o principal aplicativo utilizado para proposta de implementação deste trabalho.

Conforme Moraes, Sales e Costa (2011) o R foi idealizado por Robert Gentleman e Ross Ihaka e desenvolvido por esforço colaborativo, sendo muito utilizado por universidades e com crescente aceitação em empresas. Sua utilização é bastante difundida na literatura recente e é altamente expansível com o uso dos pacotes desenvolvidos pela comunidade no mundo inteiro, que são bibliotecas para funções específicas ou áreas de estudo específicas.

Para Amaral (2016), existem três conceitos elementares no aprendizado de máquina: os atributos, instâncias e classes. Comparando-se à uma planilha convencional, as colunas com valores semânticos podem ser consideradas atributos ou dimensões, enquanto as linhas com informações coletadas serão as instâncias. Para as classes ou variáveis de interesse tem-se os elementos de previsão ou classificação desejada. A Figura 5 demonstra por meio de um exemplo o conceito descrito.

Figura 5 – Atributos, instância e classes



Fonte: Adaptado de Amaral (2016).

Conforme explica Amaral (2016) e Debuse (2007), a mineração de dados pressupõe uma série de etapas, que vão desde entender o negócio até sua implementação. O padrão mais conhecido é o *Cross Industry Standard Process for Data Mining* (CRISP-DM). Este contempla seis etapas:

- a) entendimento do negócio: compreender as características do ambiente e seus objetivos;
- b) entendimento dos dados: observação de estrutura, quantidade, qualidade e acesso às informações;
- c) preparação dos dados: organização, limpeza e seleção dos dados;
- d) modelagem: criação do modelo;
- e) avaliação: avaliação do modelo por meio de matriz de confusão;
- f) implementação: onde processo de mineração é implantado.

Conforme salienta Debuse (2007), o *framework* CRISP-DM é o mais utilizado para a mineração de dados, concorrendo apenas com o *Sample, Explore, Modify, Model and Assess* (SEMMA), a diferença principal entre eles se refere a inexistência de aspectos de entendimento do negócio no modelo SEMMA, sendo seu principal foco apenas o modelamento de tarefas computacionais para mineração de dados.

Para construção do modelo no CRISP-DM, normalmente os dados históricos são divididos em dois grupos: uma parte utilizada para treinamento do modelo (70%) e outra para testar (30%). Uma vez submetidos os dados de treino ao algoritmo, o modelo é construído e os dados de teste são verificados. Por serem dados históricos, os dados de teste já são classificados, podendo assim observar-se a efetividade do modelo comparando-se a informação conhecida com o previsto pelo algoritmo em uma matriz de confusão (AMARAL, 2016). O Quadro 3 apresenta um exemplo de matriz de confusão onde os dados previstos pelo algoritmo estão nas colunas e os dados reais nas linhas.

Quadro 3 – Exemplo de matriz de confusão

| Dados | Previsão | | |
|-------|----------|-----|-----|
| | | Sim | Não |
| | Sim | 732 | 40 |
| Não | 29 | 199 | |

Fonte: Adaptado de Amaral (2016).

É possível observar no exemplo acima a incidência de 29 falsos-positivos e 40 falsos-negativos, gerando uma precisão de 93% (total de acertos divididos pelo total de instâncias) e uma taxa de erro de 7% (total de erros divididos pelo total de instâncias).

A aprendizagem de máquina pode ser dividida em três grandes processos: classificação, agrupamento e associação. Na classificação, tem-se histórico de instâncias e suas respectivas classes, uma nova instância é coletada e se deseja assim classificá-la. Caso não haja classificações, pode-se utilizar o agrupamento, separando grupos com atributos semelhantes. De mesma forma, pode-se desejar avaliar relações entre os atributos observados, encontrando associações de causa e efeito por classes (AMARAL, 2016).

Amaral (2016) ressalta que cada um dos processos apresenta técnicas de se resolver uma tarefa de aprendizado de máquina diferentes e por sua vez, cada técnica, apresenta algoritmos diferentes para sua aplicação. Com objetivo de demonstrar as principais técnicas e algoritmos aplicados em cada processo, o Quadro 4 pode ser observado:

Quadro 4 – Processos, técnicas e algoritmos

| Processos | Técnicas | Algoritmos |
|----------------------|----------------------|------------------------|
| Classificação | Bayes | NaiveBayes BaysNet |
| | Rules | Party DecisionTable |
| | Árvores de Decisão | Random Forest J48 |
| Agrupamento | Por Densidade | DBSCAN |
| | Baseado em Protótipo | K-means K-medoids |
| Regras de Associação | | Apriori FP Growth |

Fonte: Adaptado de Amaral (2016).

A técnica de árvores de decisão (*decision trees*) desenvolve um modelo onde uma árvore com nodos é construída. A cada nodo, o valor da instância é avaliado, seguindo por nodos internos de acordo com a definição anterior, até sua condição de parada. Uma árvore de decisão possui um nodo raiz, nodos internos e nodos folha, onde a classificação é definida. Para estabelecer a sequência de particionamento e a ordem de testes dentro da árvore, utiliza-se algoritmos de expressão lógica, o método pode ser aplicado em instâncias nominais ou de valores numéricos (AMARAL, 2016).

Baseado na teoria de Thomas Bayes o algoritmo Naive Bayes avalia o quanto cada atributo contribui para classificar uma instância, construindo uma tabela de probabilidades. Na classificação o modelo soma os índices obtidos, e o valor que possuir maior índice é selecionado e sua classificação é efetivada. Apesar de simples, segundo Amaral (2016) a técnica pode gerar elevados índices de precisão.

Um terceiro método pode ser utilizado para o processo de classificação: as Redes Neurais Artificiais (RNAs). Segundo Braga, Carvalho e Ludermir (2000), as RNAs são sistemas paralelos distribuídos por unidades de processamento simples que calculam funções matemáticas. Estas unidades são dispostas em camadas e interligadas por um elevado número de conexões, geralmente unidirecionais. As conexões estão associadas a pesos, os quais armazenam conhecimento e servem para ponderar a entrada recebida por cada neurônio da rede. Este funcionamento é inspirado em uma estrutura física concebida pela natureza: o cérebro humano.

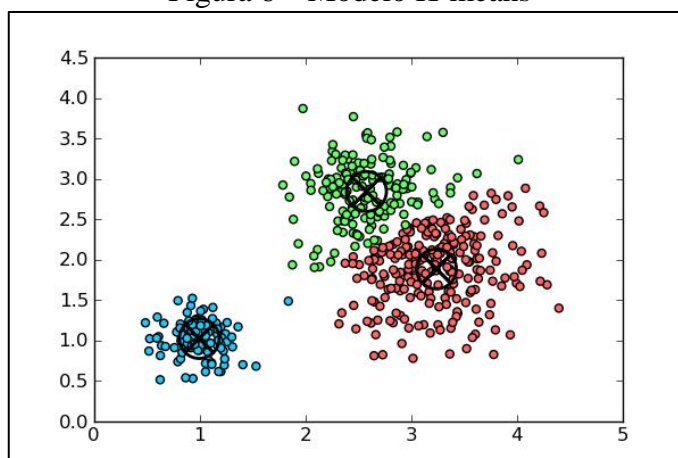
Os modelos baseados em RNAs, em função de sua forma de rede e de seu paralelismo, apresentam um desempenho superior aos modelos convencionais. O procedimento inicia por meio da fase de aprendizado, onde um conjunto de exemplos é apresentado à rede, a qual extrai automaticamente as características necessárias para representar a informação fornecida. Estas características são utilizadas posteriormente para gerar respostas para o problema (BRAGA; CARVALHO; LUDERMIR; 2000).

Segundo Amaral (2016), dentro dos métodos de classificação, tem-se ainda a regressão e a correlação, a regressão busca prever um valor numérico baseado em dados históricos, desde que haja relação matemática entre estes. Esta relação pode ser observada por meio da correlação, definindo sua força e direção. A correlação, quanto testada, retorna valores entre 1 e -1, quanto mais próxima dos extremos mais forte ela é, de mesma forma, o sinal do valor identifica se uma correlação é positiva ou negativa. As regressões comumente utilizadas podem ser formuladas por meio de retas, chamadas de lineares; ou exponenciais, tratadas pelo mesmo nome.

Para o agrupamento, as tarefas buscam reunir instâncias com características comuns em grupos, que posteriormente serão classificados. Existem dois principais modelos de algoritmos para o agrupamento, sendo um baseado em formas geométricas, avaliando a proximidade da instância a um centroide e o outro caracterizado pela densidade, agrupando elementos conforme uma função de proximidade.

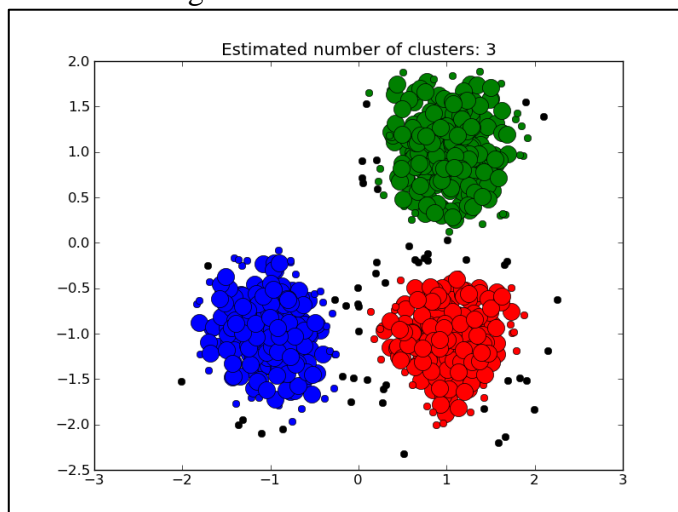
Para agrupamento por centroide tem-se os algoritmos K-means e K-medoid, nestes o número de grupos a ser formado é definido pelo usuário e todas as instâncias dispostas são agrupadas. Para densidade de dados tem-se o HBSCAN, onde a quantidade de grupos é definida automaticamente pelo algoritmo e dados com densidade baixa são eliminados do estudo, sendo considerados ruídos (AMARAL, 2016), essas técnicas são comuns na área de estatística e são encontradas em *softwares* comerciais como o SPSS e Minitab. As Figuras 6 e 7 ilustram os agrupamentos de cada modelo.

Figura 6 – Modelo K-means



Fonte: Amaral (2016).

Figura 7 – Modelo DBSCAN



Fonte: Amaral (2016).

Para associação, dois algoritmos são populares: Apriori e FP-Grow. Eles são baseados na relação entre dois eventos e a probabilidade destes acontecerem em sequência. Muito utilizados pelo comércio eletrônico, eles identificam em base histórica a relação: quem comprou A também comprou B, então quando o cliente comprar A deve-se oferecer B. Os algoritmos de aprendizado de máquina mineram as transações em busca de associação entre itens, porém, o número de regras pode ser muito grande, por este motivo duas métricas principais são utilizadas: suporte e confiança. Sendo o suporte a proporção de transações que contém todos itens e a confiança a proporção de transações que contendo A, contém B. Para um algoritmo, estes parâmetros são inseridos com um valor mínimo esperado, caso a regra encontrada tenha valores inferiores ao desejado, ela não será utilizada.

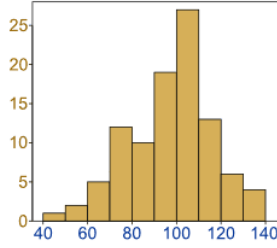
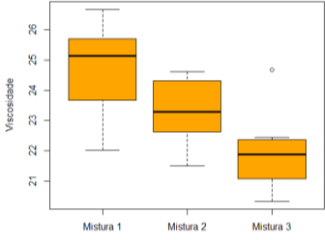
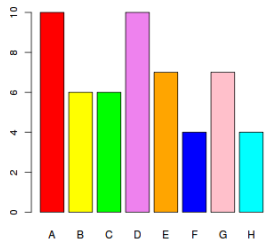
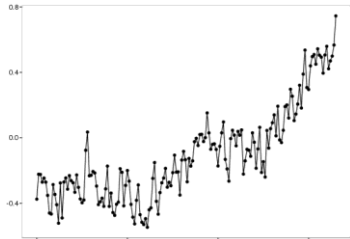
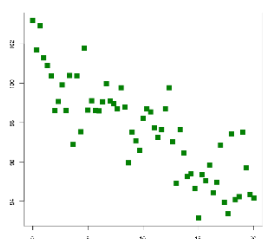
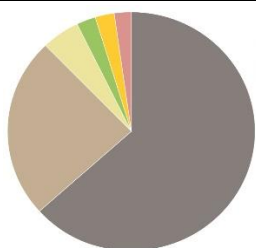
Para Amaral (2016), uma última regra de análise de dados apresenta destaque: a mineração de texto, ela é relevante pois 80% dos dados produzidos estão nesse formato. Inserindo textos de diversas fontes, como *e-mails*, internet e rede sociais, este algoritmo irá primeiramente remover as chamadas *stop words*, que são palavras sem valor semântico, como: aos, seus, quem, nas e me. Posteriormente tem-se o *steming*, buscando extrair todos os radicais livres, com o objetivo de agrupar palavras com mesmo significado. Remove-se também a pontuação, numeração, símbolos e linguagens de marcação. Findando os tratamentos, a mineração de texto pode produzir uma matriz de termos e suas frequências, pode ser utilizada para classificação de documentos, análise de sentimentos, construção de nuvem de palavras, entre outros.

2.3 VISUALIZAÇÃO DE DADOS

A visualização de dados refere-se às tecnologias que dão suporte à interpretação de dados e informações ao longo da cadeia de processamento. Esta etapa pode contemplar imagens, gráficos, realidade virtual, representações dimensionais, vídeos e animações, ajudando a identificar relações, como no caso de tendência (TURBAN et al., 2009).

Segundo Padua, Dias e Lima (2015), estas representações nos permitem analisar um grande volume de dados ao mesmo tempo. Para que a informação seja transmitida de forma clara e objetiva, é necessário escolher técnicas de visualização apropriadas. O Quadro 5 traz os principais tipos de gráficos utilizados pela ciência de dados e suas respectivas aplicações.

Quadro 5 – Tipos de elementos gráficos adequados

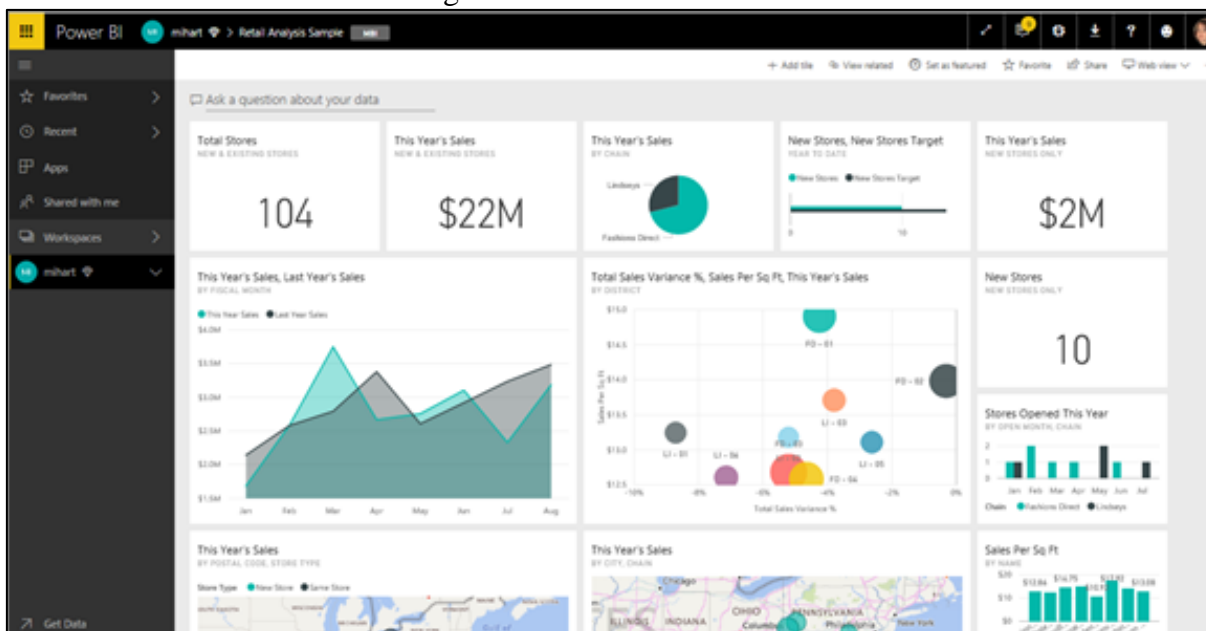
| Tipo de Gráfico | Utilização | Exemplo |
|-----------------------|--|---|
| Histograma | Mostrar a distribuição de um único dado |  |
| Diagrama de Caixa | Mostrar a distribuição de um ou mais dados |  |
| Gráfico de Barras | Quando uma das variáveis é nominal |  |
| Séries Temporais | Quando dados quantitativos são coletados regularmente em uma escala de tempo |  |
| Gráficos de Dispersão | Correlacionar duas variáveis quantitativas |  |
| Gráfico de Setores | Comparar partes de um total |  |

Fonte: Adaptado de Amaral (2016).

Dashboards também são comumente utilizados na análise e monitoramento de áreas específicas. Para Turban et al. (2009), eles proporcionam exibições visuais de informações

importantes, que são consolidadas e organizadas em uma tela única para serem absorvidas e exploradas facilmente, um modelo de *dashboard* pode ser observado na Figura 8.

Figura 8 – Modelo *dashboard*



Fonte: Hart (2017).

Os *dashboards* podem ser operacionais, usados por supervisores para monitorar dados resumidos e atualizados com frequência ao longo do dia. *Dashboards* táticos, utilizados por gerentes e analistas no acompanhamento diário ou semanal gerados por processos e projetos departamentais. E por fim, *dashboards* estratégicos, disponíveis a gerentes e executivos mensalmente para visualização de informações e objetivos estratégicos (TURBAN et al., 2009).

Para Amaral (2016), a visualização tem um papel fundamental na análise de dados e no seu papel em descrever um fato, visualizar dados permite resumir informações, comunicar de forma mais efetiva, compreender, explorar, interpretar e analisar. Por esta razão a forma de representar a informação gerada deve ser estudada e adequada ao formato de conhecimento esperado.

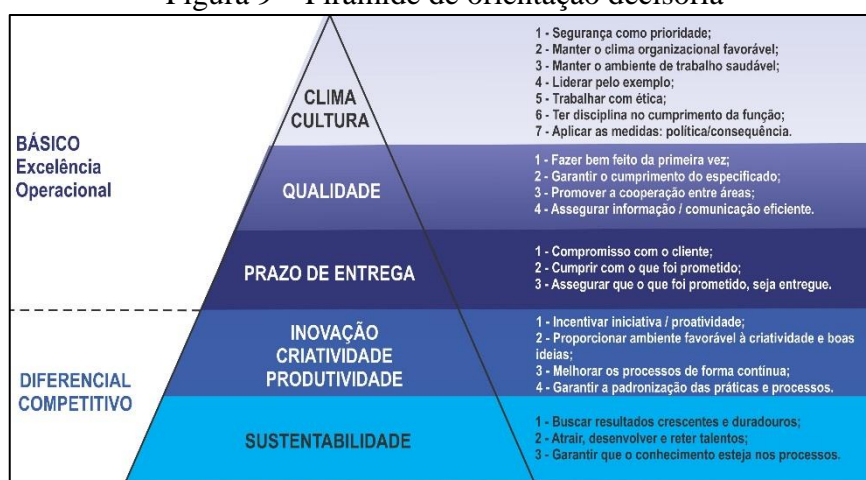
3 PROPOSTA DE TRABALHO

Como descrito anteriormente, o trabalho teve como objetivo principal o desenvolvimento de um algoritmo de predição de ocorrências de acidentes ocupacionais por meio da mineração de dados da mão de obra direta da empresa metalúrgica de Caxias do Sul, Randon S/A Implementos e Participações. Para tanto, buscou-se descrever neste capítulo o cenário atual do ambiente de aplicação do trabalho frente a análise de dados gerados.

3.1 CENÁRIO ATUAL

A gestão da Randon S/A Implementos e Participações possui uma política clara referente à priorização de aspectos de clima e cultura, conforme visualizado na pirâmide de orientação decisória apresentada como Figura 9. Para tanto, a companhia apresenta em seu site fabril um centro médico, destinado ao atendimento clínico e funcional de todos os funcionários do grupo Randon.

Figura 9 – Pirâmide de orientação decisória



Fonte: Banco de Imagens Randon (2017).

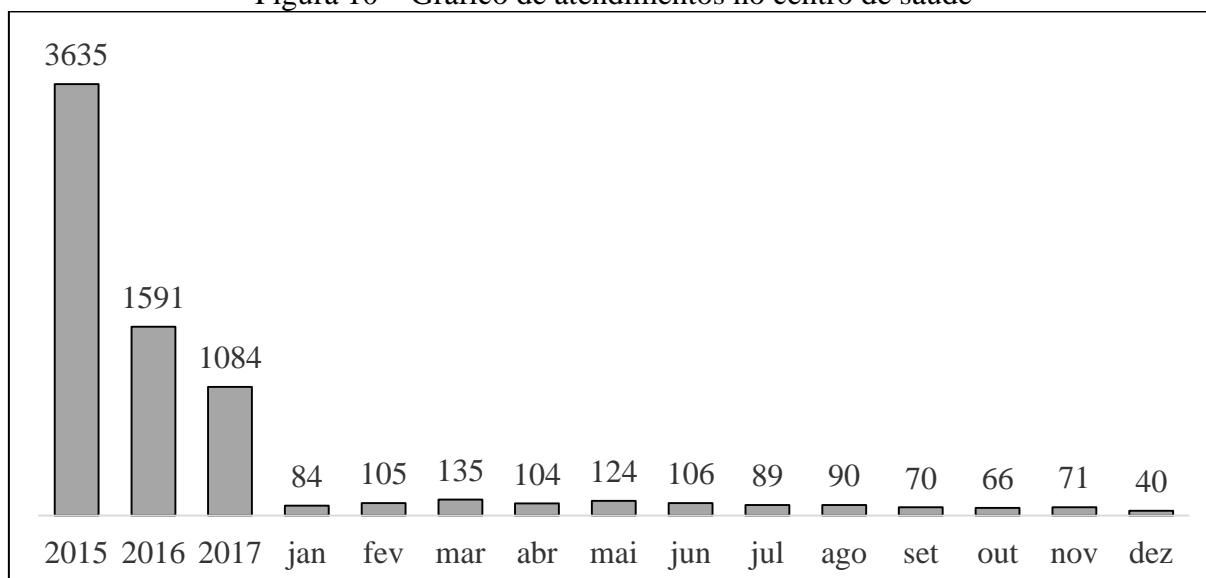
O centro de saúde, assim chamado, conta com médicos clínicos, um médico do trabalho, um enfermeiro, dois técnicos de enfermagem, fonoaudiólogos, dentistas e uma Unidade de Tratamento Intensivo (UTI) móvel, destinada ao resgate de infortúnios graves ocorrentes dentro da fabricação dos produtos Randon. A assistência não compreende apenas questões ocupacionais, todos os colaboradores podem utilizar o serviço de forma clínica, sendo este assunto não abordado no trabalho em função de seu vínculo particular e abrangência de escopo.

Para o registro de atendimentos, o centro utiliza o módulo de recursos humanos do ERP empregado pelo Grupo Randon, *software* de gestão fornecido pela companhia alemã SAP. Neste módulo, o médico responsável pelo atendimento registra o ocorrido via campo de texto, e de mesma forma, relaciona ao funcionário por meio de sua matrícula no sistema. As informações cadastradas em banco de dados relacionais não são analisadas posteriormente, apenas o número total de atendimentos durante o dia é observado. De maneira geral, a empresa não realiza ações corretivas ou utiliza-se destas informações de atendimentos disponíveis para prevenção de casos mais graves.

Em sua estrutura, a entidade conta também com uma área de Saúde, Segurança e Meio Ambiente (SSM), este setor responde tecnicamente por todas aspectos legais e diretrizes organizacionais referentes aos assuntos de clima e cultura da empresa. Seu quadro funcional conta com um engenheiro de segurança e seis técnicos, estes atendem áreas específicas da organização e acompanham diariamente os Atendimentos Ocupacionais (AT) realizados no centro de saúde, os Acidentes Sem Perda de Tempo (ASPT) e os Acidentes com Perda de Tempo (ACPT), sendo a diferença entre os dois casos, o afastamento ou não do funcionário de suas atividades após 24 horas do ocorrido.

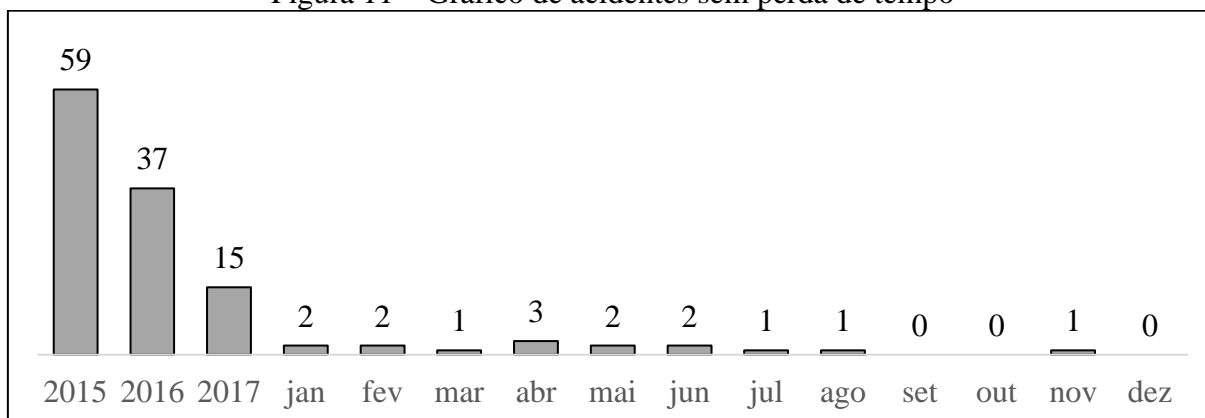
Como indicadores operacionais, tem-se o número de AT, ASPT e ACPT ocorridos ao longo do tempo, conforme apresentado respectivamente nos gráficos ilustrados como Figuras 10, 11 e 12.

Figura 10 – Gráfico de atendimentos no centro de saúde



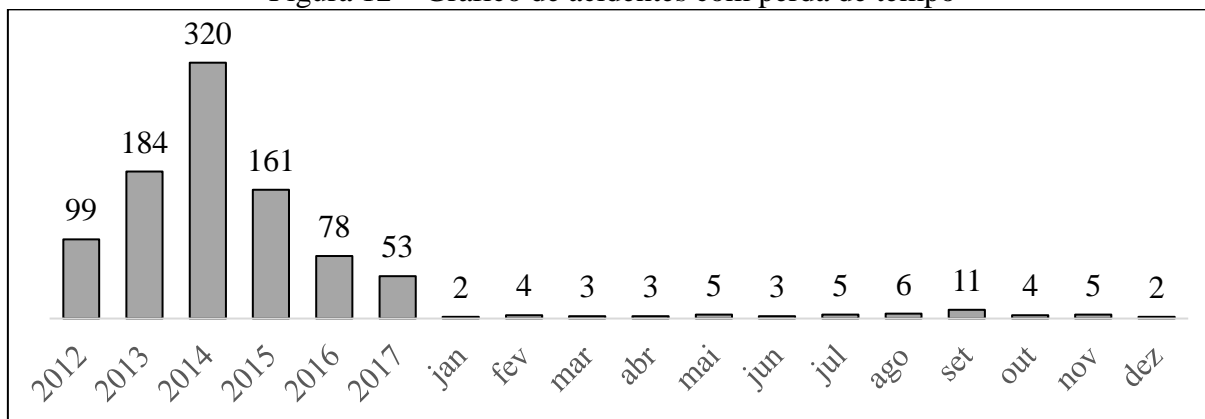
Fonte: SSM Randon (2017).

Figura 11 – Gráfico de acidentes sem perda de tempo



Fonte: SSM Randon (2017).

Figura 12 – Gráfico de acidentes com perda de tempo



Fonte: SSM Randon (2017).

Para os acidentes de ambos tipos, o registro ocorre via *Microsoft Excel*. Na planilha eletrônica, o técnico de segurança encarregado pelo centro de trabalho desenvolve um documento chamado alerta de segurança. Neste arquivo, a ocorrência é descrita, bem como suas ações corretivas, um exemplo deste registro pode ser observado como Anexo A.

Outro aspecto importante de se avaliar são os registros efetuados pelos funcionários dentro do parque fabril, esta atividade é processada por meio de *software* de registro de ponto, *Ponto Soft*, da empresa Ponto Soft Enterprise. Dentro do sistema, os funcionários são cadastrados via valor numérico único, chamado de matrícula, e estão relacionados à área de trabalho que atuam via código de centro de trabalho.

A Randon Implementos conta com 67 centros de trabalho onde 1375 funcionários de mão de obra direta registram diariamente o horário de entrada na empresa, seu início de atividade, sua entrada no refeitório, seu término de atividade e saída do parque fabril por meio de catracas e relógios ponto disponibilizados em locais estratégicos da planta.

Estas informações são armazenadas em banco de dados relacionais e são utilizadas única e exclusivamente para cálculo de folha de pagamento, observando o tempo trabalhado por cada funcionário e a quantidade de refeições realizadas por este durante o mês.

Como pode ser observado, os dados gerados por três processos distintos da companhia: atendimentos centro de saúde, registro de acidentes de trabalho e registro de relógio ponto, não são relacionados e não são utilizados com o intuito de prever comportamentos, sua avaliação se dá apenas de forma corretiva. Esta necessidade de ultrapassar a barreira da correção para prevenção ou previsão é observada pela gestão da companhia e foi abordada na proposta de trabalho apresentada a seguir.

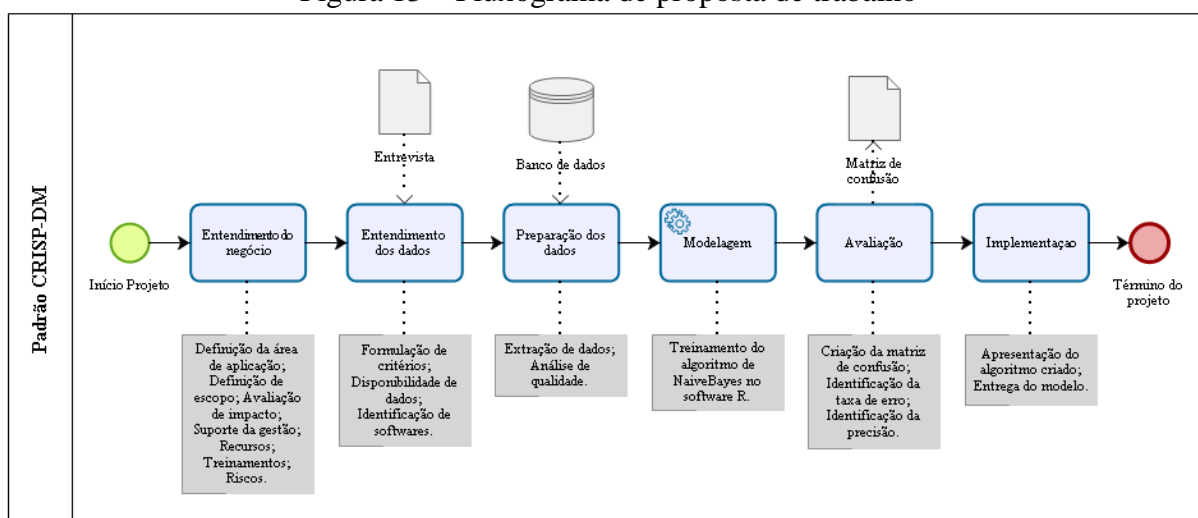
3.2 PROPOSTA DE TRABALHO

O trabalho teve como proposta a criação de um algoritmo de predição de ocorrências de acidentes por meio da mineração de dados da base histórica de informações relacionadas à segurança ocupacional dos funcionários de manufatura da Randon S/A Implementos e Participações.

Para tal, utilizou-se o método de mineração de dados CRISP-DM, apresentado por Rennolls e Al-Shawabkeh (2008) como *framework* conceitual dominante empregado por empresas de mineração de dados, seguindo as etapas de entendimento do negócio, entendimentos dos dados, preparação dos dados, modelagem, avaliação e implementação.

A sequência de atividades descritas a seguir pode ser melhor compreendida no fluxograma apresentando como Figura 13.

Figura 13 – Fluxograma de proposta de trabalho



Fonte: o autor (2017).

3.2.1 Entendimento do negócio

O modelo *Cross Industry Standard Process for Data Mining* parte do entendimento do negócio. Esta etapa, segundo Debusse (2007), tem fundamental importância dentro do processo de mineração de dados e pode significar a falha de projeto caso não seja adequadamente observada. Neste momento, todas características de negócio foram avaliadas por meio de observação do cenário atual, de mesma forma, as expectativas de resultados, estabelecidas por meio de um objetivo geral, foram validadas com a gestão da companhia.

Como componentes dessa fase, destacarem-se:

- a) definição da área de aplicação considerando a pirâmide de orientação decisória apresentada como Figura 9;
- b) definição de escopo observando-se o principal indicador da área de aplicação;
- c) entendimento do impacto do processo de mineração de dados dentro da empresa, avaliando-se as áreas e pessoas envolvidas pelo escopo;
- d) validação de apoio por parte da gestão por meio de reunião de abertura de projeto;
- e) descrição de recursos computacionais e humanos necessários para efetivação do trabalho;
- f) identificação de necessidade em treinamentos em função de utilização de tecnologia inédita dentro da companhia;
- g) listagem de riscos do projeto.

3.2.2 Entendimento dos dados

A segunda etapa contemplou o entendimento dos dados. Esta foi composta de duas sub-etapas da seguinte forma:

- a) formulação de critérios relacionados a acidentes ocupacionais dada por meio de entrevista à gerência industrial da Randon Implementos, cargo responsável pela área de Segurança, Saúde e Meio Ambiente da organização, que em conjunto ao engenheiro de segurança da empresa definiu os possíveis critérios que ocasionam acidentes de trabalho;
- b) observação da existência de registros em bancos de dados referentes aos critérios definidos no item anterior, identificando os *softwares* envolvidos no processo.

3.2.3 Preparação dos dados

Na etapa de preparação dos dados, se realizou a extração dos dados relacionados aos critérios definidos anteriormente e a análise de qualidade dos mesmos.

Para esta atividade foram extraídos manualmente, via exportação de arquivos de texto em formato *Comma-Separated Values* (CSV), os dados de cada *software* identificado no item 3.2.2.

Para análise de qualidade, os dados extraídos foram importados em planilha *Microsoft Excel*, onde se observou a estruturação de atributos e instâncias, eliminando registros incoerentes e desnecessários e organizando as instâncias conforme necessidade. Por fim, foi exportada a planilha resultante em formato CSV, possibilitando a importação de dados no software de modelagem.

3.2.4 Modelagem

Segundo Debuse (2007), o processo de mineração de dados através do método CRISP-DM é fundamentalmente influenciado pelo software de modelagem utilizado, pois problemas com grandes volumes de dados estão além das capacidades padrões de aplicativos de processamento analítico em tempo real e ferramentas de bancos de dados relacionais. Deve-se gerenciar de maneira apropriada os algoritmos utilizados e empregar com cuidado os dados de treinamento selecionados.

Para tanto, seguindo o método, os dados extraídos e preparados nos itens anteriores foram segregados, 70% do volume foi utilizado para treinamento do modelo e os outros 30% para teste de validação. Este treinamento foi realizado por meio de importação do documento de texto CSV no *software* de mineração de dados *open source* R, e aplicação do *toolkit* de classificação por meio do algoritmo de NaiveBayes disponível no aplicativo R.

Como resultado deste processo o *software* R gerou um algoritmo de classificação capaz de identificar possíveis acidentes ocupacionais por meio de inserção de novas instâncias.

3.2.5 Avaliação

Conforme modelo de avaliação descrito por Amaral (2016) e apresentado na fundamentação teórica, a proposta deste trabalho utilizou uma matriz de confusão para validação do algoritmo gerado pelo *software* R no item anterior.

Nesta etapa, foram utilizados os 30% de dados não aproveitados no treinamento, estas novas instâncias passaram pela análise do algoritmo criado, retornando os possíveis acidentes. Por se tratar de dados históricos, pôde-se confirmar os resultados, criando-se registros de falsos-positivos e falsos-negativos, obtendo-se a precisão e taxa de erro do algoritmo.

3.2.6 Implementação

Após avaliação do algoritmo criado, por meio de sua precisão e taxa de erro, o processo de mineração foi apresentado para gestão da empresa, podendo ser implementado em formato de sistema em tempo real de acompanhamento por meio de um *dashboard*. O processo de implementação não foi abordado no presente trabalho conforme delimitação de escopo, encerrando a proposta na entrega do algoritmo de previsão de acidentes para a companhia.

4 IMPLEMENTAÇÃO

Os seguintes tópicos relacionam a sequência de atividades realizadas observando o método de mineração de dados CRISP-DM para efetivação dos objetivos gerais e específicos descritos no trabalho.

4.1 ENTENDIMENTO DO NEGÓCIO

A metalúrgica Randon S/A Implementos e Participações de Caxias do Sul apresenta em seu modelo de gestão uma pirâmide de orientação decisória, ilustrada como Figura 9, que direciona a priorização de novos projetos e implementações dentro da companhia. Sua principal responsabilidade está relacionada ao clima e cultura, sendo segurança a prioridade de referência. Em consequência, a área definida para realização da proposta apresentada neste trabalho, foi a de Segurança, Saúde e Meio Ambiente (SSM).

Para acompanhamento e avaliação do aspecto segurança dentro da empresa, a SSM tem como indicador operacional principal o número de acidentes com perda de tempo. Este indicador representa o número de infortúnios ocorridos dentro do parque fabril em que o funcionário acometido tenha se afastado mais de 24 horas de suas atividades. Este parâmetro foi considerado como classificação alvo para o algoritmo de NaiveBayes, ou seja, a classificação de novas instâncias dentro do modelo deve prever a ocorrência de novos casos.

Esta antecipação a acontecimentos indesejados dentro da área industrial deve ter seus impactos discutidos, não cabendo no escopo de trabalho, definir os procedimentos de gestão realizados após identificação de possíveis acidentes. Esta etapa deve ser avaliada pela área de recursos humanos, entendendo que o algoritmo indicará individualmente um funcionário, podendo este ser afastado ou não das atividades no período crítico apontado. Este movimento influencia fortemente em diversos fatores humanos do ambiente, sendo estes o comportamento da gestão, do próprio funcionário, do seu local de trabalho e da área de segurança. Como consequência de responsabilidade do projeto, tem-se a identificação do operário com possível ocorrência de acidente, esta predição considera fatores humanos e apresenta taxa de erro, espera-se apenas que quando implementado o algoritmo seja capaz de identificar, e com consequentes ações, evitar eventualidades que impactarão em perdas físicas e ao ambiente social dos trabalhadores.

Visto a amplitude de repercussão do trabalho e da necessidade de utilização de recursos internos, o apoio e validação do projeto por parte da gestão responsável se fez

necessário. Antes de iniciar qualquer movimento, no dia 22 de agosto de 2017, em uma reunião privada marcada via *Microsoft Outlook*, realizada na sala da gerência industrial da Randon S/A Implementos e Participações unidade Caxias do Sul, o analista de engenharia de processos apresentou o escopo do projeto ao gerente industrial da planta, este responde em nível tático pela área de segurança, saúde e meio ambiente da companhia.

Após aprovação e solicitação de sigilo, o trabalho teve início. Durante o período do projeto, o funcionário responsável utilizou uma estação de trabalho (*workstation*) da marca *Hewlett-Packard* (HP), modelo Z400, com processador *Intel Dual Core Xeon W3520*, com 16 *gigabytes* de memória RAM, disco rígido de 500 *gigabytes* e sistema operacional *Windows 7*, sendo este o recurso computacional destinado pela empresa à engenharia de produto para modelagem e simulação de novos projetos. De mesma forma, foi solicitado e aprovado que o empregado fosse cadastrado como administrador do equipamento, esta etapa foi indispensável para instalação do *software R* e *toolkits* necessários ao longo do trabalho.

Para aplicação do *software R* não foram encontradas pessoas capacidades ou conhecedoras da solução dentro da Randon Implementos. Para o treinamento necessário, foi adquirido o livro “Introdução à Ciência de Dados” de Amaral (2016), este apresenta conceitos em sua primeira parte e aulas práticas de R em uma segunda.

Por fim, dentro da etapa de entendimento do negócio do método CRISP-DM, tem-se a listagem de riscos do projeto. Compreendidos neste trabalho como a possibilidade de não encontrar padrões entre as eventualidades registradas no ano de 2017, caracterizando uma alta taxa de erro e baixa precisão evidenciadas por meio da matriz de confusão gerada na etapa de avaliação. Neste caso, não seria possível a criação de um modelo de predição válido para aplicação e o objetivo geral do trabalho não seria atendido.

4.2 ENTENDIMENTO DOS DADOS

A fase de entendimentos dos dados do método CRISP-DM contemplou a formulação de um *cluster* de critérios que podem influenciar acidentes de trabalho dentro da empresa e a avaliação da existência de registros relacionados em *softwares* da companhia.

A criação dos parâmetros considerou o conhecimento empírico de dois entrevistados, as informações sobre os participantes e sobre a reunião podem ser observadas no Quadro 6. O evento foi gravado utilizando-se um *smartphone* e posteriormente teve os critérios transcritos no presente trabalho.

Quadro 6 – Reunião de formulação de critérios

| Cargo | Tempo no Cargo | Formação | Data | Local Reunião | Tempo de Duração |
|-------------------------|-----------------------|---|-------------|--|-------------------------|
| Gerente Industrial | 6 anos | Tecnólogo em Automação Industrial Mestrando Engenharia de Produção | 14/12/2017 | Sala de Reuniões da Segurança em Caxias do Sul | 1 hora |
| Engenheiro de Segurança | 7 anos | Engenheiro Químico Pós graduado em Segurança do Trabalho | | | |

Fonte: o autor (2018).

Para identificação dos *softwares* relacionados à cada critério, alinhamentos via telefonemas e *e-mails* com a área de soluções em engenharia da Tecnologia da Informação (TI) foram efetuados.

No Quadro 7 são listados os treze critérios observados e considerados na reunião anteriormente descrita, estes foram construídos a partir da apresentação do escopo de trabalho e discussão livre entre os participantes, é possível encontrar também os respectivos softwares que contém seus registros.

Quadro 7 – Critérios

(continua)

| Item | Critério | Descrição Critério | Equipamento | Software |
|-------------|--|--|--------------------|-------------------|
| 1 | Chegada anterior ao horário regular de trabalho | Avalia a efetivação de horas extras no dia observado (nova instância) | Relógio ponto | <i>Ponto Soft</i> |
| 2 | Chegada posterior ao horário regular de trabalho | Avalia a chegada em atraso do funcionário no dia observado (nova instância) | Relógio ponto | <i>Ponto Soft</i> |
| 3 | Saída anterior ao horário regular de trabalho | Avalia a saída antecipada do funcionário na data antecedente ao dia observado (nova instância) | Relógio ponto | <i>Ponto Soft</i> |
| 4 | Saída posterior ao horário regular de trabalho | Avalia a realização de horas extras na data antecedente ao dia observado (nova instância) | Relógio ponto | <i>Ponto Soft</i> |
| 5 | Refeição realizada fora da planta fabril | Avalia se o funcionário saiu da unidade fabril para realizar sua refeição no dia observado (nova instância) | Relógio ponto | <i>Ponto Soft</i> |
| 6 | Consumo de refeição extra | Avalia se o funcionário consumiu sobremesa ou proteína extra no dia observado (nova instância) | Relógio ponto | <i>Ponto Soft</i> |
| 7 | Alteração do local de trabalho | Avalia se o funcionário registrou o início de suas atividades em relógio ponto divergente ao utilizado na data antecedente ao dia observado (nova instância) | Relógio ponto | <i>Ponto Soft</i> |
| 8 | Ausência de trabalho | Avalia se o funcionário se ausentou do trabalho no dia antecedente ao dia observado (nova instância) | Relógio ponto | <i>Ponto Soft</i> |

(conclusão)

| | | | | |
|----|---------------------------|--|------------|------------------------|
| 9 | Atendimentos ocupacionais | Avalia se o funcionário realizou atendimentos ocupacionais no centro de saúde no dia antecedente, nos últimos sete ou trinta dias em relação ao dia observado (nova instância) | Computador | ERP SAP |
| 10 | Dia da semana | Avalia se em dias específicos da semana funcionários possuem maior propensão de sofrer acidentes de trabalho | Computador | <i>Microsoft Excel</i> |
| 11 | Quinzena do mês | Critério que avalia se em quinzenas do mês divergentes (primeira e segunda) funcionários possuem maior propensão de sofrer acidentes de trabalho | Computador | <i>Microsoft Excel</i> |
| 12 | Tempo de empresa | Avalia se o tempo de empresa de funcionários influencia na propensão dos mesmos sofrerem acidentes de trabalho | Computador | ERP SAP |
| 13 | Idade | Avalia se a idade de funcionários influencia na propensão dos mesmos sofrerem acidentes de trabalho; registro realizado via módulo de recursos humanos do ERP SAP | Computador | ERP SAP |

Fonte: o autor (2018).

Os itens observados partiram de um *brainstorming*, estes não são definitivos e tem seus pesos avaliados para modelagem. Porém, os itens de 1 a 8 possuem relevâncias relacionadas à variação do padrão de trabalho definido oficialmente pela empresa, sendo esta definição a entrada regular de trabalho às 7:12, almoço no refeitório da companhia, saída 17:00 horas e utilização do mesmo relógio ponto diariamente. Para o item 9 busca-se entender se funcionários com atendimentos ocupacionais no centro de saúde possuem maior possibilidade para acidentes. Do 10 ao 13 deseja-se avaliar padrões em características específicas.

4.3 PREPARAÇÃO DOS DADOS

A extração de dados foi realizada contemplando três etapas diferentes, esta divisão está relacionada diretamente ao *softwares* em que os dados dos critérios descritos no tópico anterior estão armazenados. Este processo contempla o objetivo específico de busca ao histórico por meio de informações nos bancos de dados disponíveis da companhia.

4.3.1 Registros de acidentes em *Microsoft Excel*

Para os registros de acidentes com perda de tempo acontecidos na Randon Implementos na unidade de Caxias do Sul em 2017, foram compilados de forma manual os dados informados nos 97 alertas de segurança desenvolvidos em *Microsoft Excel* pelo setor de Saúde, Segurança e Meio Ambiente. As informações do número do alerta, data do ocorrido, horário, matrícula do funcionário acometido e descrição geraram uma nova planilha, apresentada em partes como Apêndice A. Alertas com acidentes sem perda de tempo, dano material ou ambiental, foram desconsiderados em função do escopo de classificação alvo de ACPT. As descrições foram ocultadas em função da natureza do assunto tratado e da política de privacidade das empresas do grupo Randon.

4.3.2 Registros de atendimentos ao centro de saúde

Aos dados de atendimentos ocupacionais realizados no centro de saúde da Randon Implementos, se aplicou a transação zehs031 do módulo de recursos humanos do ERP SAP utilizado pela companhia.

Neste relatório, exportado em arquivo de extensão csv e apresentado em partes como Apêndice B, foram observados 1084 registros no ano de 2017, contemplando a data e hora do atendimento, matrícula do funcionário, seu centro de trabalho e o tipo e descrição da assistência realizada.

4.3.3 Registros de acessos no *Ponto Soft*

Para os registros de acessos e marcações de relógio ponto dos funcionários da companhia foram exportados em arquivos textos de extensão txt os dados mensais de todas as catracas e relógios presentes na empresa no ano de 2017 e gerenciados pelo *software Ponto Soft*.

Em função do volume de dados, contemplando mais de 2,5 milhões de registros, foi necessária a separação por relatórios mensais. Nestes relatórios foi possível encontrar a matrícula e nome do funcionários, seu centro de trabalho, a data e hora do registro, o código de referência e descrição da catraca ou relógio utilizado. Para cada marcação uma instância foi criada, resultando nos valores observados como Quadro 8. Ao mínimo, entende-se que um funcionário deva realizar 5 registros ao dia: entrada e saída do parque fabril, início e final de

atividade e marcação de refeição. Esta quantidade pode variar para mais ou menos dependendo das movimentações e particularidades de cada dia, por exemplo: se o funcionário realizou sua refeição no ambiente externo à empresa, não existirá a marcação do refeitório mas sua saída e entrada da empresa será contabilizada. Para melhor entendimento desta realidade, criou-se o Quadro 9.

Quadro 8 – Volume de dados

| Mês | Volume de Registros |
|--------------|----------------------------|
| jan/17 | 153348 |
| fev/17 | 166216 |
| mar/17 | 210279 |
| abr/17 | 176808 |
| mai/17 | 331892 |
| jun/17 | 327682 |
| jul/17 | 206317 |
| ago/17 | 218353 |
| set/17 | 183747 |
| out/17 | 206577 |
| nov/17 | 205901 |
| dez/17 | 130679 |
| Total | 2517799 |

Fonte: o autor (2018).

Quadro 9 – Registros mensais

| Mês | Volume de Registros | Qt. Funcionários com Registro | Média Mensal | Média Diária |
|------------|----------------------------|--------------------------------------|---------------------|---------------------|
| jan/17 | 153348 | 1818 | 84,3 | 4,2 |
| fev/17 | 166216 | 1845 | 90,1 | 4,5 |
| mar/17 | 210279 | 1838 | 114,4 | 5,7 |
| abr/17 | 176808 | 1939 | 91,2 | 4,6 |
| mai/17 | 331892 | 2926 | 113,4 | 5,7 |
| jun/17 | 327682 | 2964 | 110,6 | 5,5 |
| jul/17 | 206317 | 1986 | 103,9 | 5,2 |
| ago/17 | 218353 | 1885 | 115,8 | 5,8 |
| set/17 | 183747 | 1856 | 99,0 | 5,0 |
| out/17 | 206577 | 1942 | 106,4 | 5,3 |
| nov/17 | 205901 | 1952 | 105,5 | 5,3 |
| dez/17 | 130679 | 1986 | 65,8 | 3,3 |

Fonte: o autor (2018).

No quadro 9 foram contabilizados todos os funcionários que realizaram marcações no sistema, englobando mão de obra direta, indireta e administrativa. Tem-se a média de registros mensais e diários por funcionário, considerando, para todos os meses, 20 dias úteis.

4.4 MODELAGEM

Em função do volume de registros considerados e do objetivo do trabalho, para a análise das informações extraídas utilizou-se o *software* de mineração de dados *open source* R x64 versão 3.5.1., aplicando-se o processo de classificação e a técnica de Bayes.

4.4.1 Classificação: NaiveBayes

Para criação do modelo, os mais de 2,5 milhões de registros extraídos das três fontes: *Microsoft Excel* (acidentes), ERP SAP (atendimentos no centro de saúde) e *Ponto Soft* (dados de acessos) foram unificados e tratados em uma planilha csv.

Considerou-se apenas os funcionários da mão de obra direta do turno diurno da companhia, visto a dificuldade de validação de parâmetros de critérios para a troca de datas em um mesmo período de trabalho.

A consolidação dos dados gerou uma tabela com 241.238 instâncias, contemplando as matrículas dos empregados e os dias do ano, ou seja, para cada funcionário da mão de obra direta e do turno diurno, criou-se o registro de dias úteis, segunda-feira à sexta-feira, de 2017.

Como atributos, foram considerados os critérios definidos no tópico 4.2 e suas variações, as possibilidades de cada critério podem ser observadas como Quadro 10.

Quadro 10 – Variações critérios (atributos)
(continua)

| Critério | Variação |
|-----------------------|---|
| Chegada | a) entrada hora extra b) entrada normal c) entrada atrasada d) sem entrada |
| Saída dia antecedente | a) saída antecipada b) saída normal c) saída hora extra d) sem saída |
| Refeição | a) alimentação interna b) sem registro |

(conclusão)

| | |
|--------------------------------|---|
| Refeição extra | a) proteína b) sobremesa c) proteína e sobremesa d) sem registro |
| Alteração do local de trabalho | a) mesmo relógio b) relógio diferente |
| Ausência de trabalho | a) ausente dia anterior b) regular |
| Atendimentos ocupacionais | a) regular b) atendimento dia anterior c) atendimento últimos sete dias d) atendimento últimos trinta dias |
| Dia da semana | a) seg b) ter c) qua d) qui e) sex |
| Quinzena do mês | a) primeira quinzena b) segunda quinzena |
| Tempo de empresa | a) menos de 1 ano b) entre 1 e 5 anos c) entre 5 e 7 anos d) entre 7 e 12 anos e) entre 12 e 38 anos |
| Idade | a) entre 18 e 27 anos b) entre 27 e 32 anos c) entre 32 e 38 anos d) entre 38 e 45 anos e) entre 45 e 61 anos |
| Classe | a) acidente b) sem acidente |

Fonte: o autor (2018).

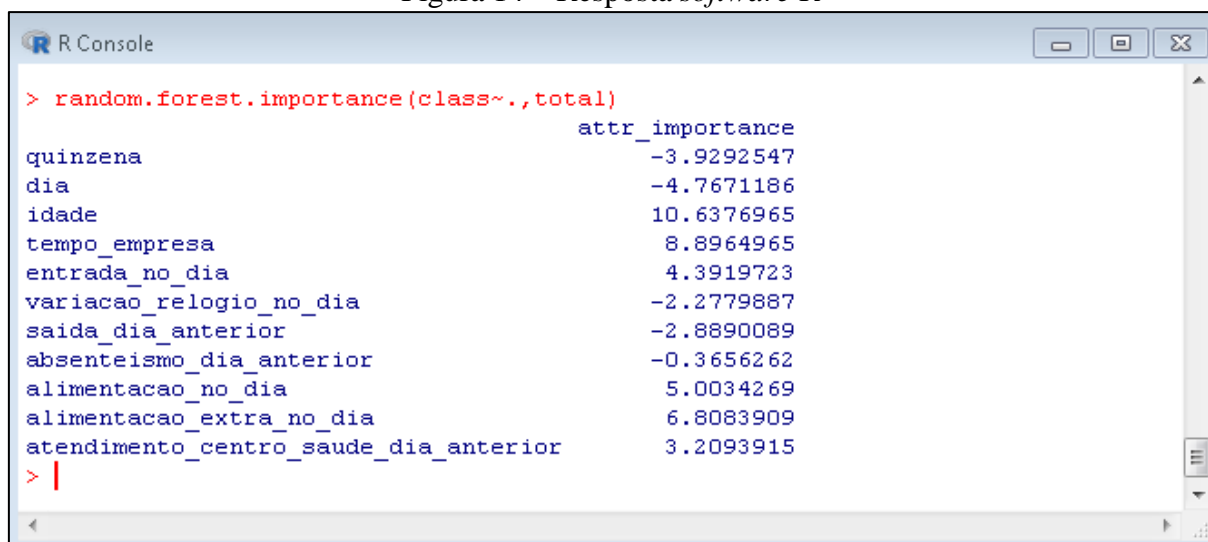
O último atributo, chamado de classe, é a variável de interesse sendo o elemento de previsão desejada.

4.4.2 Seleção de atributos

Para seleção dos atributos a serem considerados na modelagem de NaiveBayes foi utilizado o pacote “*FSelector*” do *software* R, este é baseado no algoritmo *random forest* e atribui pesos aos critérios, identificando a importância dos mesmos na classificação de uma nova instância.

Um *script* foi utilizado para implementação do algoritmo, considerando a totalidade dos dados, seu retorno no *prompt* de comando do *software* R pode ser observado como Figura 14.

Figura 14 – Resposta *software* R

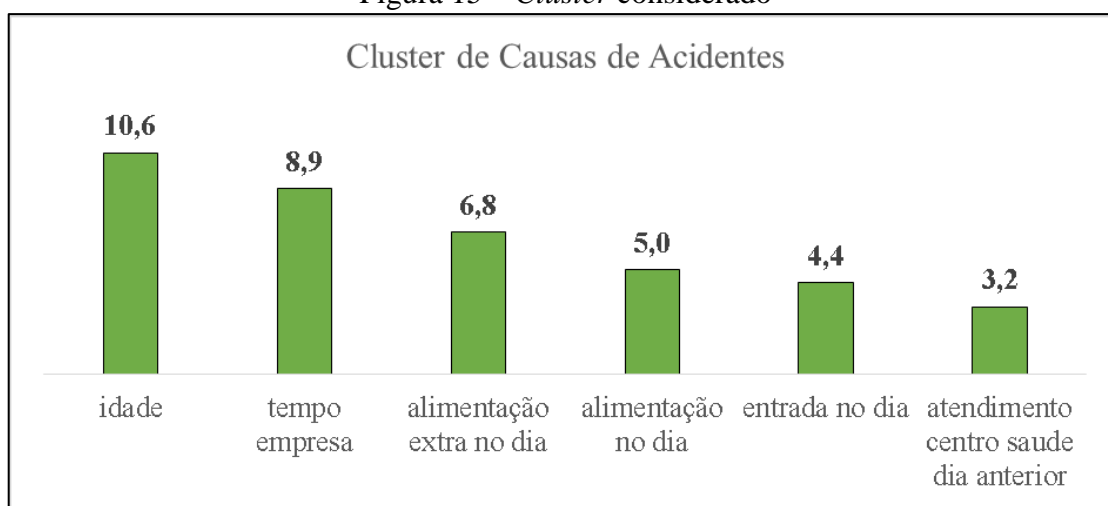


Fonte: *software* R (2018).

A seleção de atributos gerada como resultado apresentou pesos negativos, esta categorização indica critérios que não auxiliam na classificação e podem prejudicar a modelagem do algoritmo de NaiveBayes. Por tanto, os critérios de quinzena, dia, variação do relógio ponto, saída e ausência de trabalho no dia anterior, serão desconsiderados.

Desta forma, cria-se o *cluster* de possíveis causas de acidentes de trabalho, os atributos estão apresentados como Figura 15 em forma gráfica com ordem decrescente de relevâncias para classificação de uma nova instância.

Figura 15 – *Cluster* considerado



Fonte: o autor (2018).

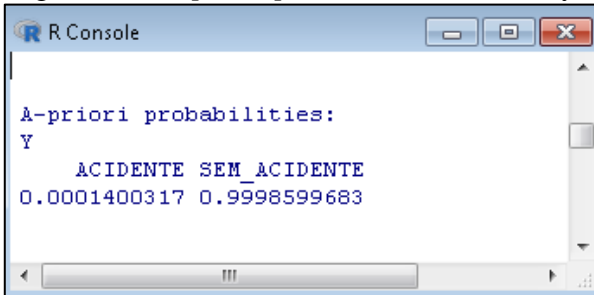
Os seis critérios apresentados foram considerados para a modelagem do algoritmo de NaiveBayes descrita a seguir.

4.4.3 Modelagem NaiveBayes

Para treino do modelo utilizando a técnica de Bayes foram segregados 70% da totalidade de instâncias dos dados agrupados, esta divisão considerou os 149.966 primeiros registros em função de sua data. Os pacotes do *software* R aplicados foram: “*foreign*” e “*e1071*”.

O modelo retornado pelo *software* R entregou dois grupos de probabilidades, o primeiro chamado de *A-priori probabilities* considera o volume total de dados e sua classificação, este parâmetro pode ser visualizado como Figura 16.

Figura 16 – *A-priori probabilities* NaiveBayes



```
R Console
A-priori probabilities:
Y
  ACIDENTE SEM_ACIDENTE
0.0001400317 0.9998599683
```

Fonte: *software* R (2018).

O valor baixo de probabilidade para acidentes pode ser entendido em função da baixa quantidade de dados com este tipo de classificação, visto que das 149.966 instâncias apenas 21 apresentam este registro.

O segundo grupo de estatística, com maior aplicação para o trabalho, entrega as probabilidades condicionais, estas avaliam o impacto da variação dos atributos selecionados sobre a classificação. Este resultado pode ser observado como Figura 17.

Figura 17 – Conditional probabilities NaiveBayes

```

R Console
Conditional probabilities:
idade
Y      ENTRE_18_E_27 ENTRE_27_E_32 ENTRE_32_E_38 ENTRE_38_E_45 ENTRE_45_E_61
ACIDENTE 0.1428571 0.2857143 0.1904762 0.2380952 0.1428571
SEM_ACIDENTE 0.2004468 0.1923572 0.2042616 0.2030678 0.1998666

tempo_empresa
Y      ENTRE_1_E_5_ANOS ENTRE_12_E_38_ANOS ENTRE_5_E_7_ANOS ENTRE_7_E_12_ANOS MENOS_DE_1_ANO
ACIDENTE 0.1904762 0.0952381 0.2380952 0.1428571 0.3333333
SEM_ACIDENTE 0.2052219 0.2355197 0.1731568 0.2271166 0.1589850

entrada_no_dia
Y      ENTRADA_ATRASADA ENTRADA_HORA_EXTRA ENTRADA_NORMAL SEM_ENTRADA
ACIDENTE 0.000000000 0.047619048 0.904761905 0.047619048
SEM_ACIDENTE 0.013938444 0.008503118 0.797592451 0.179965988

alimentacao_no_dia
Y      ALIMENTACAO_INTERNA SEM_REGISTRO
ACIDENTE 0.4285714 0.5714286
SEM_ACIDENTE 0.7809930 0.2190070

alimentacao_extra_no_dia
Y      PROTEINA PROTEINA_SOBREMESA SEM_REGISTRO SOBREMESA
ACIDENTE 0.000000000 0.000000000 0.952380952 0.047619048
SEM_ACIDENTE 0.035212911 0.008856581 0.809696889 0.146233619

atendimento_centro_saude_dia_anterior
Y      ATENDIMENTO_DIA_ANTERIOR ATENDIMENTO_ULTIMOS_30_DIAS ATENDIMENTO_ULTIMOS_7_DIAS REGULAR
ACIDENTE 0.000000000 0.047619048 0.000000000 0.952380952
SEM_ACIDENTE 0.002587615 0.045063190 0.013471606 0.938877588
> |

```

Fonte: *software* R (2018).

O retorno observado no console do *software* R representa o peso de cada variação nos atributos para classificação de uma nova instância, estas probabilidades encontradas serão discutidas no tópico de avaliação.

4.5 AVALIAÇÃO

Seguindo o método CRIS-DM, a avaliação do algoritmo NaiveBayes criado foi efetuada por meio de uma matriz de confusão com os 30% de registros restantes, contabilizando 64.272 instâncias, este processo foi realizado no *software* R e seu retorno pode ser observado como Figura 18.

Figura 18 – Retorno matriz de confusão

```

R Console
      predicao
      ACIDENTE SEM_ACIDENTE
ACIDENTE      0         8
SEM_ACIDENTE  0      64264
> taxaacerto = (confusao[1] + confusao[4]) / sum(confusao)
> taxaerro = (confusao[2] + confusao[3]) / sum(confusao)
> taxaacerto
[1] 0.9998755
> taxaerro
[1] 0.000124471
> |

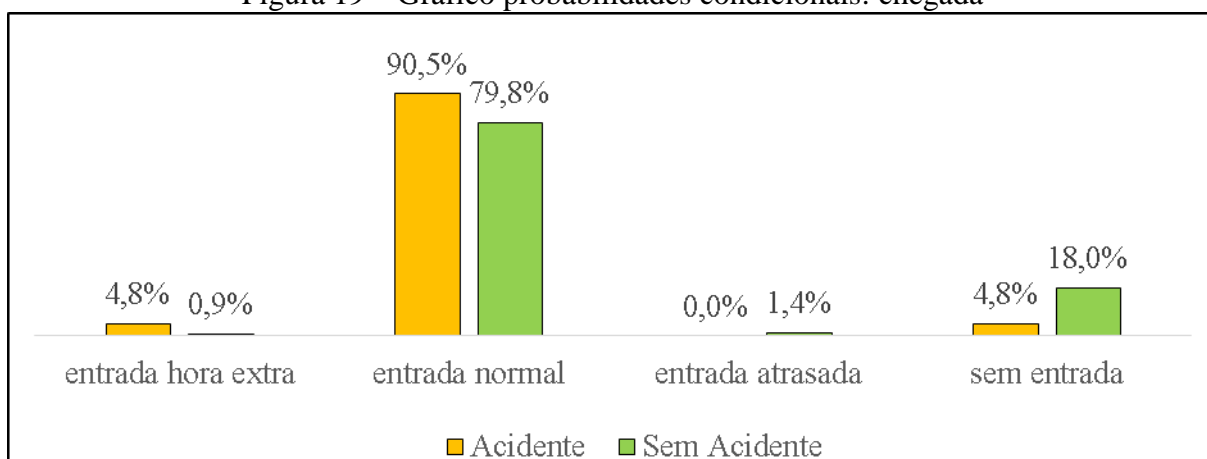
```

Fonte: software R (2018).

Em função da baixa probabilidade *a-priori*, de 0,00014, era esperada a dificuldade de se prever acidentes com o algoritmo, todas as instâncias de teste foram classificadas como sem acidente, sendo destas, oito falsos-negativos, onde acidentes aconteceram. Em consequência ao elevado valor de instâncias reais sem acidente a taxa de acerto alcançada foi de 99,99% e a taxa de erro de 0,012%, contemplando as 8 ocorrências não previstas dos 64.272 registros.

Entretanto, o modelo entregou as probabilidades condicionais para cada critério considerado, podendo-se avaliar o perfil de acidentados da mão de obra direta dentro da companhia. Para melhor visualização, estas probabilidades estão ilustradas graficamente nas imagens a seguir, inicialmente observa-se o critério chegada como Figura 19.

Figura 19 – Gráfico probabilidades condicionais: chegada

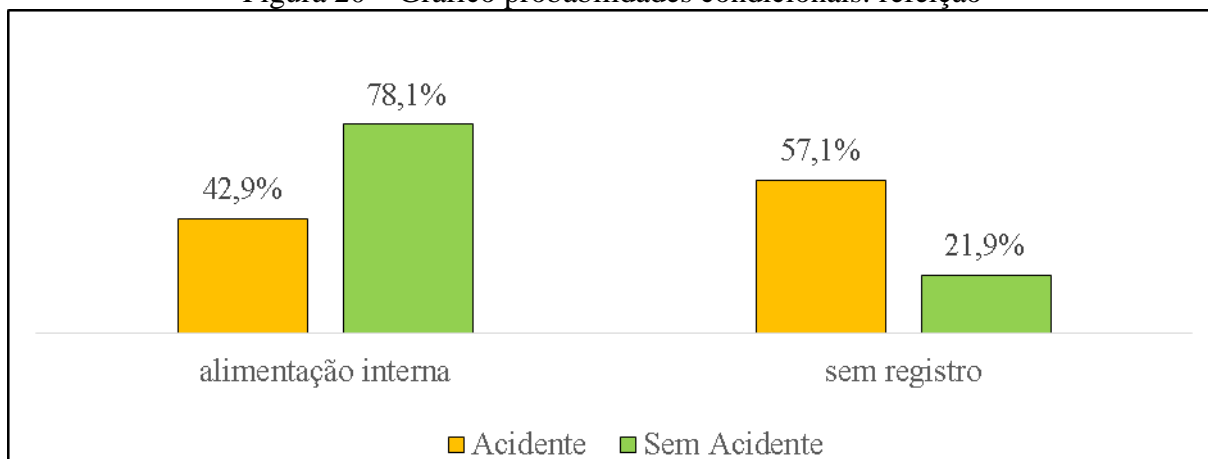


Fonte: o autor (2018).

Para ocorrência de acidentes a variação com maior porcentagem é a de entrada normal do funcionário, com 90,5% de peso sobre o resultado do algoritmo, valor inferior a classificação sem acidente, com 79,8%. Esta relação ocorre também na entrada em hora extra, porém com

uma diferença de 460%, enquanto a entrada normal apresenta 13%. A Figura 20 apresenta a mesma análise para o critério refeição.

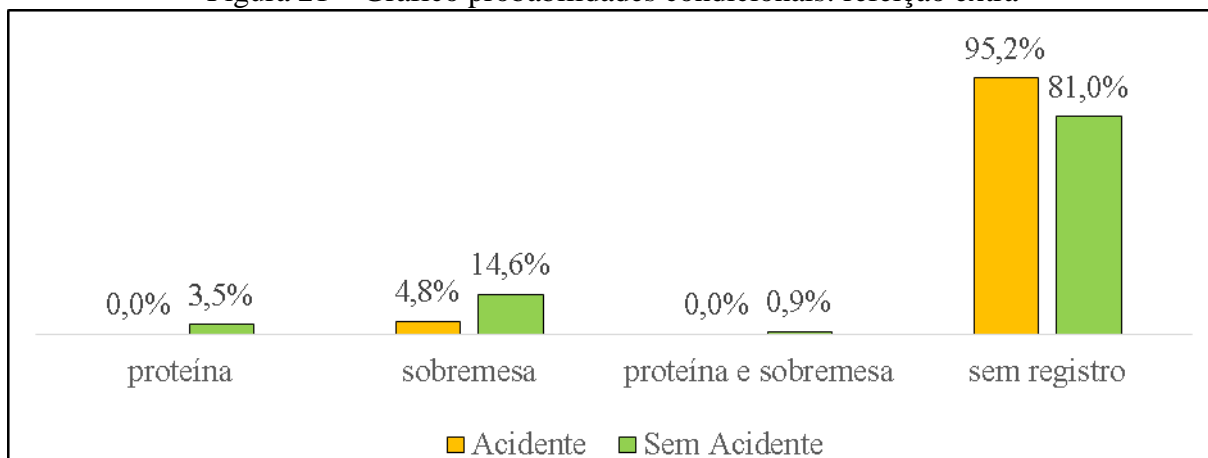
Figura 20 – Gráfico probabilidades condicionais: refeição



Fonte: o autor (2018).

Na avaliação do impacto sobre o atributo refeição, o maior peso está relacionado à variação de sem registro, ou seja, quando o funcionário não realiza sua alimentação no ambiente oferecido pela companhia, seu valor é de 57,1%. A diferença para a classificação de sem acidente é de 161%. Para a refeição extra o gráfico ilustrado como Figura 21 foi desenvolvido.

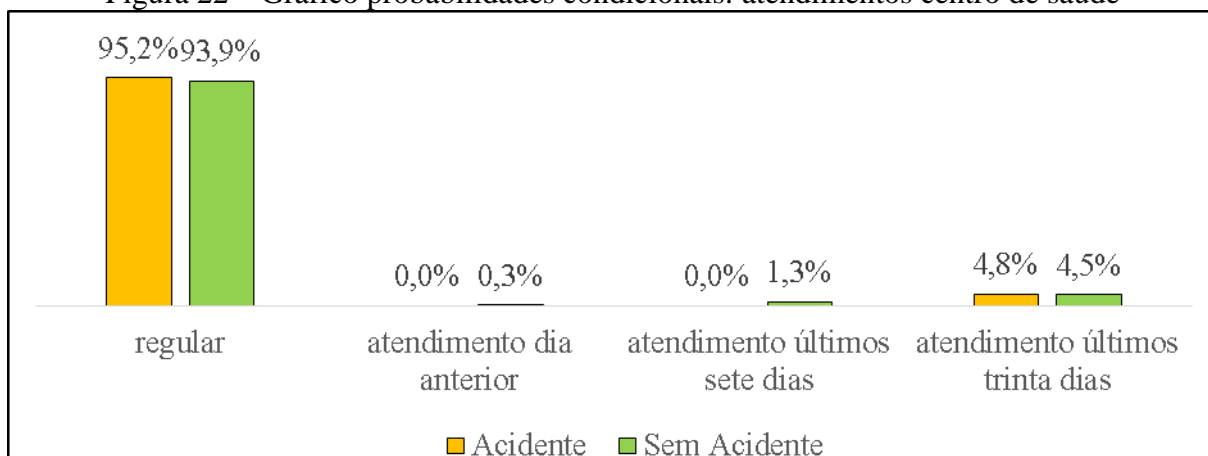
Figura 21 – Gráfico probabilidades condicionais: refeição extra



Fonte: o autor (2018).

O peso de 95,2% com uma diferença de 18% foi encontrado para o critério de sem registro, ou seja, quando o funcionário não consome alimentação extra oferecida no refeitório da empresa. Em relação aos atendimentos no centro de saúde, a Figura 22 pode ser observada.

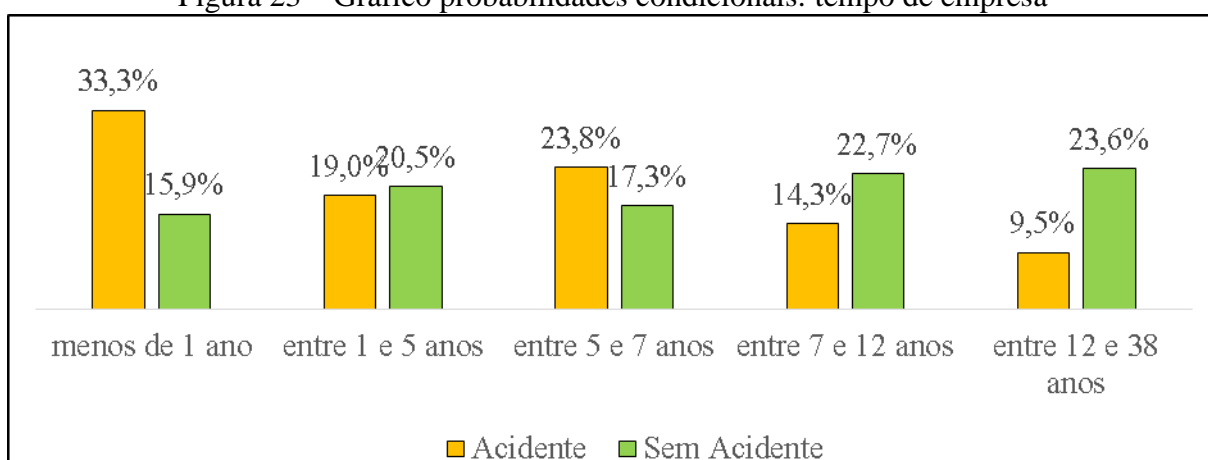
Figura 22 – Gráfico probabilidades condicionais: atendimentos centro de saúde



Fonte: o autor (2018).

Duas variações de atributos apresentaram maior peso para classificação de acidentes, sendo o regular, onde não há registros de atendimentos, com valor de 95,2% e diferença de 1% e o atendimento nos últimos trinta dias com 4,8% e diferença ao peso de sem acidentes de 6%. Para o tempo de empresa, os grupos estão apresentados como Figura 23.

Figura 23 – Gráfico probabilidades condicionais: tempo de empresa

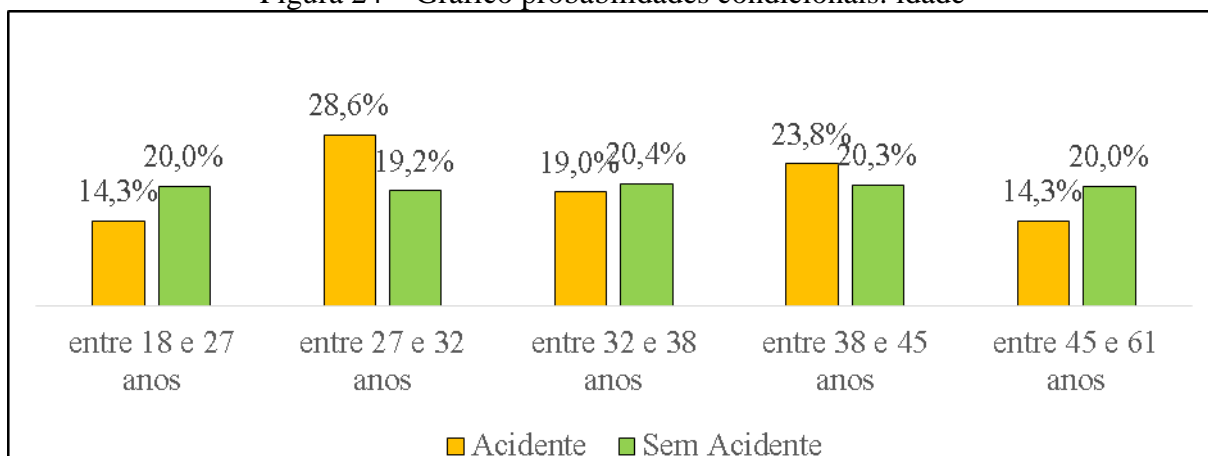


Fonte: o autor (2018).

O grupo com maior peso é observado entre os funcionários com menos de um ano de atividade dentro da companhia, com valor de 33,3% estes apresentam diferença de 110%. Um segundo grupo apresenta peso maior para classificação de acidentes do que sem acidentes, o de empregados com tempo de empresa entre 5 e 7 anos, com variação de 38%.

Por fim, as variações de idade podem ser consideradas como Figura 24.

Figura 24 – Gráfico probabilidades condicionais: idade



Fonte: o autor (2018).

Para a idade, os funcionários entre 27 e 32 anos e os de 38 e 45 anos possuem maior probabilidade de sofrer um acidente, os primeiros com 28,6% e diferença de 49% e os segundos com 23,8% e variação de 17%.

Como pode ser observado, os atributos de maior relevância estão relacionados ao tempo de empresa dos funcionários, à realização de refeição no ambiente externo à empresa e à efetivação de horas extras do dia observado. Estes apresentam respectivamente às diferenças de 110%, 161% e 460% entre as classificações de acidentes ou sem acidente. Estes resultados foram apresentados aos gestores envolvidos em um fórum específico descrito no tópico a seguir.

4.6 APRESENTAÇÃO

Como última etapa do método CRISP-DM considerada na delimitação de escopo do presente trabalho, tem-se a apresentação do modelo criado para a gestão da companhia, esta considerará e efetivará a implementação do algoritmo dentro da empresa caso entenda como válido o estudo realizado.

Para tanto, uma reunião marcada previamente foi concretizada, utilizou-se o aplicativo *Microsoft PowerPoint* para demonstrar a efetivação das etapas do processo padrão inter-indústrias na mineração de dados. A reunião contou com cinco participantes, o Quadro 11 apresenta e caracteriza os integrantes do fórum.

O primeiro tema discutido foi a seleção de atributos realizada pelo algoritmo, demonstrando os critérios desconsiderados pelo modelo, estes sendo: saída dia antecedente, alteração do local de trabalho, ausência de trabalho, dia da semana e quinzena do mês. Estas variáveis foram apontadas por participantes deste evento, na reunião de formulação de critérios.

Desta forma, o algoritmo foi capaz de desfazer de forma quantitativa, caracterizações não reais consideradas pelos gestores da companhia.

O segundo assunto abordado entre os participantes foi o perfil de acidentados criados pelo algoritmo, este contemplando funcionários com 27 a 30 anos, como menos de 1 ano de empresa, sem atendimentos no centro de saúde, com entrada em horário regular e refeição realizada fora da companhia sem consumir alimentação extra.

Quadro 11 – Reunião de apresentação modelo

| Cargo | Tempo no Cargo | Formação | Data | Local Reunião | Tempo de Duração |
|--------------------------------------|-----------------------|---|-------------|--|-------------------------|
| Gerente Industrial | 6 anos | Tecnólogo em Automação Industrial Mestrando Engenharia de Produção | 24/05/2018 | Sala de Reuniões da Engenharia de Processos em Caxias do Sul | 1 hora |
| Gerente Industrial | 1 ano | Engenheiro de Produção Mestre em Engenharia Mecânica | | | |
| Coordenador de Engenharia Industrial | 2 anos | Engenheiro Mecânico | | | |
| Coordenador de Segurança | 1 ano | Engenheiro Mecânico | | | |
| Engenheiro de Segurança | 7 anos | Engenheiro Químico Pós graduado em Segurança do Trabalho | | | |

Fonte: o autor (2018).

A partir da descrição construída, a gerência responsável solicitou a revisão dos treinamentos de segurança para novos funcionários e a criação de um procedimento de gestão para diálogos de segurança com operários após a realização de refeição no ambiente externo à empresa. Esta solicitação demandará a criação de um sistema de acompanhamento de refeições dos funcionários, algo a ser desenvolvido pelo setor de Segurança, Saúde e Meio Ambiente em conjunto à área de Tecnologia da Informação da companhia.

5 CONCLUSÃO

O objetivo deste trabalho consistiu na criação de um modelo de predição de ocorrências de acidentes aplicando a técnica da Bayes, especificamente o algoritmo de NaiveBayes, no *software* de mineração de dados R. Os resultados obtidos foram avaliados e considerados para criação de um perfil de funcionários acidentados dentro da empresa Randon S/A Implementos e Participações de Caxias do Sul. Com a descrição construída, espera-se prever e direcionar ações buscando a redução de infortúnios e danos pessoais aos colaboradores, de mesma forma, iniciar e criar a cultura de trabalhos analíticos com o intuito de prever acontecimentos por meio do vasto banco de dados disponível dentro da companhia.

O método aplicado, CRISP-DM, teve como etapas a análise do negócio, entendimento e preparação dos dados, modelagem, avaliação e apresentação. O algoritmo criado, em função de sua baixa probabilidade *a-priori*, não foi capaz de prever acidentes considerando os dados de teste. Porém, a avaliação de critérios relacionados às casualidades de 2017 clarificou o impacto de atributos na ocorrência de eventos indesejados, este conhecimento possuía característica exclusivamente empírica, baseado somente na experiência da gestão da área de Segurança, Saúde e Meio Ambiente.

O modelo criado descartou os critérios antes considerados de quinzena, dia da semana, variação do relógio ponto, saída e ausência de trabalho no dia anterior, demonstrando numericamente que tais comportamentos não impactam na ocorrência de acidentes e desfez considerações existentes dentro da área de SSM sobre a importância de atributos específicos. Entretanto, os aspectos de idade, tempo de empresa, início de atividade, refeição, consumo de refeição extra e atendimentos ocupacionais, resultaram em um perfil único, que por meio de pesos definiu e caracterizou funcionários com maior probabilidade de acidentes. Esta descrição, apresentada para empresa, pôde ser definida como: colaboradores com 27 a 30 anos, como menos de 1 ano de empresa, sem atendimentos no centro de saúde, com entrada em horário regular e refeição realizada fora da companhia sem consumir alimentação extra.

A caracterização teve grande aceitação na reunião de apresentação do método para gestão, direcionando duas ações principais buscando-se a redução de acontecimentos: revisão de treinamentos em segurança para novos funcionários, relacionando o atributo de tempo de empresa; e criação de sistema para identificação da realização de refeição fora do refeitório da companhia, possibilitando que a gestão direta do empregado avalie e oriente o comportamento seguro do mesmo.

A maior dificuldade encontrada na implementação do trabalho esteve relacionada à extração e preparação dos dados, visto o volume de registros e a falta de automação e relações entre os sistemas. Porém, o projeto deixa perspectivas de movimentos futuros e lições aprendidas para criação de modelos de análise de dados com o objetivo de prever acontecimentos, aproveitando-se o conhecimento implícito já existente na companhia.

Por meio da conceituação de ciência de dados, é possível afirmar que os objetivos do trabalho foram alcançados, já que os dados antes coletados e armazenados foram analisados criando-se a informação, esta informação apresentada gerou o conhecimento que aplicado prediz o perfil de acidentados dentro da empresa Randon.

REFERÊNCIAS

AMARAL, Fernando. **Introdução à ciência de dados**: mineração de dados e big data. Rio de Janeiro: Alta Books, 2016.

BRAGA, Antônio de Pádua; CARVALHO, André Ponce de Leon F. de; LUDERMIR, Teresa Bernarda. **Redes neurais artificiais**: teoria e aplicações. Rio de Janeiro: Livros Técnicos e Científicos Editora S.A., 2000.

DAVENPORT, Thomas H.; HARRIS, Jeanne G. **Competição analítica**: vencendo através da nova ciência. Rio de Janeiro: Elsevier, 2007.

DEBUSE, J. C. W.. Extending data mining methodologies to encompass organizational factors. **Systems Research and Behavioral Science**, Brisbane, v. 24, n. 2, p.183-190, maio 2007. Disponível em: <<http://rdcu.be/yoM4/>>. Acesso em: 11 out. 2017.

FERNÁNDEZ, Alberto et al. Big data with cloud computing: an insight on the computing environment, mapreduce, and programming frameworks. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, [s.l.], v. 4, n. 5, p. 380-409, set. 2014. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.706.4988&rep=rep1&type=pdf>>. Acesso em: 17 set. 2017.

FREDERICK, Donna Ellen. Libraries, data and the fourth industrial revolution: data deluge column. **Library Hi Tech News**, Bingley, v. 33, n. 1, p. 9-12, jul. 2016. Disponível em: <<https://doi.org/10.1108/LHTN-05-2016-0025>>. Acesso em: 19 ago. 2017.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas S.A., 2008.

HART, Michele. **Dashboard tiles in power BI**. 2017. Disponível em: <<https://powerbi.microsoft.com/en-us/documentation/powerbi-service-dashboard-tiles/>>. Acesso em: 08 out. 2017.

LI, Guoping; HOU, Yun; WU, Aizhi. Fourth industrial revolution: technological drivers, impacts and coping methods. **Chinese Geographical Science**, Changchun, v. 27, n. 4, p. 626-637, jul. 2017. Disponível em: <<https://link-springer-com.ez314.periodicos.capes.gov.br/content/pdf/10.1007%2Fs11769-017-0890-x.pdf>>. Acesso em: 17 set. 2017.

MORAES, Rafael Ferraz de; SALES, Paulo Roberto de Holanda; COSTA, José Fabiano da Serra. Utilização do software estatístico R na previsão de séries temporais: série prêmio de resseguro. **Cadernos do IME – Série Estatística**. Rio de Janeiro, v. 30, n. 1, p. 1-17, jun. 2011. Disponível em: <<http://www.e-publicacoes.uerj.br/index.php/cadest/article/view/15727/11898>>. Acesso em: 10 out. 2017.

MUTULA, Stephen. Big data industry: implication for the library and information sciences. **African Journal of Library, Archives and Information Science**, Ibadan, v. 26, n. 2, p. 93-96, out. 2016. Disponível em: <<http://web.b-ebsohost->

com.ez314.periodicos.capes.gov.br/ehost/pdfviewer/pdfviewer?vid=1&sid=6be3f474-6f1b-44ec-89b0-033c3aeb14e0%40sessionmgr4008> Acesso em 24 set. 2017.

PADUA, Mariana Cantisani; DIAS, Guilherme Ataíde; LIMA, Thiago Lucas Castor de. Dados, formas, cores e informação: um estudo sobre construção e análise na infografia. **Liinc em Revista**, Rio de Janeiro, v. 11, n. 1, p. 287-309, maio 2015. Disponível em: <<http://revista.ibict.br/liinc/article/view/3627/3092>>. Acesso em 01 out. 2017.

RABELO, Emerson; CAMPOS, Fernando Celso de. Big data e kdd: novas descobertas. In: XXXIV ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 34., 2014, Curitiba. **Anais...** Disponível em: <http://www.abepro.org.br/biblioteca/enegep2014_TN_STP_202_144_25705.pdf>. Acesso em: 17 set. 2017.

RENNOLLS, Keith; AL-SHAWABKEH, Abdallah. Formal structures for data mining, knowledge discovery and communication in a knowledge management environment. **Intelligent Data Analysis**, Amsterdam, v. 12, n.2, p. 147-163, 2008.

SCHWAB, Klaus. **The fourth industrial revolution**. Nova York: Crown Business, 2017.

SILVEIRA, Marcio; MARCOLIN, Carla Bonato; FREITAS, Henrique Mello. Uso corporativo do big data: uma revisão de literatura. **Revista de Gestão e Projetos**, São Paulo, v. 6, n. 3, p. 44-59, dez. 2015. Disponível em: <<http://www.revistagep.org/ojs/index.php/gep/article/view/369/pdf>>. Acesso em: 24 set. 2017.

STĂNCIOIU, Alin. The fourth industrial revolution: industry 4.0. **Fiabilitate și Durabilitate**, Târgu Jiu, n. 1, p. 74-78, maio 2017. Disponível em: <[http://www.utgjiu.ro/rev_mec/mecanica/pdf/2017-01/11_Alin STĂNCIOIU - THE FOURTH INDUSTRIAL REVOLUTION „INDUSTRY 4.0”.pdf](http://www.utgjiu.ro/rev_mec/mecanica/pdf/2017-01/11_Alin_STĂNCIOIU_-_THE_FOURTH_INDUSTRIAL_REVOLUTION_„INDUSTRY_4.0”_pdf)>. Acesso em: 13 ago. 2017.

TURBAN, Efraim et al. **Business intelligence**: um enfoque gerencial para a inteligência do negócio. Porto Alegre: Bookman, 2009.

VICTORINO, Marcio de Carvalho et al. Uma proposta de ecossistema de big data para a análise de dados abertos governamentais conectados. **Informação e Sociedade: Estudos**, João Pessoa, v. 27, n. 1, p. 225-242, jan. 2017. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/29299/17505>>. Acesso em 30 set. 2017.

APÊNDICE A – REGISTROS ACPT RANDON IMPLEMENTOS CAXIAS DO SUL

(continua)

| Alerta | Data | Horário | Mês | Matrícula |
|---------------|-------------|----------------|------------|------------------|
| 1 | 12/01/2017 | 22:00 | Janeiro | 6946 |
| 4 | 19/01/2017 | 14:00 | Janeiro | 22896 |
| 12 | 15/02/2017 | 15:30 | Fevereiro | 6230 |
| 13 | 20/02/2017 | 18:15 | Fevereiro | 16435 |
| 14 | 20/02/2017 | 18:30 | Fevereiro | 11645 |
| 16 | 23/02/2017 | 12:50 | Fevereiro | 8482 |
| 18 | 09/03/2017 | 21:50 | Março | 4894 |
| 20 | 21/03/2017 | 01:20 | Março | 1648 |
| 21 | 28/03/2017 | 01:00 | Março | 8228 |
| 22 | 04/04/2017 | 19:30 | Abril | 3293 |
| 27 | 26/04/2017 | 10:30 | Abril | 12152 |
| 30 | 15/05/2017 | 14:10 | Maio | 13125 |
| 32 | 22/05/2017 | 08:40 | Maio | 22896 |
| 34 | 24/05/2017 | 16:00 | Maio | 2173 |
| 35 | 25/05/2017 | 09:30 | Maio | 17816 |
| 36 | 29/05/2017 | 09:00 | Maio | 20949 |
| 41 | 14/06/2017 | 01:45 | Junho | 23024 |
| 42 | 22/06/2017 | 08:20 | Junho | 5546 |
| 47 | 30/06/2017 | 07:50 | Junho | 22912 |
| 49 | 14/07/2017 | 09:00 | Julho | 15405 |
| 50 | 18/07/2017 | 14:45 | Julho | 10527 |
| 53 | 25/04/2017 | 19:20 | Abril | 6364 |
| 54 | 31/07/2017 | 14:00 | Julho | 5884 |
| 55 | 31/07/2017 | 08:15 | Julho | 10605 |
| 56 | 31/07/2017 | 10:25 | Julho | 6466 |
| 58 | 02/08/2017 | 11:10 | Agosto | 22950 |
| 59 | 17/08/2017 | 22:30 | Agosto | 19732 |
| 60 | 21/08/2017 | 10:00 | Agosto | 1651 |
| 61 | 22/08/2017 | 11:30 | Agosto | 22937 |
| 62 | 28/08/2017 | 10:40 | Agosto | 23601 |
| 63 | 28/08/2017 | 17:40 | Agosto | 22932 |
| 65 | 02/09/2017 | 16:15 | Setembro | 22447 |
| 66 | 05/09/2017 | 08:50 | Setembro | 22539 |
| 67 | 12/09/2017 | 11:40 | Setembro | 5966 |
| 68 | 13/09/2017 | 15:20 | Setembro | 13625 |
| 69 | 18/09/2017 | 18:50 | Setembro | 5753 |
| 70 | 19/09/2017 | 10:30 | Setembro | 24783 |
| 71 | 20/09/2017 | 00:10 | Setembro | 11614 |



(conclusão)

| | | | | |
|----|------------|-------|----------|-------|
| 73 | 25/09/2017 | 10:10 | Setembro | 1856 |
| 74 | 26/09/2017 | 16:15 | Setembro | 20946 |
| 75 | 28/09/2017 | 01:20 | Setembro | 1849 |
| 76 | 28/09/2017 | 01:45 | Setembro | 691 |
| 79 | 05/10/2017 | 16:00 | Outubro | 7576 |
| 80 | 05/10/2017 | 16:45 | Outubro | 22538 |
| 81 | 10/10/2017 | 15:40 | Outubro | 23037 |
| 82 | 17/10/2017 | 20:00 | Outubro | 17831 |
| 88 | 20/11/2017 | 18:30 | Novembro | 22910 |
| 90 | 21/11/2017 | 03:30 | Novembro | 6946 |
| 91 | 23/11/2017 | 08:40 | Novembro | 1376 |
| 92 | 24/11/2017 | 20:00 | Novembro | 2430 |
| 94 | 29/11/2017 | 09:15 | Novembro | 11545 |
| 95 | 04/12/2017 | 23:00 | Dezembro | 20535 |
| 97 | 20/12/2017 | 16:50 | Dezembro | 23800 |

APÊNDICE B – REGISTRO ATENDIMENTOS CENTRO DE SAÚDE

| Data | Hora | Matrícula | CT | Prest.Assistência |
|-------------|-------------|------------------|-----------|--------------------------|
| 05/01/2017 | 11:50 | 17937 | LPE | Contusão |
| 10/01/2017 | 2:00 | 19774 | LAS | Escoriação |
| 10/01/2017 | 12:00 | 12267 | ENG | Contusão |
| 10/01/2017 | 21:00 | 20936 | DPR | Escoriação |
| 10/01/2017 | 21:00 | 13731 | LMS | Escoriação |
| 11/01/2017 | 0:10 | 6943 | DOB | Contusão |
| 11/01/2017 | 0:30 | 22968 | LDE | Ferimento Cortante |
| 11/01/2017 | 16:20 | 491 | LNO | Escoriação |
| 11/01/2017 | 22:30 | 13486 | PRE | Escoriação |
| 12/01/2017 | 8:20 | 10605 | LMM | Queimadura 1º Grau |
| 12/01/2017 | 14:00 | 651 | LMS | Entorse |
| 12/01/2017 | 19:30 | 13325 | LDE | Contusão |
| 12/01/2017 | 22:00 | 6946 | LAS | Contusão |
| 13/01/2017 | 0:40 | 11230 | LCC | Contratura Muscular |
| 13/01/2017 | 9:40 | 5807 | LDE | Queimadura Ocular |
| 13/01/2017 | 10:35 | 7114 | LPE | Queimadura Ocular |
| 14/01/2017 | 0:00 | 1277 | LMM | Queimadura Ocular |
| 14/01/2017 | 0:30 | 23055 | SPE | Corpo Estranho no Olho |
| 16/01/2017 | 9:00 | 9377 | LPE | Ferimento Corto-Contuso |
| 16/01/2017 | 10:00 | 2446 | LDE | Contusão |
| 16/01/2017 | 13:15 | 20496 | TRA | Ferimento Cortante |
| 17/01/2017 | 1:10 | 22940 | PLM | Contusão |
| 17/01/2017 | 9:00 | 17823 | LSE | Queimadura 1º Grau |
| 17/01/2017 | 10:30 | 7146 | PTE | Corpo Estranho no Olho |
| 17/01/2017 | 11:30 | 2996 | LVF | Queimadura 1º Grau |
| 17/01/2017 | 12:40 | 22917 | SPI | Corpo Estranho no Olho |
| 17/01/2017 | 15:20 | 23060 | LDO | Contusão |
| 17/01/2017 | 23:00 | 22978 | LVF | Queimadura 1º Grau |
| 18/01/2017 | 9:55 | 22932 | MBA | Queimadura 1º Grau |
| 18/01/2017 | 15:30 | 4357 | LDE | Escoriação |
| 18/01/2017 | 15:30 | 497 | MOD | Contusão |
| 18/01/2017 | 18:00 | 22916 | LVF | Corpo Estranho no Olho |
| 19/01/2017 | 2:00 | 2410 | LDE | Queimadura Ocular |
| 19/01/2017 | 7:20 | 23060 | LDO | Queimadura Ocular |
| 19/01/2017 | 11:50 | 11001 | PRE | Contusão |
| 19/01/2017 | 14:00 | 22896 | LCH | Contusão |
| 20/01/2017 | 0:00 | 22898 | LDE | Corpo Estranho no Olho |
| 20/01/2017 | 1:15 | 22916 | LVF | Corpo Estranho no Olho |
| 20/01/2017 | 1:20 | 6240 | LMM | Contusão |

ANEXO A – ALERTA DE SEGURANÇA

|  | | ALERTA SSMA <i>Segurança, Saúde e Meio Ambiente</i> | |
|---|---|---|-----------------------------|
| <i>Planta Industrial: Randon Implementos Caxias do Sul</i> | | <i>Nº</i> | <i>080/2017</i> |
| Evento | | | |
| <input checked="" type="checkbox"/> | Acidente com perda de tempo | <input type="checkbox"/> | Acidente com dano material |
| <input type="checkbox"/> | Acidente sem perda de tempo | <input type="checkbox"/> | Acidente com dano ambiental |
| <input type="checkbox"/> | Doença Ocupacional | | |
| *Caso não esteja definido o evento por motivo de avaliação médica | | | |
| Setor: | SSS | Seção: | LMS |
| Equipamento: | Dispositivo de acoplamento | Local: | Box 03 |
| Gerente: | Rodrigo Fantinel | Coordenador: | Nataníel Depont |
| Contato SSMA: | Diego Santos de Quadros | Ramal SSMA: | 9-9974-6435 |
| Data: | 05/10/2017 | Horário: | 16:45hs |
| Descrição: | Funcionário matrícula 22538 estava realizando montagem de suspensão (acoplamento de conjunto de suspensão no produto), quando ao manusear o gabarito que realiza o acoplamento, ocorreu, de forma brusca, o movimento pendular do conjunto, vindo a atingir 1º dedo da mão E contra o gabarito. | | |
| Fotos: |  | | |
| Condições gerais do acidentado: | Funcionário atendido no Centro de Saúde encaminhado ao HCO para atendimento. | | |
| Ações de contenção: | Instalação de calço no gabarito a fim de eliminar o risco de deslocamento da suspensão; | | |
| Relato | Empresas | | |

