

**UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE EDUCAÇÃO ONLINE
ESPECIALIZAÇÃO EM BIG DATA, DATA SCIENCE E DATA ANALYTICS**

Alexandre Ribeiro Frasson

**DESCOBRINDO CONHECIMENTO EM BASE DE DADOS DE
OCORRÊNCIAS DE MANUTENÇÃO DE UMA INDÚSTRIA DE CELULOSE**

São Leopoldo / RS

2019

**UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE EDUCAÇÃO ONLINE
ESPECIALIZAÇÃO EM BIG DATA, DATA SCIENCE E DATA ANALYTICS**

Alexandre Ribeiro Frasson

**DESCOBRINDO CONHECIMENTO EM BASE DE DADOS DE
OCORRÊNCIAS DE MANUTENÇÃO DE UMA INDÚSTRIA DE CELULOSE**

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção do título de Especialista em *Big Data*, *Data Science* e *Data Analytics*, pelo curso de Pós-Graduação Lato Sensu em *Big Data*, *Data Science* e *Data Analytics* da Universidade do Vale do Rio dos Sinos – UNISINOS.

Orientador: Profa. Dra. Rosemary Francisco

São Leopoldo / RS

2019

DESCOBRINDO CONHECIMENTO EM BASE DE DADOS DE OCORRÊNCIAS DE MANUTENÇÃO DE UMA INDÚSTRIA DE CELULOSE

Alexandre Ribeiro Frasson

Universidade do Vale do Rio dos Sinos (UNISINOS)

Av. Unisinos, 950, Bairro Cristo Rei, São Leopoldo, RS – Brasil

Resumo. *O desafio atual da manutenção e gestão de ativos em indústrias de processo contínuo é antecipar-se às falhas para evitar a indisponibilidade e conseqüente perdas de produção. Conseqüentemente, novos conhecimentos, que podem ser obtidos a partir da exploração de dados armazenados em sistemas informatizados, são requeridos para apoiar nesta função. A extração de conhecimentos de uma base de dados para fins estratégicos do negócio, no entanto, não se obtém de forma direta, é necessário utilizar metodologias e ferramentas adequadas e efetivas. Nesse sentido, foi desenvolvido um estudo de caso em uma indústria de celulose da região do nordeste brasileiro com o objetivo de aplicar a metodologia de descoberta do conhecimento sob base de dados visando identificar oportunidades para a melhoria da gestão da manutenção. Os resultados obtidos apresentam descobertas importantes como a identificação precisa das principais áreas, processos, disciplinas da manutenção e, sobretudo, os equipamentos que mais impactam em termos de frequência de anomalias e perdas de produção ao longo dos anos investigados. Através de algoritmos classificadores, também é possível obter dados faltantes, tão importantes para a melhoria das informações e análises da função manutenção.*

Abstract. *The current challenge of maintenance and asset management in continuous process industries is to anticipate failures to avoid downtime and consequent production losses. Consequently, new knowledge, which may be obtained from the exploitation of data stored in computerized systems, is required to support this function. The extraction of knowledge from a database for strategic business purposes, however, is not obtained directly, it is necessary to use appropriate and effective methodologies and tools. In this sense, a case study was developed in a pulp industry in the northeastern region of Brazil with the objective of applying the knowledge discovery in databases methodology to identify opportunities to improve maintenance management. The obtained results present important discoveries such as the precise identification of the main areas, processes, maintenance disciplines and, above all, the equipment that most impact in terms of frequency of anomalies and production losses over the years investigated. Through classifying algorithms, it is also possible to obtain missing data, so important for the improvement of information and analysis of the maintenance function.*

1. Introdução

Em uma grande empresa de produção de celulose, objeto da presente pesquisa e que aqui será nomeada empresa CEL, a manutenção assume relevância equivalente à produção. Isso é uma característica das empresas de processo contínuo, ou seja, “quando a matéria prima entra num lado do sistema e o produto final sai do outro lado continuamente” [Ribeiro 2001, p.17]. É o caso da maioria das empresas de celulose, petroquímica e siderurgia. Trabalhando em conjunto, produção e manutenção, devem garantir e entregar os produtos com a quantidade e prazos indicados, qualidade requerida, com custos e produtividade competitivos, o mínimo de impacto ambiental e em conformidade com a legislação, além de perseguir continuamente a redução zero de acidentes de trabalho.

Devido à importância estratégica na estrutura das empresas com reflexos diretos ao nível de operação e logística, o enfoque atual da manutenção e gestão de ativos é de atuar na prevenção das falhas [Fuentes 2006] evitando a quebra ou indisponibilidade dos equipamentos, que sempre trazem algum risco ambiental ou de acidente de trabalho. Com isso, evita-se paradas de produção garantindo a continuidade operacional sem interrupções por motivos de falhas em equipamentos. Ou seja, antecipar-se às falhas diminuindo a taxas mínimas de manutenções corretivas em equipamentos de alta criticidade é o grande desafio da gestão da manutenção e ativos. E isso está no sentido oposto ao do paradigma de que a execução de um bom reparo significava sucesso na manutenção [Pinto e Ribeiro 2002].

A perda de produção proveniente de uma parada inesperada em uma determinada linha de produção é um dos principais problemas enfrentados pelas indústrias. Consertar uma máquina após sua quebra não é viável, sendo mais sensato evitar que isto aconteça, fazendo as inspeções preventivas ou preditivas [Kardec 2001]. Por isso, um dos principais indicadores de performance da indústria de processo contínuo é a disponibilidade operacional, que deve ser garantida por estratégias, políticas e planos de uma manutenção essencialmente proativa.

A empresa CEL, por exemplo, mesmo com todos os esforços das manutenções preventivas, preditivas ou proativas, vem tendo uma perda média anual de mais de 25 mil toneladas de celulose seca ao ar (TSA) ou no inglês, *air dried ton (ADT)*, que é a unidade padrão adotada por todo o segmento para essa commodity [Bachmann & Associados 2013]. Isso significa, com base nos preços atuais da TSA, um prejuízo anual de aproximadamente USD 25 milhões.

Na maioria dessas empresas, assim como na empresa CEL, toda ocorrência de defeito ou falha em equipamentos críticos, ou seja, aqueles que provocam perdas de produção, redução da qualidade do produto, impacto ambiental e acidente de trabalho é considerado uma anomalia, devendo ser analisada utilizando metodologias adequadas para identificação de suas causas, bem como a definição das ações corretivas. Tudo isso, visa impedir que se repitam no futuro. Tais tratamentos ou análises e seus respectivos planos de ações e acompanhamentos são amplamente registrados e armazenados nos mais diferentes repositórios das redes corporativas contendo arquivos de textos, planilhas, imagens e vídeos.

Os registros realizados, desde a notificação e ordens de manutenção para intervenção até os tratamentos de falhas, correspondem aos dados armazenados em sistemas informatizados cujo o volume cresce rapidamente e que contém informações valiosas. Porém, a recuperação dessas informações não é de forma direta, muita informação e conhecimento úteis podem estar sendo desperdiçados, ficando ocultos dentro das bases de dados com a falta de ferramentas adequadas para sua extração, associada a limitação humana de analisar extensas bases de dados e extrair relações entre elas [Silva 2007].

Há uma demanda crescente, voltada a decisões estratégicas, por novos conhecimentos obtidos a partir da exploração de dados armazenados referentes às ocorrências ou eventos de alto impacto no processo de produção. Nesse contexto, a descoberta de conhecimento em bases de dados de ocorrências, anomalias, defeitos ou falhas de equipamentos, utilizando técnicas de análise de dados para avaliar os eventos relevantes, estabelecendo um modelo de como são endereçados, pode ser uma ferramenta importante para o processo de tomada de decisão realizados pelos gestores da manutenção de ativos. Para a empresa CEL, essa demanda também está na ordem do dia.

Do inglês, Knowledge-Discovery in Databases (KDD) [Fayyad et al. 1996] é uma área da Inteligência Artificial e pode ser visto ainda como o processo da descoberta de novas correlações, padrões e tendências significativas por meio da análise minuciosa de grandes conjuntos de dados [Silva 2007].

A partir desses pressupostos, portanto, pode-se apresentar a questão de pesquisa: como aplicar a metodologia de descoberta de conhecimentos em base de dados relacionados às ocorrências de manutenção, anomalias ou falhas nos equipamentos da empresa CEL nos últimos 4 anos para obtenção de insights que contribuam com a melhoria da gestão da função manutenção?

Isso significa que o principal objetivo deste estudo é avaliar a aplicação da metodologia de descoberta do conhecimento em base de dados no escopo ou universo de dados e equipamentos estabelecido nesta pesquisa efetuada na empresa de celulose CEL, entre os anos de 2015 e 2018, período em que as perdas de produção atribuídos a problemas de manutenção de equipamentos totalizaram 103.310 toneladas de celulose, e, com isso, identificar oportunidades que possam contribuir para melhorar a eficiência da gestão da manutenção.

Para tanto, foram elencados os seguintes objetivos específicos:

- i. Definir o processo mais adequado para extração do conjunto de dados alvos, seu pré-processamento e seleção por relevância;
- ii. Identificar o melhor método de análise de dados a ser aplicado para a descoberta de conhecimentos dentro do contexto investigado;
- iii. Avaliar como o método de descoberta de conhecimento pode contribuir com os direcionamentos estratégicos da gestão da manutenção.

A importância desse estudo é demonstrar que a metodologia da descoberta de conhecimentos aplicada aos dados de manutenção industrial pode contribuir de forma significativa para o estabelecimento de direcionamentos estratégicos, uma vez que permite análises mais minuciosas daqueles eventos de alta criticidade e impacto.

Para este artigo, na seção 2 foi elaborada a fundamentação teórica da função manutenção industrial e alguns conceitos normativos relevantes para esta pesquisa e da metodologia de descoberta de conhecimento em base de dados. A seção 3 descreve os métodos e procedimentos utilizados, os procedimentos de coleta e análise de dados, bem como os softwares aplicados. Na seção 4 são apresentados os resultados do estudo de caso. Na seção 5 são realizadas as análises e discussões dos resultados. Por fim, na seção 6, são apresentadas as considerações finais e futuros desdobramentos.

2. Referencial Teórico

Nesta seção são apresentados os principais conceitos utilizados para embasar o estudo realizado.

2.1 Manutenção Industrial

De acordo com Cabral (2006), a manutenção industrial pode ser definida como o conjunto de ações destinadas a assegurar um adequado funcionamento de equipamentos, máquinas e instalações garantindo que elas sejam submetidas a ações ou intervenções nas oportunidades e com a abrangência adequada visando evitar avarias ou que baixem de rendimento.

Visando a padronização internacional, a norma DIN EN 13306 (2018) apresenta a manutenção como sendo “a combinação de todas as ações técnicas, administrativas e gerenciais durante o ciclo de vida de um item, destinadas a manter ou recolocar o item em um estado no qual possa desempenhar a função requerida”.

Outros termos relacionados à manutenção, falha e disponibilidade empregados neste estudo são definidos pela mesma norma:

- Gerenciamento de Manutenção: todas as atividades gerenciais que determinam os requisitos de manutenção, objetivos, estratégias e responsabilidades e sua implementação por meios como planejamento de manutenção, controle de manutenção e melhoria das atividades de manutenção e economia;
- Estratégia de Manutenção: método de manutenção utilizado para atingir os objetivos de manutenção;
- Manutenção Preventiva: manutenção realizada destinada a avaliar e ou mitigar a degradação e reduzir a probabilidade de falha de um item;
- Manutenção Preditiva: manutenção baseada na condição realizada seguindo uma previsão derivada de análise repetitiva ou conhecimento de características e avaliação de parâmetros significativo da degradação do item;
- Manutenção Corretiva: manutenção realizada após o reconhecimento da falha e destinada ao restabelecimento do item para um estado no qual ele pode desempenhar uma função requerida;
- Item: parte, componente, dispositivo, subsistema, unidade funcional, equipamento ou sistema que pode ser descrito e considerado individualmente;
- Falha: perda da capacidade de um item em desempenhar uma função requerida;
- Função Requerida: função, combinação de funções, ou uma totalidade de funções de um item considerada necessária para cumprir um dado requisito;
- Modo de Falha: forma na qual ocorre ou se manifesta a incapacidade de um item para executar uma função requerida. Um modo de falha pode ser definido pela função perdida ou pela transição de estado ocorrido e pode ser um único evento ou estado que causa uma falha funcional;
- Disponibilidade: capacidade de um item ou ativo estar em condições de operar como e quando necessário, sob condições especificadas, assumindo que os recursos externos requeridos sejam fornecidos.

A norma ISO 13372 (2012) também contém a definição de outros termos importantes utilizados neste trabalho:

- Anomalia ou Anormalidade: desvio de uma condição padrão;
- Defeito: condição de um item que ocorre quando um de seus componentes ou peças degrada ou exibe um comportamento anormal, o que pode evoluir para a falha do item.
- Manutenção Proativa: Tipo de manutenção que enfatiza a detecção e correção rotineiras das condições de causa raiz que, de outra forma, levariam à falha.

Como desafio estratégico, o enfoque atual da manutenção é atuar na prevenção da falha ou da indisponibilidade dos equipamentos e outros ativos industriais evitando assim impactos indesejados na produção, meio ambiente, segurança no trabalho, etc. [Fuentes 2006, Pinto e Ribeiro 2002].

No entanto, enquanto função estratégica, a manutenção não consiste apenas em evitar a falha ou reparar, mas também contribuir com a melhoria em termos de custos, segurança, meio ambiente e produtividade. Isso significa que a manutenção tem muita relevância para aumentar a competitividade das organizações. Nesse sentido, a eficácia dos custos aplicados e a

efetividade ou acurácia das atividades são critérios fundamentais para uma boa manutenção. [Matsumota 2016, Kardec et al. 2001, Pinto 2002].

Em um mundo em transformação com o advento da era digital em uma economia globalizada e altamente competitiva, a manutenção enquanto função estratégica também está sob os desafios das grandes mudanças, da complexidade cada vez maior das instalações produtiva, da alta tecnologia embarcada e da necessidade de profissionais que estejam à altura dessa realidade. A busca por novas ferramentas que auxiliem no gerenciamento de uma manutenção cada vez mais proativa é de fundamental importância para a competitividade das empresas, pois impacta na qualidade e produtividade de seus produtos, processos e serviços [Kardec e Nascif, 2001].

Por isso, há demandas cada vez maiores por conhecimentos, técnicas e práticas inovadoras que contribuam para a melhoria na gestão da manutenção, que é o caso da aplicação da metodologia da descoberta de conhecimentos sob base de dados e respectivas ferramentas para análise dos dados.

2.1 Análise de Dados

De acordo com Amaral (2016, p. 61), “analisar dados é aplicar algum tipo de transformação nos dados em busca de conhecimento” e podem ser classificados em dois tipos, análise explícitas e implícitas:

1. **Análises explícitas.** A informação e o conhecimento estão disponíveis explicitamente nos dados e, normalmente, só é necessária alguma operação, de baixa complexidade, para ressaltar o dado e produzir a informação tais como: aplicação de um filtro, ordenação de registros, criação de colunas calculadas, etc. “A informação está lá, explicitamente, é preciso apenas destaca-la na “multidão””.
2. **Análises implícitas.** A informação não está disponível claramente no conjunto de dados, mesmo tentando olhar os dados de diversas formas, filtrando ou calculando a informação só será produzida com o uso de alguma função mais sofisticada como aprendizado de máquina ou um modelo matemático ou estatístico.

Ainda conforme Amaral (2016, p. 63), “antes de analisar dados, é preciso conhecê-los”. Para isso, “a técnica mais consagrada é a **análise exploratória**, proposta pelo estatístico John Wilder Tukey na obra *Exploratory Data Analysis* de 1977, cujos conceitos são válidos até hoje. O objetivo é conhecer os dados antes de tentar analisá-los, para depois, usando técnicas explícitas ou implícitas, por exemplo, tirar conclusões”. [Amaral 2016, p.64]

A análise exploratória pode usar tanto técnicas quantitativas (medidas de dispersão e posição, média, mediana, amplitude, desvio padrão, etc.) como visuais (diagrama de dispersão, diagrama de caixa, diagrama de colunas, histogramas, nuvem de barra, etc.). Importante observar que há uma diferença entre análises exploratórias e as análises explícitas: “os conceitos se sobrepõem de diversas formas, e a diferença está mais nos objetivos que nas técnicas: enquanto a primeira busca conhecer os dados, a segunda tem um objetivo claro e específico, por exemplo, resumir as vendas do mês, verificar notas faltantes, checar o cálculo do imposto, etc.” [Amaral 2016, p.64]

2.2 Mineração de Dados e Descoberta de Conhecimento sob Base de Dados

Segundo Pinheiro (2008, p.97), mineração de dados, ou *Data Mining*, é o processo de descoberta de padrões e tendências existentes em repositórios de dados. Este processo visa basicamente a análise de dados com o objetivo principal de descoberta de conhecimento. Os maiores benefícios pela execução de processos de mineração de dados é a criação de uma

inteligência de negócios sobre determinado assunto. O processo de mineração de dados não é apenas um processo computacional ou automático, mas sim um esforço cooperativo entre pessoas e máquinas. Também referenciado como KDD – *Knowledge Discovery in Databases* – ou descoberta de conhecimento sob base de dados, os processos de *data mining* focam na aplicação de técnicas estatísticas e de inteligência artificial para a análise interativa dos dados, visando a identificação de padrões de comportamento, tendências ou predição. “Em termos de nomenclatura, existem algumas diferenças entre os conceitos de *KDD* e *data mining*, a depender do autor. Para alguns, *data mining* é apenas a aplicação de técnicas de inteligência artificial com o objetivo de descobrir relações entre os dados analisados, sendo um componente de descoberta de conhecimento dentro do processo de KDD”. [Pinheiro 2008, p.97]

Ainda de acordo com Pinheiro (2008, p. 98) o processo de descoberta de conhecimento em base de dados pode ser descrito como sendo composto por seis etapas distintas a saber:

1. A primeira etapa corresponde ao entendimento do problema, quando se busca a compreensão dos objetivos do projeto e suas necessidades;
2. A segunda etapa está relacionada com a extração dos dados visando criar um conjunto de dados alvos onde a descoberta deverá ser efetuada;
3. A terceira etapa, de limpeza dos dados, é realizado o pré-processamento dos dados, ou seja, tratamento dos dados ausentes, inconsistentes ou fora dos padrões;
4. A quarta etapa diz respeito à seleção daqueles dados que podem influenciar nos resultados do modelo a ser construído;
5. A quinta etapa é voltada para a escolha do método de mineração de dados, que envolve as fases de levantamento dos objetivos do processo, a identificação da melhor técnica a ser aplicada e a abordagem para a aplicação dos modelos.
6. A sexta e última etapa é a interpretação dos resultados: apresentação das descobertas obtidas, determinação da melhor forma de utilizar as informações na tomada de decisão, definição das vantagens e desvantagens dos modelos, reavaliação do processo como um todo.

Simplificadamente, o processo de descoberta de conhecimento em base de dados pode ser representado pela Figura 1 com base da obra de [Fayyad et al. 1996].

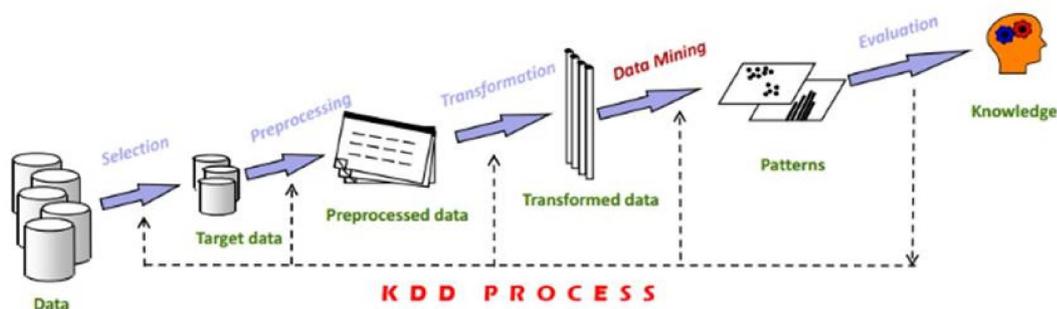


Figura 1. Processo de Descoberta de Conhecimento em Base de Dados

2.3 Aprendizado de Máquina

Normalmente a mineração de dados está associada ao aprendizado de máquina: uma área da inteligência artificial que desenvolve algoritmos capazes de fazer com que o computador aprenda a partir do passado “usando dados de eventos que já ocorreram”. [Amaral 2016, p.61]. O aprendizado de máquina é capaz de identificar padrões que dificilmente seriam

identificados a “olho nu” por técnicas triviais de análise de dados ou exploratórias, por pivôs, filtros, junções ou agrupamentos.

Segundo o mesmo autor, o que em banco de dados é uma tabela, em aprendizado de máquina é chamado de relação, que contém as características do negócio, onde cada coluna é um atributo e cada linha é uma instância. Em relação às tarefas, o aprendizado de máquina pode ser dividido em quatro grandes grupos: classificação, regressão, agrupamentos e regras de associação:

- **Classificação:** objetiva prever um atributo especial chamado “classe” a partir dos demais atributos. A classe é um tipo de dado nominal ou categórico. Usa-se, por exemplo, para prever uma fraude ou um tipo de doença.
- **Regressão:** é um tipo de classificação onde a classe é numérica. Prever o peso ou altura de uma pessoa é um exemplo de regressão.
- **Agrupamento:** não existe classe e o objetivo é criar grupos e atribuir instâncias a estes grupos a partir das características ou atributos destas instâncias. Exemplo: identificar grupo de clientes e direcionar campanhas de marketing.
- **Regras de Associação:** buscam a relação entre itens ou identificam causas ou conjuntos de fatores que levam a ocorrência de um fenômeno ou fato. Diagnósticos médicos e sensores, são exemplos de aplicação.

As tarefas de mineração podem ser supervisionadas, ou seja, quando há uma classe com o qual se pode comparar e validar o resultado ou não supervisionadas, onde não existe uma classe (Amaral, 2006). Por isso, classificação é um exemplo de tarefa supervisionada e agrupamentos são exemplos de tarefas não supervisionadas.

Em função das características dos dados e objetivos desta pesquisa, será utilizado somente a classificação como tarefa de aprendizado de máquina.

2.3.1 Tipos de Algoritmos de Classificação

Os dois tipos de algoritmos de classificação utilizados nesta pesquisa são o de Naive Bayes e de Árvores de Decisão:

- **Naive Bayes:** segundo Amaral (2016), o classificador é baseado na teoria das probabilidades e que supõe que os atributos vão influenciar a classe de forma independente. Durante a criação do modelo, será construída uma tabela mostrando a contribuição ou o peso que cada categoria de cada atributo contribui com cada classe. A partir do modelo construído, ao ser submetido uma nova instância para o algoritmo classificador, este irá considerar os pesos nesta tabela, somá-los e verificar qual classe atingiu um peso maior. Na Figura 2, extraído do Weka, é possível observar os pesos atribuídos a cada classe de um determinado atributo.

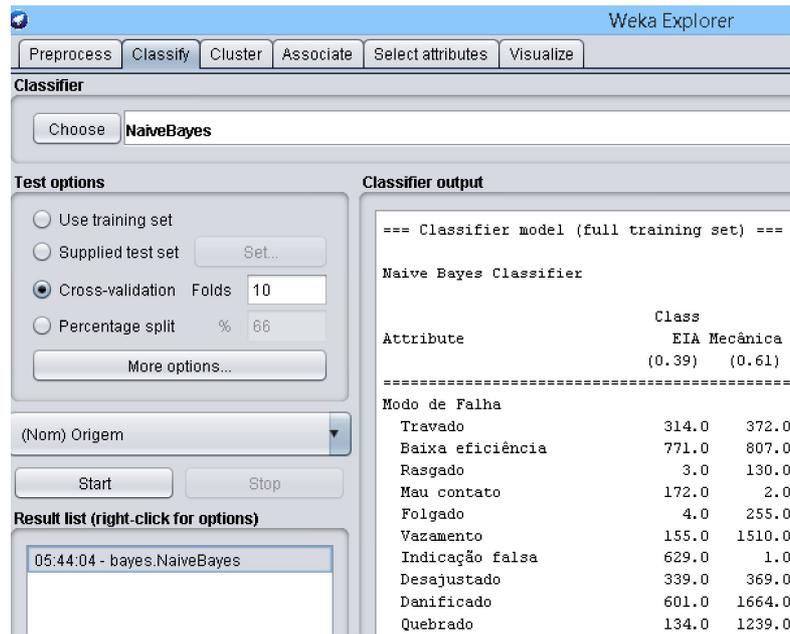


Figura 2. Classificador Naive Bayes - Weka

- Árvore de Decisão:** consiste em um modelo de predição que pode ser visualizado como uma estrutura de árvores. “Especificamente, cada ramo da árvore representa um tópico de classificação, e as folhas da árvore são partições de conjuntos de dados com suas respectivas classificações” [Pinheiro 2008, p. 127). A árvore sempre começa por um único nodo raiz e vai sendo dividida até levar a classe, conforme demonstrado na Figura 3. Um algoritmo determina como cada nodo será particionado. O tipo de partição utilizado depende do algoritmo e do tipo do dado. Dados categóricos podem gerar partições binárias ou mesmo várias partições. Dados numéricos normalmente têm condições lógicas do tipo “maior” ou “menor igual” para decidir o caminho a percorrer [Amaral 2016].

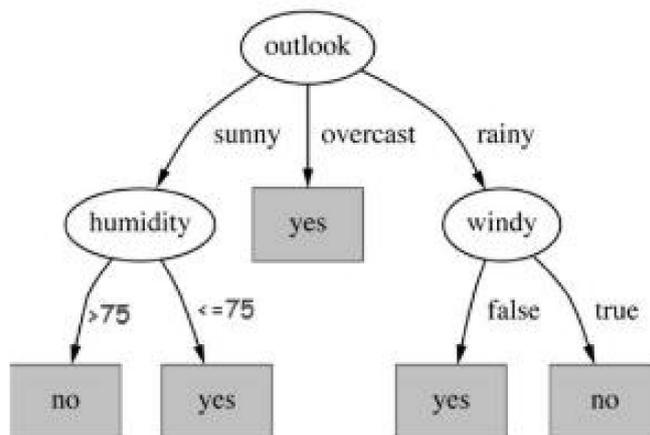


Figura 3. Classificador Árvore de Decisão - Weka

3. Métodos e Procedimentos

São apresentados nesta seção os métodos e procedimentos aplicados com base na fundamentação teórica visando responder ao problema de pesquisa.

3.1 Delineamento da Pesquisa

A abordagem desse estudo é quantitativa, que é caracterizado pelo emprego da quantificação tanto da coleta de informações quanto no tratamento delas por meio de técnicas estatísticas, que podem ser simples e complexas (RICHARDSON, 1999). Esta abordagem foi realizada a partir dos registros disponíveis na base de dados das ocorrências de anomalia e falha em equipamentos que foram registradas ao longo do tempo pelas equipes de manutenção e operação da empresa CEL, a ser apresentada a seguir.

Do ponto de vista da natureza trata-se de uma pesquisa aplicada, que segundo Gerhardt e Silveira (2009, p.35), “objetiva gerar conhecimentos para aplicação de práticas, dirigidos à solução de problemas específicos, envolvendo verdade e interesses locais”. No contexto deste trabalho, a aplicação da metodologia da descoberta de conhecimento através dos dados da manutenção da empresa CEL para identificação de oportunidades para melhoria da gestão da manutenção de equipamentos críticos ou que impactam na produção.

Em relação aos objetivos da pesquisa, esse estudo classifica-se como exploratório. “O principal objetivo da pesquisa exploratória é prover a compreensão do problema enfrentado pelo pesquisador”. Dessa forma, “é usada em casos nos quais é necessário definir o problema com maior precisão, identificar cursos relevantes de ação ou obter dados adicionais antes que se possa desenvolver uma abordagem”. [Malhotra 2001, p.105]

Gil (2008) afirma que as pesquisas exploratórias visam proporcionar visão geral, tipo aproximativo, realizado quando o tema escolhido é pouco explorado, para proporcionar maior familiaridade com o problema e para torná-lo mais explícito. Sendo assim, pode ser classificada como um tipo de pesquisa que permite a obtenção de novos conhecimentos, bem como ampliação e complementação acerca do tema abordado.

Quanto ao método de pesquisa, relacionada aos procedimentos, o escolhido foi o estudo de caso. Um estudo de caso “é uma investigação empírica que investiga um fenômeno contemporâneo dentro de seu contexto da vida real, especialmente quando os limites entre o fenômeno e o contexto não estão claramente definidos” [Yin 2010, p. 39]. Assim, o estudo de caso compreende uma estratégia de pesquisa que abrange tudo – com a lógica de planejamento incorporando abordagens específicas à coleta de dados e à análise de dados (YIN, 2010).

3.2 Unidade de Análise

O estudo foi desenvolvido junto a uma empresa industrial de produção de celulose, que por confidencialidade será chamada como CEL. Essa empresa, localizada na Região Nordeste do Brasil, produz celulose há quase 15 anos com uma média anual de produção em torno de 1,1 milhão de toneladas.

Celulose ou pasta química é a matéria-prima utilizada na fabricação de diversos tipos de papéis. Pode ser extraída de diversas fontes (matérias-primas) fibrosas, principalmente da madeira. Os produtos de celulose e papel da empresa CEL são fabricados, exclusivamente, a partir de madeira de florestas plantadas de eucalipto.

As etapas do processo produtivo da celulose da empresa CEL são basicamente as seguintes:

- Preparação da Matéria-Prima (madeira);
- Cozimento;
- Depuração e Lavagem alcalina;
- Branqueamento;
- Secagem e Embalagem.

Há um outro processo conhecido como recuperação química, que permite o reaproveitamento dos produtos químicos utilizados e da fração de madeira solubilizada no processo de obtenção da celulose (que será convertida em energia térmica e elétrica), viabilizando a fabricação de celulose em termos de custos e também em termos ambientais. As etapas do ciclo de recuperação química são as seguintes:

- Tratamento de Água da Caldeira;
- Evaporação;
- Caldeira Recuperação Química;
- Caldeira Biomassa;
- Central Elétrica;
- Caustificação;
- Sistema de Controle de Odor.

A Figura 4 (Castro 2009, p.10) representa de forma simplificada as partes que constituem um processo de celulose.

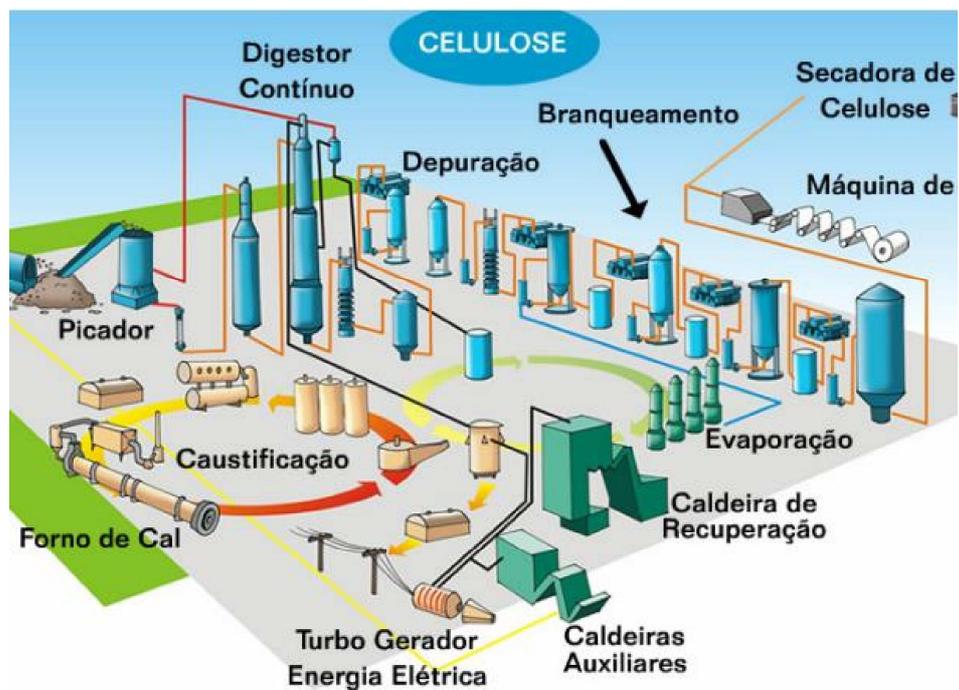


Figura 4. Processo Simplificado de Produção de Celulose

A planta de celulose da empresa CEL é composta pelas seguintes Áreas ou Macroprocessos e seus respectivos Processos:

- Linha de Fibras: Preparação de Cavacos, Cozimento, Branqueamento, Lavagem / Deslignificação e Depuração;
- Recuperação e Utilidades: Caldeira de Recuperação, Evaporação, Caldeira de Força, Desmineralização, Turbo Gerador;
- Secagem: Máquina, Cortadeira, Secador, Enfardamento, Depuração e Estocagem;
- Forno e Caustificação: Forno de Cal e Caustificação;
- ETE e ETE: Tratamento de Efluente e Tratamento de Água.

A função manutenção da planta de celulose é realizada por uma equipe de aproximadamente 180 profissionais, entre gestores, engenheiros e técnicos especializados das áreas de mecânica, elétrica, automação, instrumentação e planejamento. Esses profissionais ficam distribuídos nas diversas áreas da fábrica atuando juntamente com os profissionais da produção ou operadores.

No período sob estudo, 2015 a 2018, perdas de produção atribuídos a problemas de manutenção de equipamentos totalizaram 103.310 toneladas de celulose, uma média anual de 25.828 toneladas, ou aproximadamente 2,3 % da produção total anual. O preço da tonelada de celulose de fibra curta, que é o tipo de celulose produzida na empresa CEL, é de aproximadamente USD 970, segundo os dados do Centro de Estudos Avançados em Economia Aplicada e Economia Florestal (CEPEA-ESALQ/USP, 2017). Dessa forma, pode-se afirmar que o valor cessante equivalente à perda de todo o período sob investigação foi de USD 100 milhões, ou seja, uma média anual de aproximadamente USD 25 milhões. São valores que por si só justificam os esforços em busca de mais informações e, sobretudo, conhecimentos sobre a natureza dos problemas geradores dessas perdas.

3.3 Etapas da Pesquisa

As etapas realizadas nesta pesquisa foram as seguintes:

- Levantamento bibliográfico e fundamentação teórica do método de descoberta de conhecimentos em base de dados, que foi aplicado ao universo de eventos de anomalias em equipamentos em uma planta de celulose;
- Coleta dos dados no sistema ERP (SAP-PM) referente ao período sob investigação a partir dos registros de todas as ocorrências de defeitos e falhas em equipamentos;
- Pré-processamento dos dados, com a limpeza dos registros errôneos, tratamento aos dados ausentes e inconsistentes;
- Seleção e aplicação de ferramenta para a realização de análises exploratórias e explícitas, visando aquisição de maiores conhecimentos sobre os dados, bem como a definição de alguns direcionadores para a melhor escolha daqueles dados que podem influenciar nos resultados do modelo a ser construído.
- Escolha dos métodos de mineração de dados para análises implícitas e ferramenta utilizada para a realização do treinamento dos modelos;
- Interpretação, apresentação e avaliação dos resultados dos testes, das descobertas obtidas e sua relevância em relação aos objetivos propostos nessa investigação, ou seja, identificação de oportunidades e recomendações que possam contribuir para melhoria da eficácia e eficiência da gestão da manutenção.

3.4 Coleta e Preparação de Dados

A coleta de dados foi realizada a partir dos registros feitos pelas equipes de produção obtidos ao longo de 4 anos (2015 a 2018) no módulo de Planejamento e Manutenção (PM) do SAP, que é o fornecedor do sistema ERP, *Enterprise Resource Planning*, adotado pela empresa CEL em todas suas operações.

Os registros selecionados para análise foram todos aqueles considerados ocorrências de anomalia em equipamentos da fábrica, ou seja, eventos que tenham gerado manutenções corretivas por motivos de defeito ou falha. Através de transações configuradas pela equipe de planejamento no módulo PM do SAP, esses registros foram exportados para um arquivo excel, que posteriormente foi convertido para um arquivo CSV, permitindo assim análise tanto através do MS Power BI, como do sistema WEKA para mineração de dados.

Considerando o período estudado, 2015 a 2018, o conjunto de dados originalmente gerado em forma de planilha excel contemplava 28 colunas e 30.312 linhas, o que corresponderia a 848.736 campos para registros. Imediatamente, identificou-se 4 colunas que eram redundâncias ou não eram utilizadas e, portanto, foram desconsideradas. Portanto, restaram 24 colunas com as mesmas 30.312 linhas, ou seja, 727.488 campos para registros. Desses, foram identificados 2.485 registros faltantes, o que corresponde a 0,34 % do total de campos disponíveis para registros. Dos registros faltantes, os atributos “Origem” e “Processo” eram os que tinham a maioria dos dados faltantes, respectivamente, 1232 e 921 registros.

3.5 Pré-Processamento dos Dados

Utilizando-se o recurso do sistema de filtros ainda no formato inicial de arquivo, ou seja, no MS Excel (extensão xlsx), ao observar os registros nos atributos “Tipo de Ocorrência” e “Falha de Equipamento”, que são dados alvos do estudo, foram identificados 382 registros realizados erroneamente para os atributos “Natureza” e “Origem”, que deveriam estar registrados como “Manutenção”, mas estavam como “Operação”. Tais registros foram imediatamente corrigidos.

Para a determinação de dados faltantes foi utilizado aprendizagem de máquina através de algoritmos classificadores de uma ferramenta ou sistema chamado WEKA. Este sistema está descrito no item 3.7.1.

3.6 Técnicas para Análise Exploratória e Explícita

Para as análises do tipo exploratória e explícita foi utilizada amplamente a técnica visual através dos recursos do MS Power BI, versão desktop. Segundo o site da Microsoft, “o Power BI é um serviço de análise de negócios que fornece insights para permitir decisões rápidas e informadas”.

Com base na questão desta pesquisa, bem como os seus objetivos fez-se necessário formular algumas questões iniciais direcionadoras para a análise exploratória dos dados relacionados às ocorrências de manutenção, anomalias, falhas nos equipamentos e perdas de produção no período investigado, 2015 a 2018, a saber:

- No período sob estudo, identificar as áreas ou macroprocessos e processos da fábrica que apresentaram o maior número de ocorrências e perdas de produção decorrentes de defeitos e ou falhas em equipamentos;
- Destacar e mensurar os equipamentos com mais ocorrências de anomalia e aqueles, que por motivos de falhas, provocaram as maiores perdas de produção;

- Identificar e destacar as áreas, processos e, sobretudo equipamentos, que ao longo de cada ano do período investigado mantiveram-se entre as maiores ocorrências de anomalias e falhas;
- Mensurar a participação de cada disciplina que compõe a manutenção, ou seja, mecânica e a disciplina agrupada de elétrica, instrumentação e automação (EIA), em relação às ocorrências de anomalia e perdas de produção de cada equipamento.

3.7 Técnicas para Análise Implícita

Quanto a análise do tipo implícita, foi aplicada aprendizagem de máquina para tarefa de classificação de dados faltantes de dois atributos importante do banco de dados: “Origem” e “Processo”. O primeiro atributo se refere à disciplina de origem da anomalia motivada pela natureza manutenção formado por duas classes de disciplinas: disciplina agrupada EIA (Elétrica, Instrumentação e Automação) ou disciplina de Mecânica. O segundo atributo se refere à subárea ou processo formado por 18 classes: Evaporação, Cozimento, Caldeira de Recuperação, Secador, Caustificação, Tratamento de Efluentes, Preparação de Cavacos, Branqueamento, Forno de Cal, Tratamento de Água, Enfardamento, Desmineralização, Depuração, Cortadeira, Caldeira de Força, Depuração, Lavagem e Deslignificação e Turbo Gerador.

3.7.1 Ferramenta Utilizada para Aprendizagem de Máquina

Utilizou-se uma ferramenta denominada Weka, versão 3.9.3. Trata-se de um conhecido software *open source* de aprendizado de máquina desenvolvido em linguagem Java e mantido pela Universidade de Waikato na Nova Zelândia. O Weka pode ser operado tanto em linha de comando como através de interface gráfica, que é o que lhe confere muita popularidade. Ou seja, ao invés de digitar códigos, o usuário simplesmente seta valores e comanda funções. O Weka é basicamente um ambiente de aprendizado de máquina, embora também tenha várias funcionalidades de visualização de dados.

Conforme Figura 5, o Weka traz alguns menus e quatro aplicativos: *Explorer*, *Experimenter*, *Knowledgeflow*, *Workbench* e *Simple CLI*. Utilizou-se neste trabalho somente o principal aplicativo que é o Explorer.



Figura 5. Interface Gráfica do Software Weka

No *Explorer*, além de visualizar e explorar o conjunto de dados pode-se aplicar filtros e executar tarefas de aprendizado de máquina como classificação, agrupamento, regras de associação e seleção de atributos. A Figura 6 mostra a interface deste aplicativo.

Dentre os inúmeros recursos do Weka, importante destacar o sistema de filtros, principalmente aqueles que permitem realizar uma seleção de atributos, destacando os mais importantes e removendo os menos importantes. Isso não significa que tais recursos de seleção de atributos necessariamente vão tornar o modelo mais eficiente. De qualquer forma, “mesmo em casos em que a performance seja semelhante àquela obtida usando todos os atributos, um subconjunto de atributos reduz o custo computacional do modelo, torna a sua apresentação mais simples, além de outras vantagens, como simplificar os processos de extração, transformação e carga de dados” [Amaral 2016, p. 140]. Tal recurso, o de realizar a seleção de atributos, foi utilizado nesta pesquisa e o seu acesso na interface gráfica de pré-processamento do Weka também é mostrado na Figura 6.

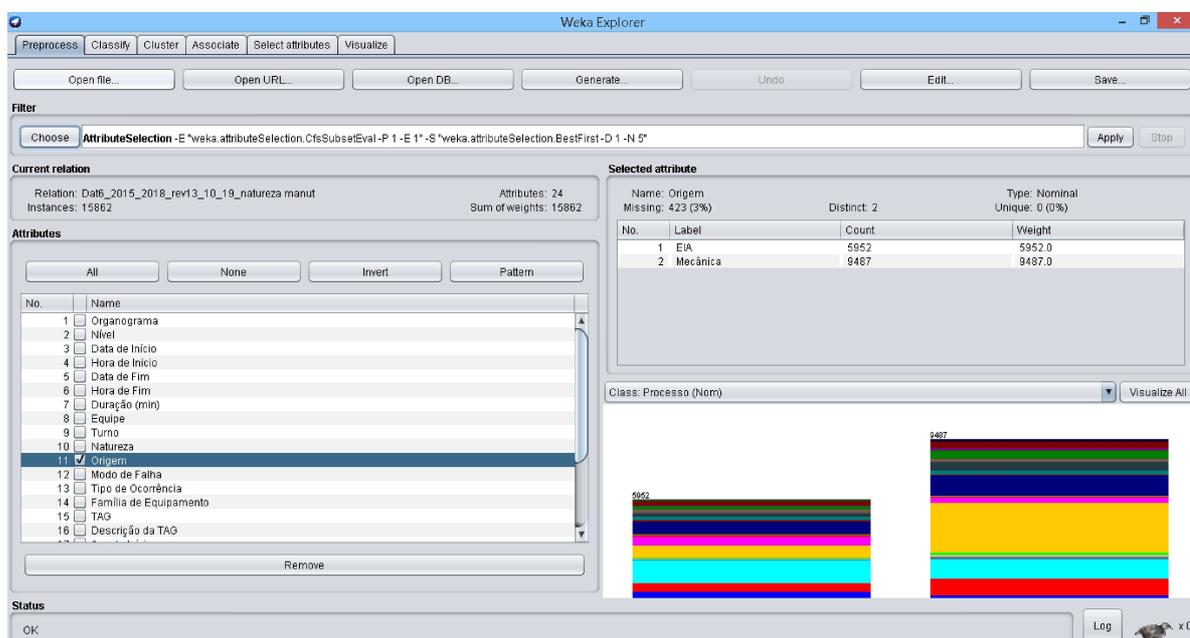


Figura 6. Interface Gráfica do Software Weka

A interface de pré-processamento, *Preprocess*, possui diversas funcionalidades para carregar dados através do menu *Open file* e também aplicar filtros, menu *filter*, bem como explorar dados. Outro comando importante é o *Edit*, onde podem ser abertos os dados previamente carregados em uma grade, permitindo tanto a visualização dos dados no seu formato natural como alteração ou correção de forma ágil e simples qualquer informação.

3.7.2 Algoritmos Classificadores Utilizados

De acordo com o que já foi apresentado na seção 2.5, os tipos de algoritmos classificadores utilizados nesta pesquisa foram o Naive Bayes e o de Árvore de Decisão. Este último, dentre as 7 alternativas existentes nesta versão do Weka, optou-se pelo algoritmo denominado J48 pois apresenta um escopo de recursos mais abrangente em termos de tipos de atributos, conforme é mostrado na Figura 7, além da obtenção dos melhores resultados quando comparado com os demais.

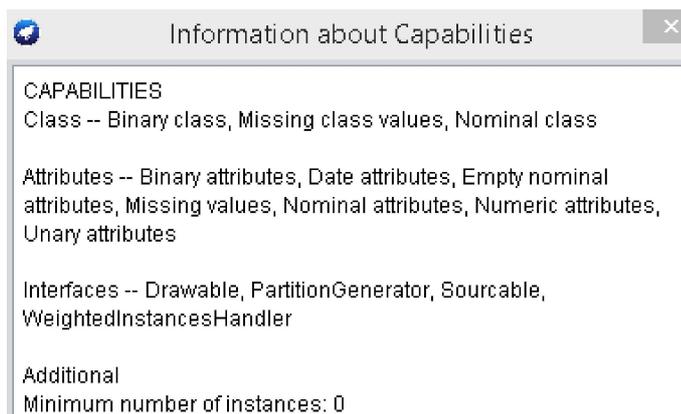


Figura 7. Recursos do Algoritmo J48

Ambos os algoritmos classificadores utilizados, podem ser acessados no menu do *Explorer*, opção *Classify*, conforme indica Figura 8. As setas em vermelho foram inseridas pelo autor deste trabalho para testar os algoritmos escolhidos.

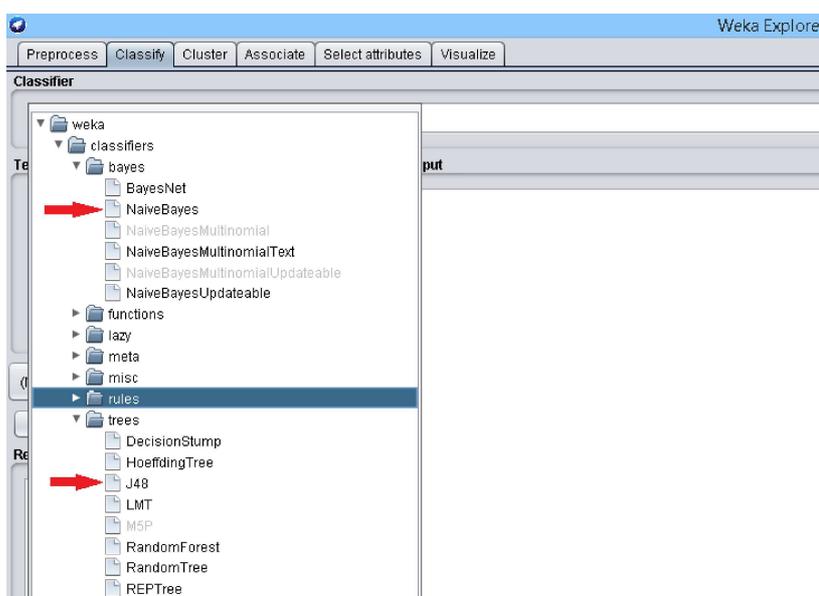


Figura 8. Algoritmos Classificadores do Weka

3.7.3 Avaliação de Desempenho dos Modelos

No aprendizado de máquina, um algoritmo aprende com dados históricos e este processo de aprendizado efetuado por um algoritmo tem como resultado a construção de um modelo. Um modelo “nada mais é do que uma fórmula criada, a partir de dados históricos, para prever, ou classificar, novos dados que sejam apresentados para o algoritmo de classificação e que ele ainda não tenha visto”. [Amaral 2016, p. 32]

Antes da utilização de um modelo em produção, no entanto, é necessário avaliar o seu desempenho, ou seja, é necessário medir o quanto o modelo construído é eficiente em prever dados que ainda não conheceu. Para isso, foram utilizados os seguintes métodos, cujas opções no WEKA podem ser visualizadas na Figura 9: Validação Cruzada (*Cross-Validation*), *Hold Out* ou *Percentage Split* e Matriz de Confusão (*Confusion Matrix*). Tanto o método da validação cruzada como o de *Hold Out* utilizam parte dos dados para criar o modelo e parte para testá-lo,

o que os tornam mais confiáveis por testar o modelo em dados que o modelo nunca tivesse visto antes [Amaral 2016]. Por isso, não se optou no Weka pelo método *Use Training Set* porque utiliza os mesmos dados para criar e testar o modelo. Quanto a matriz de confusão ela permite realizar uma avaliação mais ampla do que simplesmente o percentual de acertos conforme descrito no último parágrafo desta seção.

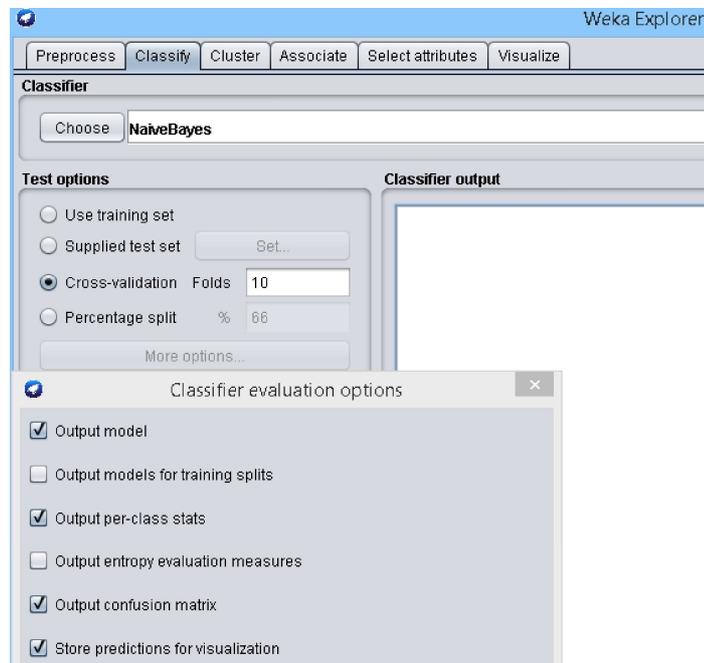


Figura 9. Opções de Testes do WEKA para Avaliação de Desempenho

- Método Validação Cruzada:** permite que o modelo seja avaliado uma série de vezes, cada série é conhecida como partição e ao final o desempenho é medido a partir da média aritmética das avaliações. Nesta pesquisa, para todos os experimentos, selecionou-se no Weka a caixa de opções de teste, *Test Options*, do classificador para validação cruzada, *Cross-Validation*, um número de 10 avaliações ou *Folds*. Isso significa que os dados históricos foram divididos em 10 partes semelhantes, de forma aleatória e sem substituição. Neste caso, o processo realiza 10 interações e a cada uma delas 9 partes são utilizadas para treino e uma parte para teste e no final, as 10 partes são utilizadas para testar o modelo conforme representa a Figura 10 [Amaral 2006, p. 35].

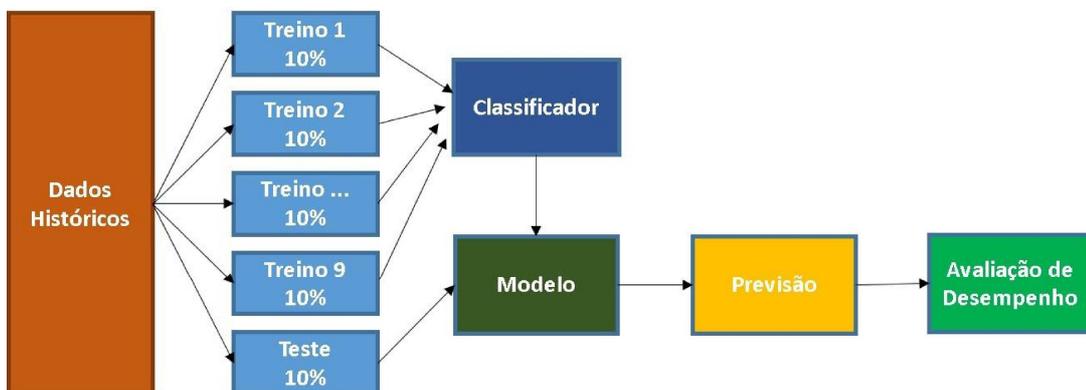


Figura 10. Método Validação Cruzada

- Método Hold Out:** este método no ambiente de aprendizado, conforme mostra Figura 11 (Amaral 2016, p. 33), consiste em separar parte dos dados para treino ou criação do modelo e parte para testá-los. Os dados de treino são submetidos ao classificador e produzem o modelo. Em seguida, os dados de teste são submetidos ao modelo para que seja feita a previsão. Da comparação entre a classe real e a classe prevista de dados históricos conhecidos, é possível realizar uma avaliação de desempenho. Nesta pesquisa, utilizou-se nos experimentos para opção de teste em Hold Out do classificador, que no Weka se apresenta como a opção *Percentage Split*, a proporção de 70% dos dados para treinar o modelo e 30% para testá-lo.

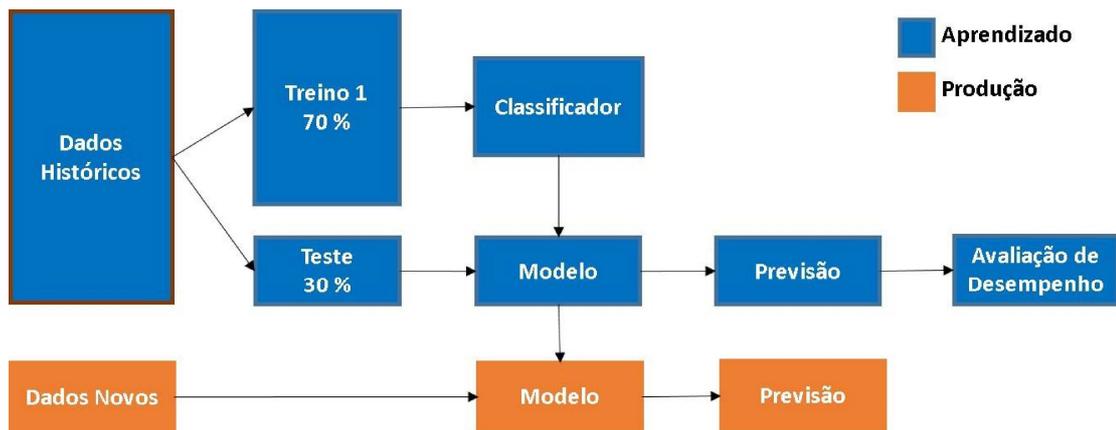


Figura 11. Método Hold Out

- Método Matriz de Confusão:** permite avaliar com quatro possibilidades na avaliação de desempenho do classificador: os verdadeiros positivos, falsos negativos, verdadeiros negativos e falsos positivos. Em todos os experimentos desta pesquisa este método foi utilizado. No WEKA, seleciona-se esta opção conforme mostra a Figura 7 em *Output confusion matrix*, que é acessado em *More Options* na caixa *Test options*.

4. Resultados do Estudo de Caso

O estudo de caso foi realizado visando identificar oportunidades de obtenção de informações relevantes e conhecimentos que possam contribuir com a gestão da função manutenção da planta de celulose da empresa CEL através da metodologia de descoberta de conhecimento em base de dados.

4.1 Análise Exploratória e Explícita dos Dados

A partir das questões direcionadoras realizou-se a análise exploratória e explícita dos dados disponíveis, por meio da ferramenta ou software da Microsoft (MS) Power BI.

4.1.1 Macroprocessos e Processos da Fábrica com Maiores Ocorrências de Manutenção

Em relação ao número de ocorrências de anomalia por motivos ou de natureza manutenção em cada área da fábrica, de 2015 a 2018, totalizou 15.851 eventos. A Linha de Fibras foi responsável por 7.569 ocorrências, o que representa 48% do total medido no período,

seguida das áreas de Secagem e Recuperação e Utilidade, conforme demonstra o gráfico de barras da Figura 12.

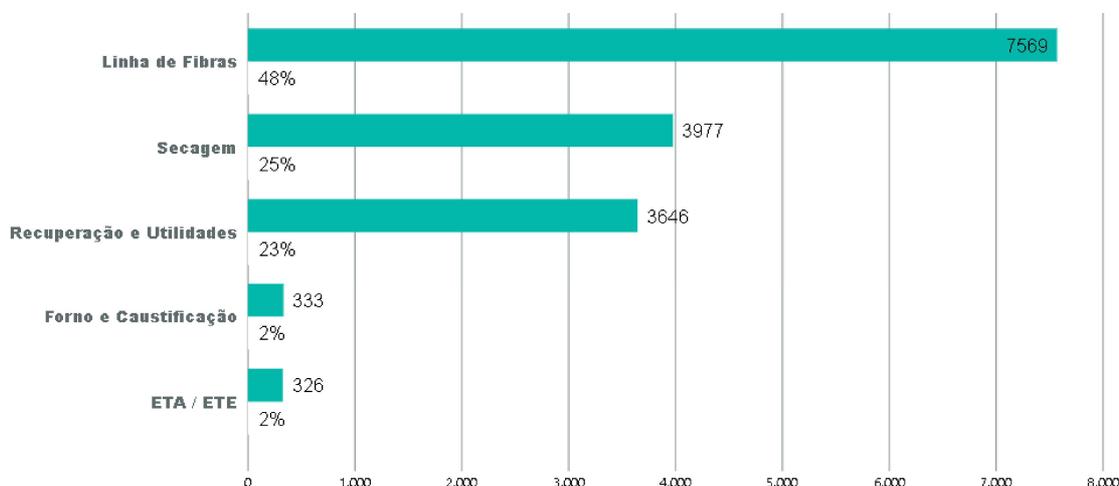


Figura 12. Ocorrências de Manutenção por Macroprocesso ou Área

Dentro da Linha de Fibras, o processo de Preparação de Cavacos, representou 53% de todas as ocorrências de manutenção ou 4.034 ocorrências, seguidos do Cozimento e Branqueamento, conforme gráfico de barras da Figura 13.

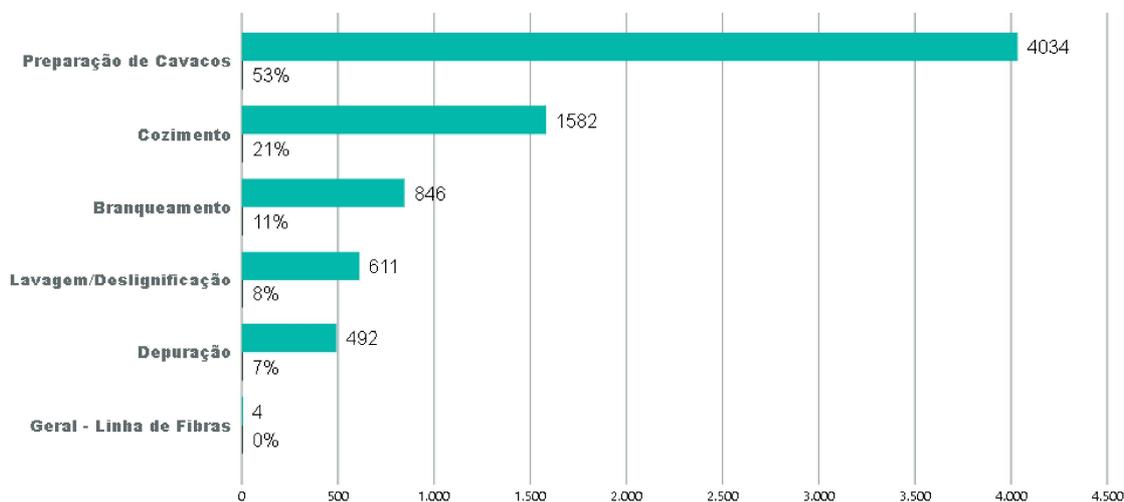


Figura 13. Ocorrências de Manutenção por Processo

O resultado da análise exploratória apresentado acima revela que o processo ou subárea Preparação de Cavacos deve ser objeto de preocupação à gestão da manutenção, pois concentra uma parte significativa dos eventos de anomalias em equipamentos da fábrica, cerca de 25%. Isso pode estar acarretando um importante impacto nos custos de manutenção, mesmo sendo considerada uma subárea de baixa criticidade em termos de efeitos sobre a produção, quando comparada, por exemplo, com a área de cozimento. Tal situação, no entanto, deverá ser comprovada como um desdobramento desta pesquisa.

4.1.2 Macroprocessos e Processos da Fábrica com Maiores Perdas de Produção por problemas de Manutenção

Para perdas de produção, a Linha de Fibras é o Macroprocesso que representou a maior perda de produção motivada por falhas em equipamentos, ou seja 68.734 TSA no período de 4 anos sob estudo nesta pesquisa de um total de 130.767 TSA, conforme demonstra o gráfico de barras da Figura 14. Isso significa que a Linha de Fibras é responsável por 52,6% do total de perdas no período.

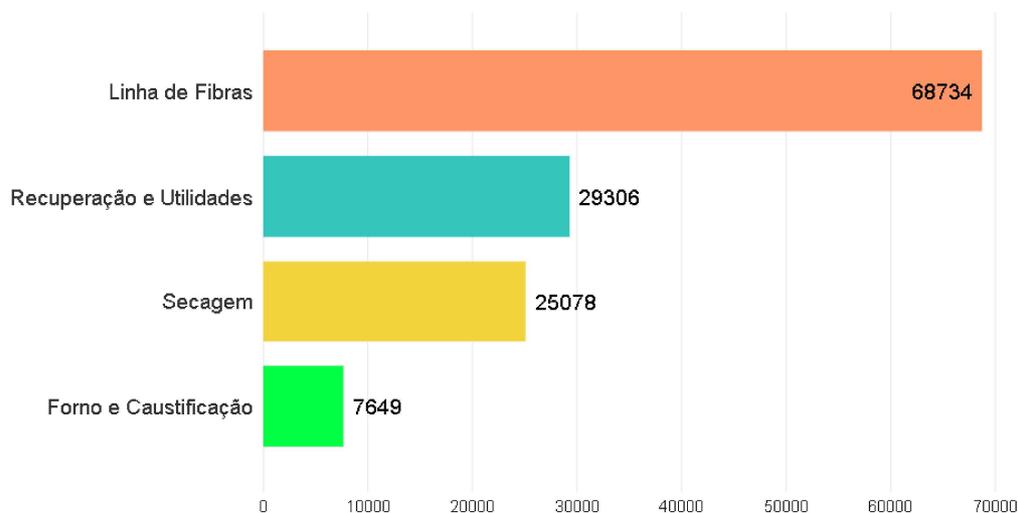


Figura 14. Perda de Produção em TSA (Tonelada Seca ao Ar) por Macroprocesso

Dentro do Macroprocesso Linha de Fibras, conforme demonstra o gráfico circular da Figura 15, o Cozimento foi o Processo onde houve maior perda de produção por falhas em equipamentos, 48.122 TSA, o que corresponde a 70% das perdas dessa Área ou 36,8% de todas as perdas das fábricas no período analisado.

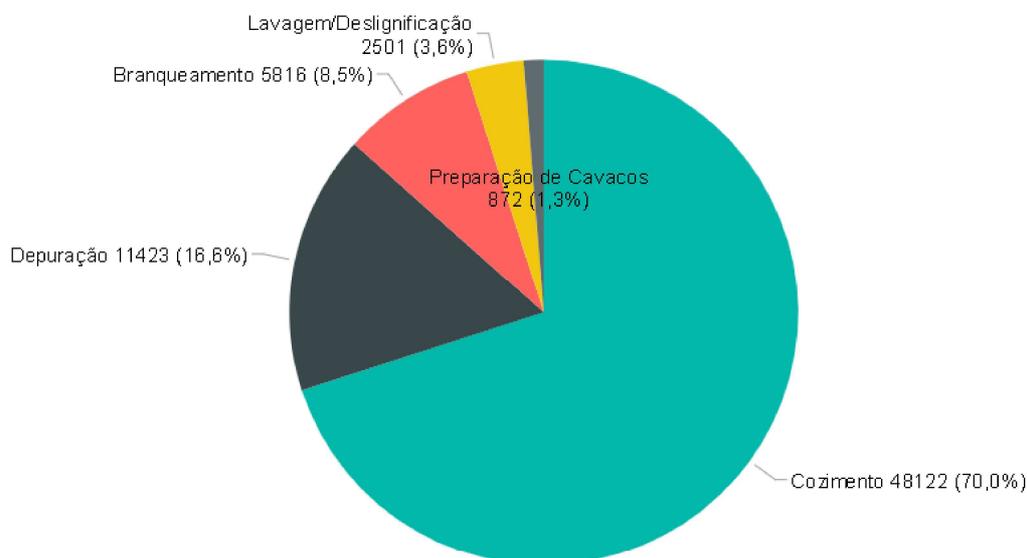


Figura 15. Perda de Produção em TSA por Processo

4.1.3 Equipamentos com mais Ocorrências de Anomalia

De todos os equipamentos da fábrica, os 10 com maior número de ocorrências (Top 10) de anomalias no período sob investigação estão relacionados conforme Tabela 1 e o gráfico de colunas da Figura 16. Destes, é importante destacar que 8 fazem parte do processo “Preparação de Cavacos” da Linha de Fibras.

Tabela 1. Equipamentos com mais Ocorrências de Anomalias por Manutenção

TAG	Descrição da TAG	Contagem de TAG	Processo	Area
20-3102-360	PICADOR L3	419	Preparação de Cavacos	Linha de Fibras
20-3102-260	PICADOR DE TORAS - LINHA 2	274	Preparação de Cavacos	Linha de Fibras
23-3102-330	TRANSPORTADOR DE TORAS L3	271	Preparação de Cavacos	Linha de Fibras
20-3253-167	CORTADEIRA LAYBOY	266	Cortadeira	Secagem
20-3102-110	TRANSPORTADOR DE DESAGUAMENTO	235	Preparação de Cavacos	Linha de Fibras
20-3102-330	MESA DE RECEBIMENTO DE TORAS	197	Preparação de Cavacos	Linha de Fibras
20-3103-080	PENEIRA DE CAVACOS 1	186	Cozimento	Linha de Fibras
23-3103-020	TRANSP ALIMENTACAO PILHA CAVACO STACKER	183	Preparação de Cavacos	Linha de Fibras
23-3102-010	TRANSPORTADOR DE TORAS LINHA 1	168	Preparação de Cavacos	Linha de Fibras
23-3102-230	TRANSPORTADOR DE ALIMENTACAO PICADOR L2	147	Preparação de Cavacos	Linha de Fibras
Total		2348		

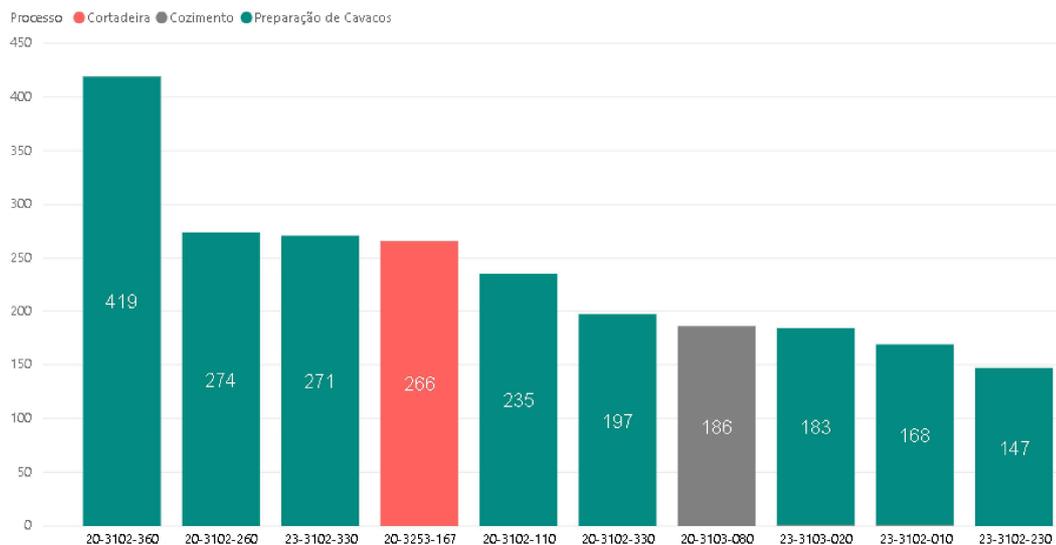


Figura 16. Tags com mais Ocorrências de Anomalias por Manutenção

Nas Figuras 17 e 18, respectivamente gráfico de coluna agrupada e do friso, é possível representar, para esses mesmos equipamentos do Top 10, o número de anomalias de cada equipamento em cada ano pesquisado, ou seja, de 2015 a 2019.



Figura 17. Ocorrências de Anomalia de Manutenção dos Equipamentos Top 10 de 2015 a 2018

No gráfico do friso da Figura 18 é possível identificar de forma mais clara e visual o comportamento de cada equipamento durante o período investigado, evidenciando que todos eles tiveram ocorrências de anomalias ao longo dos quatro anos. Chama a atenção o equipamento Tag 20-3102-360, Picador da Linha 3, que se manteve no topo em todo o período.

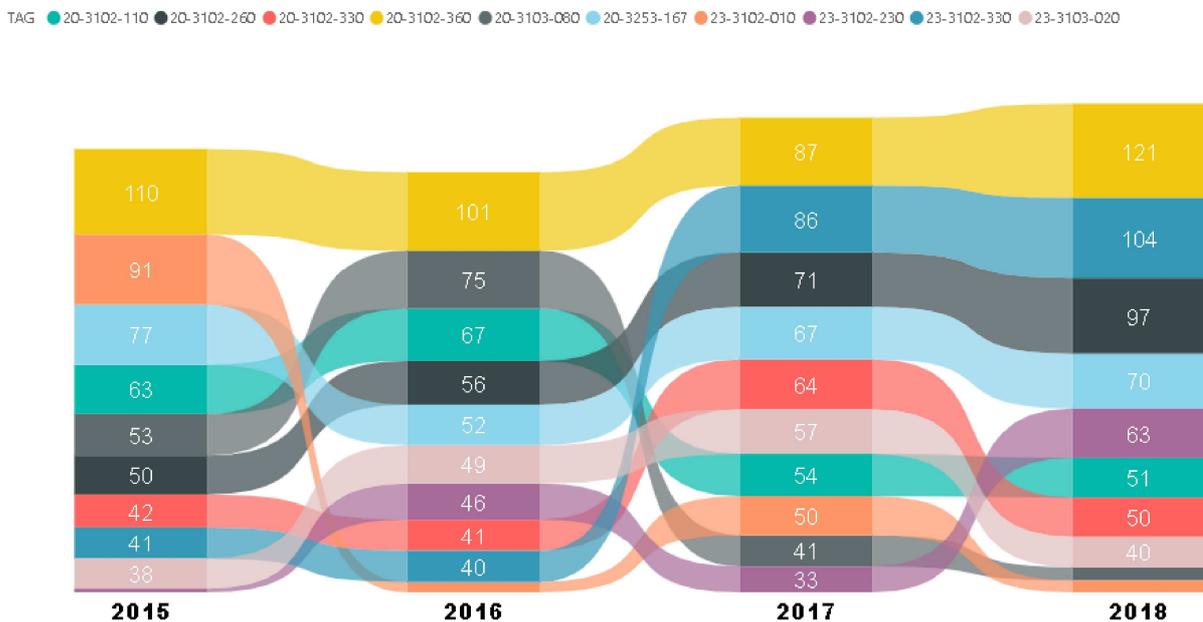


Figura 18. Evolução das Ocorrências de Anomalia de Manutenção dos Equipamentos Top 10 – 2015 a 2018

O gráfico de barras agrupadas da Figura 19 apresenta o número de ocorrências das disciplinas de manutenção que deram origem às intervenções corretivas em cada equipamento top 10. Em todos os equipamentos, a disciplina de Mecânica tem grande preponderância quando comparada com EIA, que são as disciplinas agrupadas de Elétrica, Instrumentação e Automação.

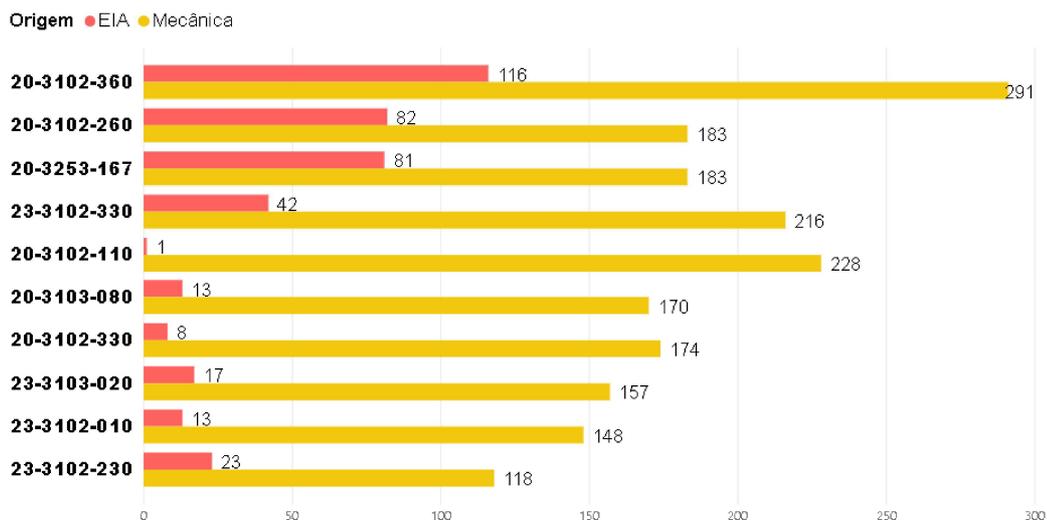


Figura 19. Ocorrências de Anomalia por Disciplinas da Manutenção de Origem nos Equipamentos Top 10

O resultado apresentado nesta seção permite observar o alto índice de ocorrências concentradas em poucos equipamentos da subárea Preparação de Cavacos, na sua maioria em decorrência de falhas mecânicas, e que se repetem ao longo de todos os anos do período avaliado. A regularidade indesejada com que a anomalia ocorre nesses equipamentos deve ser objeto de investigação através de metodologia adequada de análise de causas raiz. Esta oportunidade para investigação pode também ser um desdobramento desta pesquisa.

4.1.4 Equipamentos e Maiores Perdas de Produção do Processo Cozimento

Os equipamentos cujas falhas provocaram as maiores perdas de produção em TSA do período estudado estão relacionados no gráfico de barras da Figura 20. Na Tabela 2 são relacionados aqueles TAGs e sua descrição, onde as perdas foram superiores a mil TSA.

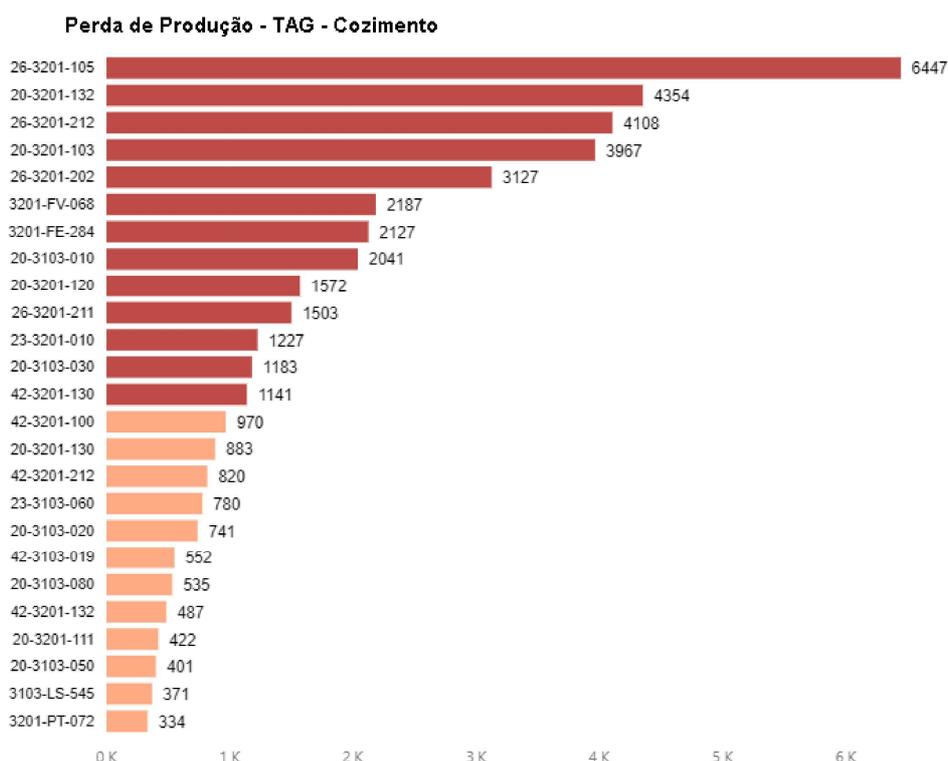


Figura 20. Equipamentos com Maiores Perdas de Produção do Processo de Cozimento em TSA

Tabela 2. Descrição dos Equipamentos com Maiores Perdas e Acima de 1 K TSA

TAG	DESCRIÇÃO
26-3201-105	Bomba de cavacos no. 1
20-3201-132	Raspador de fundo do vaso impregnacao
26-3201-212	Bomba de cavacos no. 3
20-3201-103	Medidor de cavacos com rosca dupla
26-3201-202	Bomba de resfriamento do fundo digestor
3201-FV-068	Val cont vazao para o tubo de cavacos
3201-FE-284	Med vazao de pasta marrom p/ tanque
20-3103-010	Sistema transporte/translação –reclaimer
20-3201-120	Filtro de licor negro
26-3201-211	Bomba de cavacos no. 2
23-3201-010	Transportador de cavaco para o cozimento
20-3103-030	Calha recebimento de cavacos – reclaimer
42-3201-130	Motor separ topo vaso impregnação

No gráfico de barras da Figura 21, estão relacionados os equipamentos com os maiores números de ocorrências de anomalias que provocaram perdas de produção. Importante destacar que todos os 13 equipamentos relacionados com maiores perdas de produção no período, ou seja, acima de mil TSA, estão entre os TOP 18 de maior número de ocorrências de anomalia.

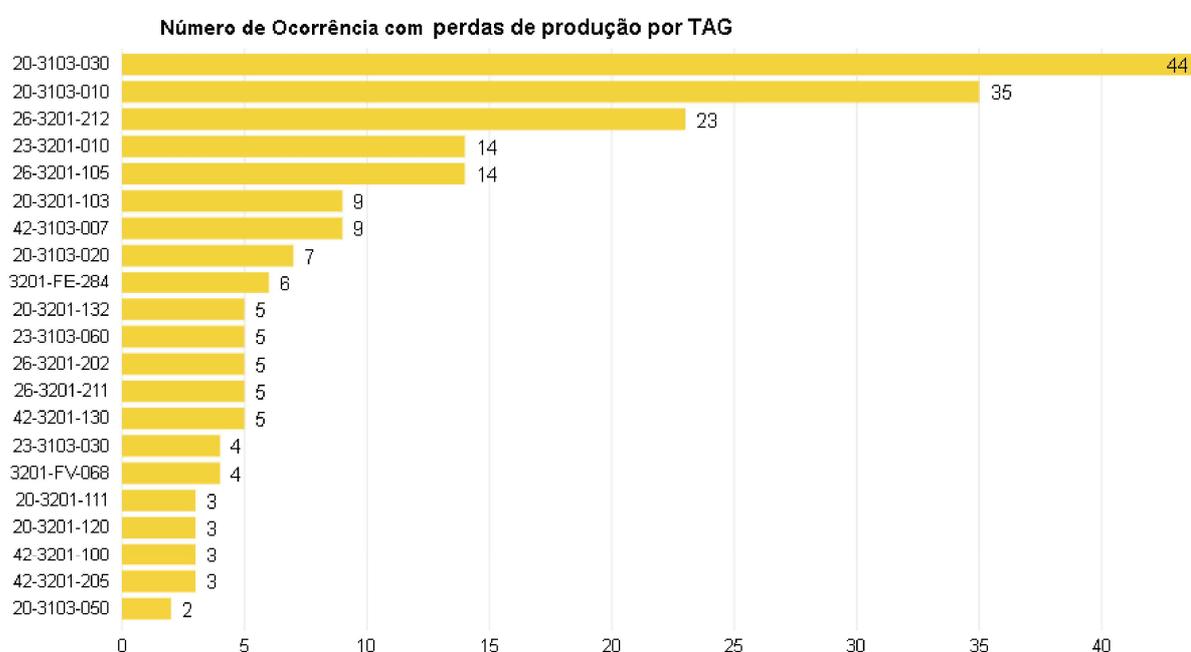


Figura 21. Equipamentos com Mais Ocorrências de Anomalias de Perdas de Produção

Conforme é apresentado no gráfico do friso da Figura 22, assim como no gráfico de colunas empilhadas da Figura 23, é possível observar ao longo dos anos com que frequência ocorrem as anomalias que provocaram perdas de produção e o tamanho dessas perdas em cada equipamento.

Esses recursos de visualização permitem observar, por exemplo, quais dos equipamentos provocaram perdas de produção em todos os anos do período, como é o caso dos TAGs 20-3103-010 e 20-3103-030 ou quais têm provocado as maiores perdas e em qual ano, como é o caso mais preocupante do TAG 26-3201-105, Bomba de Cavacos n.1, TOP 1 em perdas (6.447 TSA), que estão em ascensão ao longo dos últimos 03 anos do período investigado.

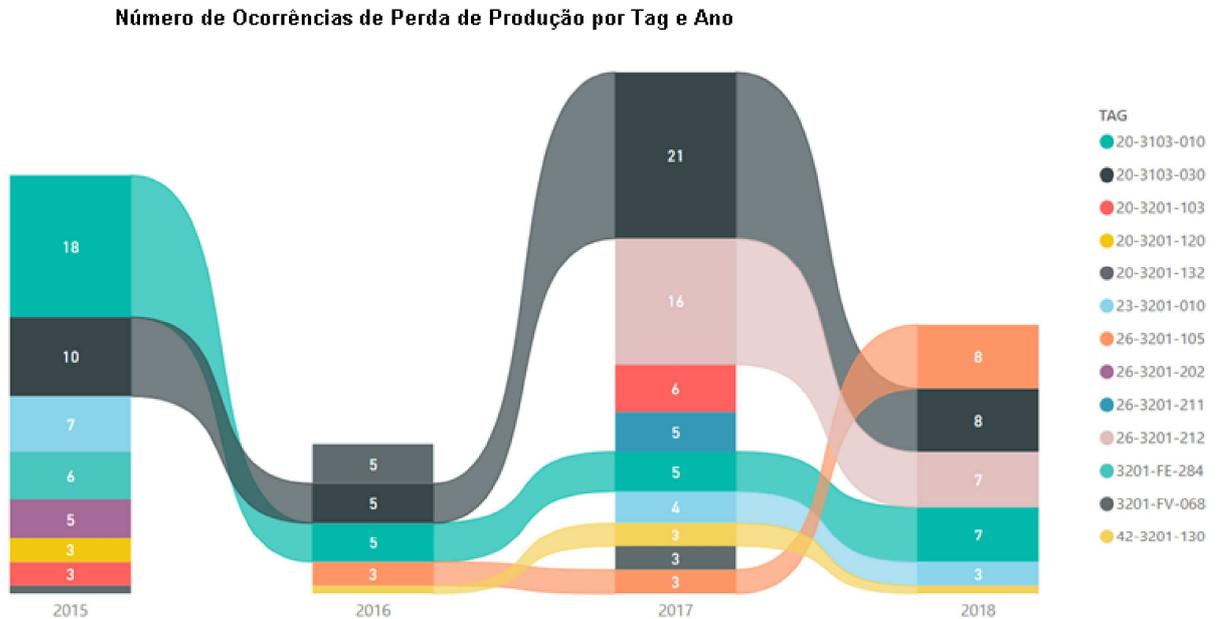


Figura 22. Evolução de Ocorrências de Anomalias de Manutenção dos Equipamentos com as Maiores Perdas de Produção – 2015 a 2018

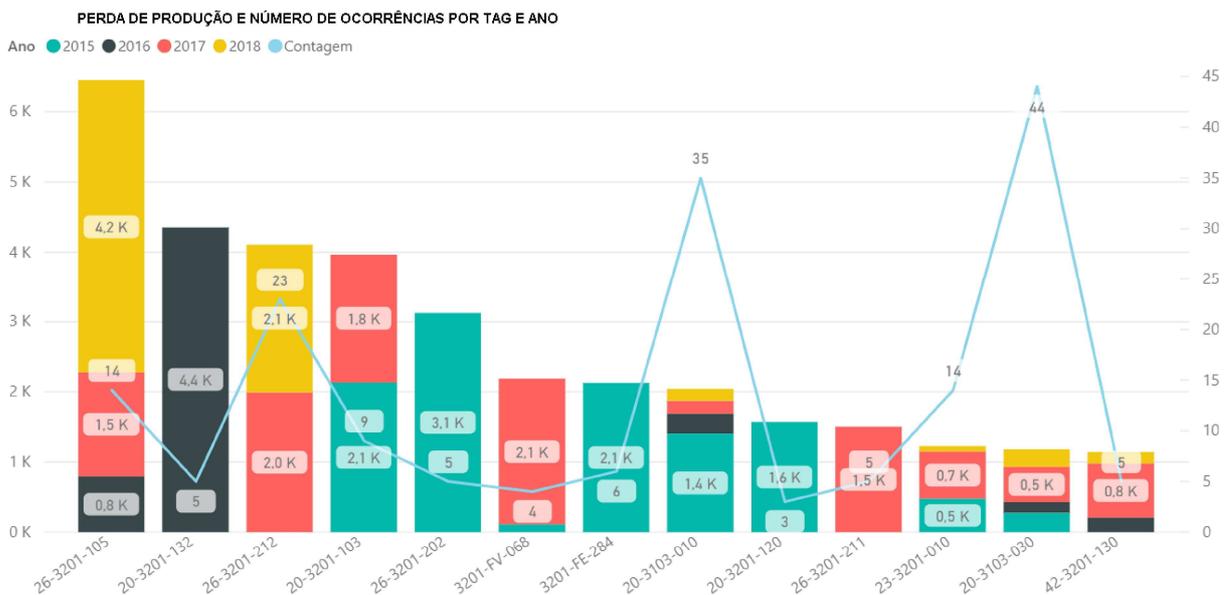


Figura 23. Perdas de Produção (TSA) e Número de Ocorrências de Anomalias de Manutenção dos Equipamentos com as Maiores Perdas de Produção – 2015 a 2018

Quanto às disciplinas de manutenção de origem dos equipamentos com maiores perdas, os gráficos da Figura 24 demonstram que a mecânica é responsável por 73,87% dos casos,

enquanto as disciplinas agrupadas em Elétrica, Instrumentação e Automação (EIA) são responsáveis por 26,13%. Dos 13 equipamentos, 10 tem predominância da disciplina de mecânica e 3 das disciplinas EIA, conforme demonstra o gráfico da Figura 25 de colunas 100% empilhadas.

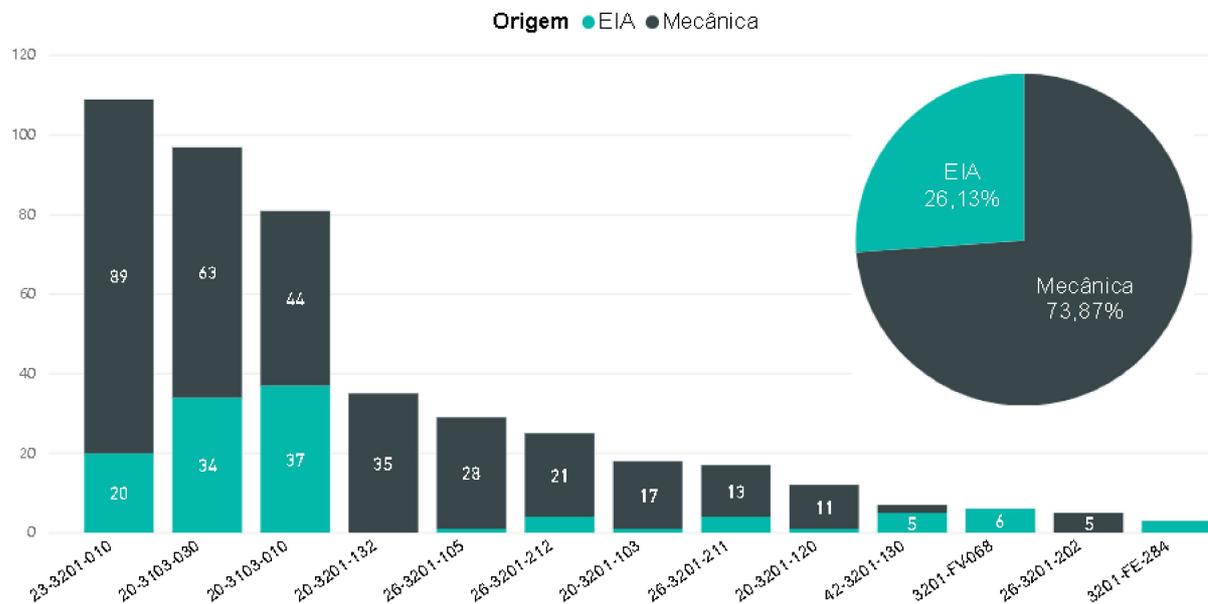


Figura 24. Ocorrências de Anomalia por Disciplinas da Manutenção de Origem nos Equipamentos com as Maiores Perdas de Produção – 2015 a 2018

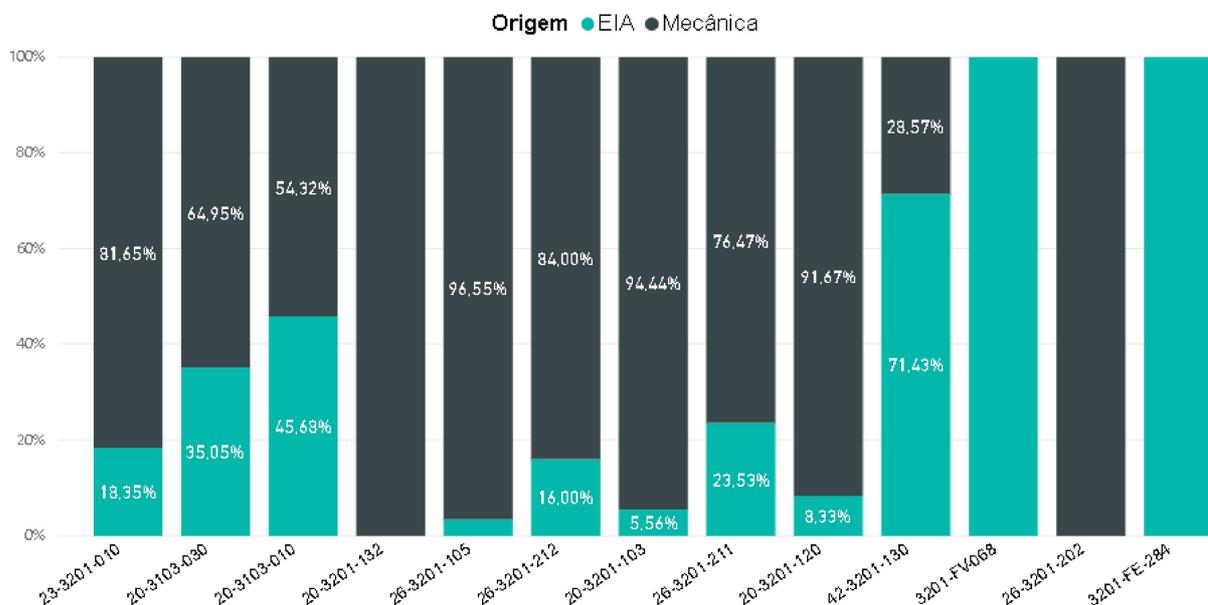


Figura 25. Percentual de Ocorrências de Anomalia por Disciplinas da Manutenção de Origem nos Equipamentos com as Maiores Perdas de Produção – 2015 a 2018

Em relação aos equipamentos que provocaram as maiores perdas de produção há um endereço certo, ou seja, a subárea de cozimento. É possível observar ainda que as perdas se concentram em alguns equipamentos de alta criticidade pelo critério de impacto na produção, que se repetem ao longo dos anos e em alguns casos em preocupante tendência de agravamento

e a maioria são decorrentes de falhas mecânicas. Como desdobramento deste estudo, também aqui e ainda com maior urgência, as falhas recorrentes devem ser objetos de investigação com metodologias adequadas de análise de causas raiz.

Importante enfatizar nesta seção o papel da ferramenta Power BI utilizada para análise exploratória e explícita. Suas características de interface intuitiva, capacidade e flexibilidade para inserir grande diversidade e quantidade de dados, com excelentes recursos de visualização e personalização, entre outras, facilitaram bastante o processo de análise e demonstração dos resultados. Em futuros desdobramentos da pesquisa, em versão corporativa, pode-se lançar mão de outros importantes recursos como os de aprendizado de máquina integrados e desenvolvimento de alertas configurados em KPIs para atualização em métricas e medições relevantes.

4.2 Análise Implícita dos Dados

Nesta seção são apresentados os resultados dos classificadores Naive Bayes e árvore de decisão J48 processados no software WEKA para classificar os dados faltantes do atributo “Origem” e “Processo”.

Antes da realização da classificação, no entanto, foi utilizado na etapa de pré-processamento um filtro para remover os atributos menos importantes para a tarefa. O filtro no Weka escolhido foi do tipo supervisionado e denominado “AttributeSelection”, que contém os recursos necessários para este fim. Após a aplicação do filtro, considerando os atributos a serem classificados, restaram apenas 5 de um total de 24 atributos originais da base de teste. Os atributos que permaneceram foram os seguintes: Origem, Modo de Falha, Família de Equipamento, TAG e Processo. Estes atributos e o tipo de filtro utilizado podem ser visualizados na Figura 26.

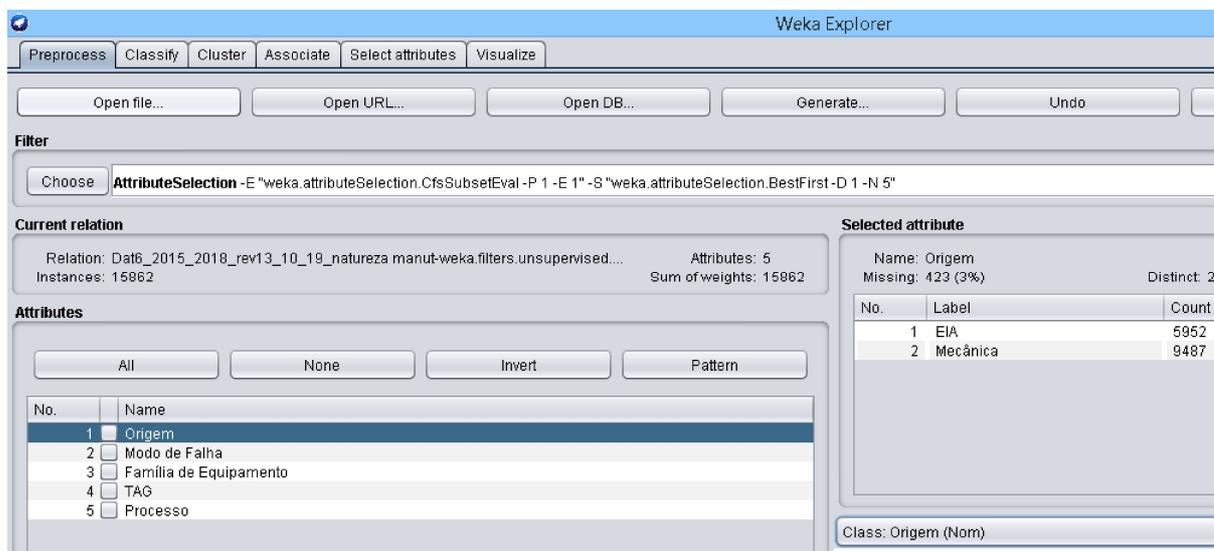


Figura 26. Atributos Filtrados no Pré-processamento

Para avaliação de desempenho dos modelos criados pelos dois algoritmos classificadores foram utilizados os métodos Validação Cruzada (*Cross-Validation*), *Hold Out* ou *Percentage Split* e Matriz de Confusão (*Confusion Matrix*).

4.2.1 Resultados do Classificador Naive Bayes

Para o atributo “Origem”, onde o algoritmo classificador deve classificar as instâncias entre duas opções de classes , EIA e Mecânica, obteve-se os seguintes resultados principais:

- 92,97% das Instâncias foram corretamente classificadas usando o modo de teste de Validação Cruzada com 10 *folds*;
- 93,13% das Instâncias foram corretamente classificadas usando o modo de teste Hold Out ou *Percentage Split 70-30*.

Alguns detalhes dos resultados obtidos no WEKA como acuracidade por classe bem como a matriz de confusão são mostrados na Figura 27.

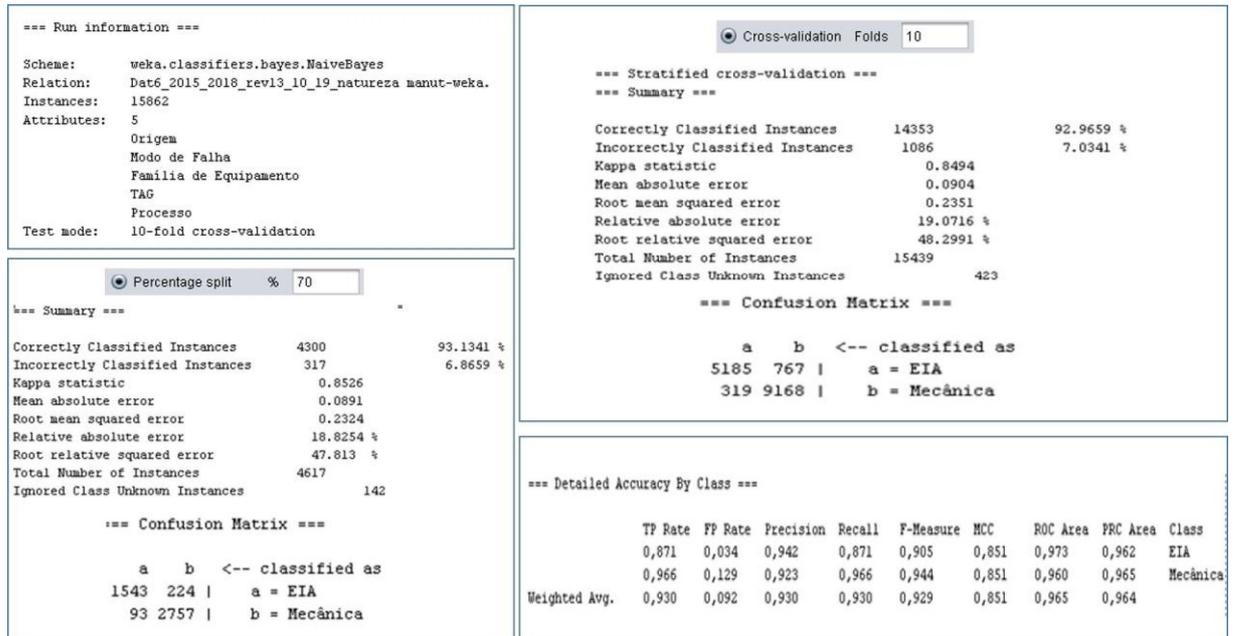


Figura 27. Resultados Weka Classificador Naive Bayes – Atributo Origem

Para o atributo “Process”, onde o algoritmo classificador deve classificar as instâncias entre vinte e cinco opções de classe, obteve-se os seguintes resultados principais:

- 79,02% das Instâncias foram corretamente classificadas usando o modo de teste de Validação Cruzada com 10 *folds*;
- 77,77% das Instâncias foram corretamente classificadas usando o modo de teste Hold Out ou *Percentage Split 70-30*.

Alguns detalhes dos resultados obtidos no WEKA como acuracidade por classe bem como a matriz de confusão são mostrados na Figura 28.

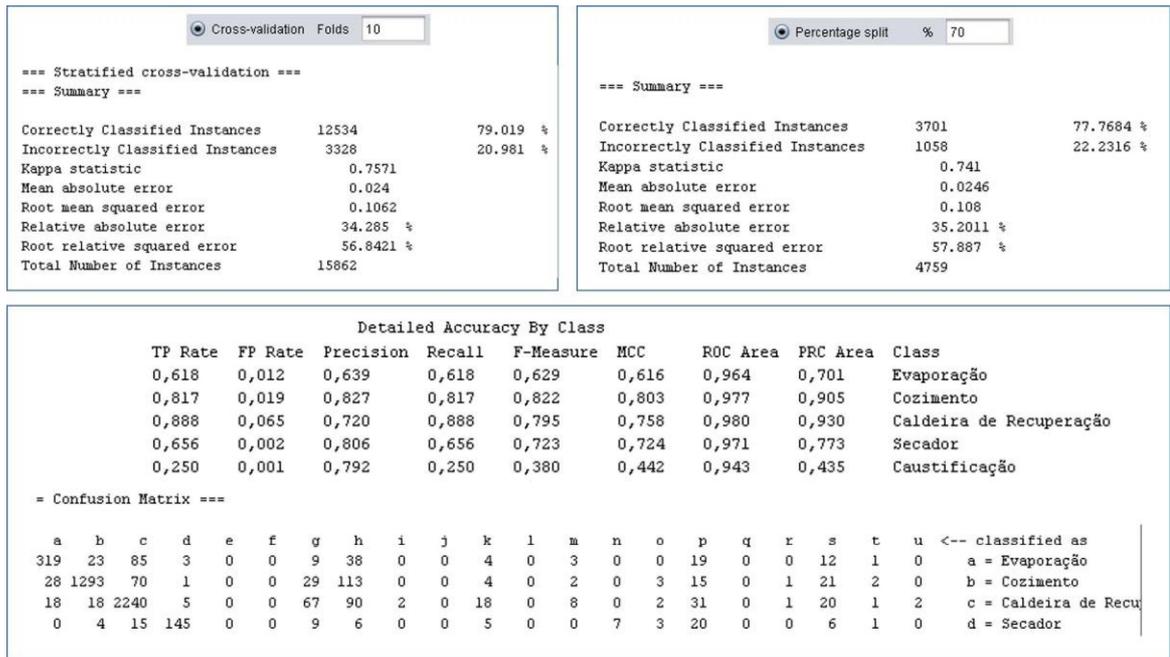


Figura 28. Resultados Weka Classificador Naive Bayes – Atributo Processo

4.2.2 Resultados do Classificador Árvore de Decisão J48

Para o atributo “Origem”, onde o algoritmo classificador deve classificar as instâncias entre duas opções de classes, EIA e Mecânica, obteve-se os seguintes resultados principais:

- 92,23% das Instâncias foram corretamente classificadas usando o modo de teste de Validação Cruzada com 10 folds;
- 91,90% das Instâncias foram corretamente classificadas usando o modo de teste Hold Out ou *Percentage Split* 70-30.

Alguns detalhes dos resultados obtidos no WEKA como acuracidade por classe bem como a matriz de confusão são mostrados na figura 29.

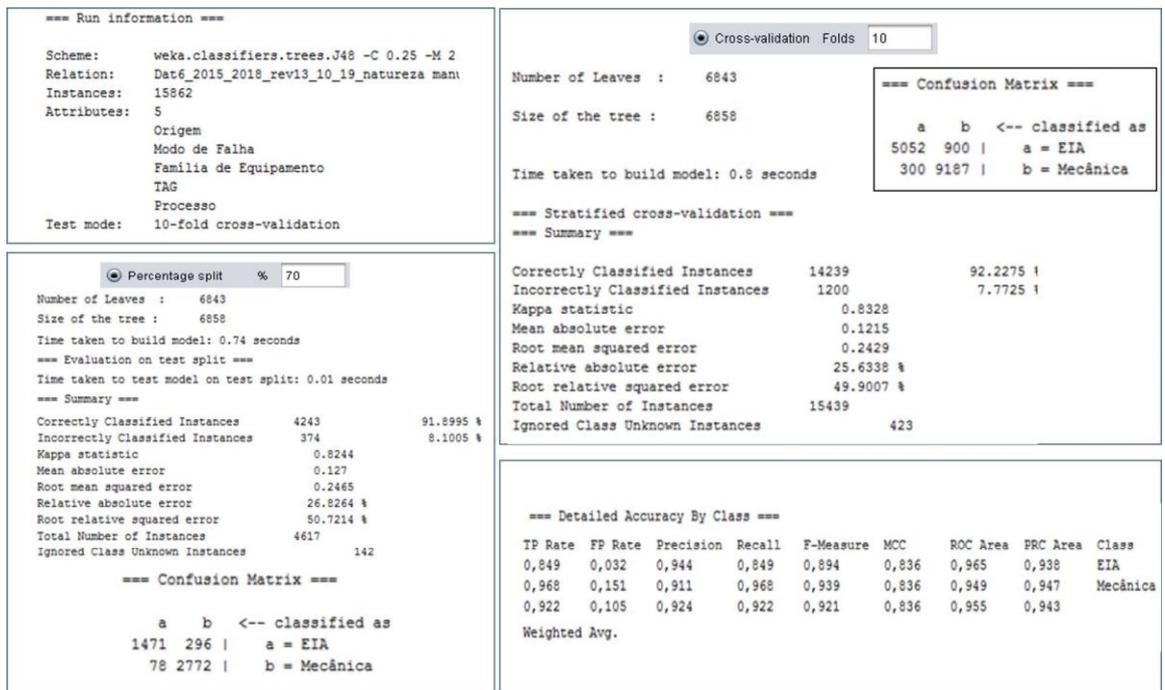


Figura 29. Resultados Weka Classificador J48 – Atributo Origem

Para o atributo “Processo”, onde o algoritmo classificador deve classificar as instâncias entre vinte e cinco opções de classe, obteve-se os seguintes resultados principais:

- 87,04% das Instâncias foram corretamente classificadas usando o modo de teste de Validação Cruzada com 10 *folds*;
- 86,17% das Instâncias foram corretamente classificadas usando o modo de teste Hold Out ou *Percentage Split 70-30*.

Alguns detalhes dos resultados obtidos no WEKA como acuracidade por classe bem como a matriz de confusão são mostrados na Figura 30.

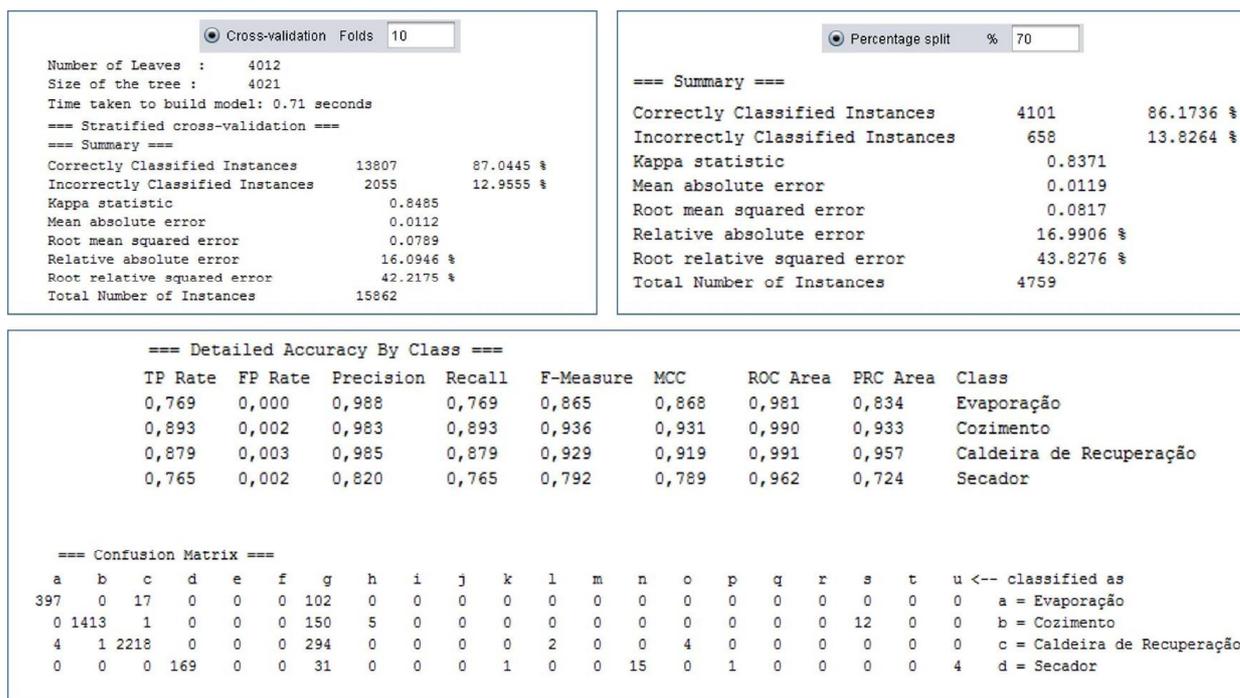


Figura 30. Resultados Weka Classificador J48 – Atributo Processo

Os resultados acima demonstram que ambos algoritmos classificadores, Naive Bayes e J48, aplicados para classificação nos atributos utilizados, “Origem” e “Processo”, foram bastante satisfatórios em termos de acuracidade, permitindo assim preencher os dados faltantes, que são tão importantes para a realização de análises de melhor qualidade, pois contribuem em solucionar lacunas das informações da manutenção.

5. Análises e Discussões dos Resultados

As informações obtidas das análises exploratória e explícita permitem apresentar as principais descobertas em forma de diagnóstico sobre as ocorrências de anomalias e perdas de produção no período sob investigação, ou seja, de 2015 a 2018, a saber:

- A maior parte das anomalias de equipamentos geradores de intervenções de manutenção e perdas de produção concentraram-se na área ou macroprocesso da Linha de Fibras. Esta área representou 48% do total das ocorrências de anomalias e 53% do total de perdas de produção ocorridas no período estudado;

- A subárea ou o Processo de Preparação de Cavacos representou 53% de todas as ocorrências de manutenção da Linha de Fibras, seguidos do Cozimento e Branqueamento, respectivamente 21% e 11%;
- A subárea ou Processo do Cozimento pertencente a área Linha de Fibras, foi responsável por 70% do total das perdas de produção, o que corresponde a 37 % das perdas de toda a fábrica no período investigado;
- Dos equipamentos TOP 10 de ocorrências de anomalias, 8 pertencem ao processo ou subárea Preparação de Cavacos da Linha de Fibras. O alto índice de ocorrências nesses equipamentos se repetem ao longo de todos os anos do período avaliado e a disciplina preponderante envolvida nas intervenções de manutenção é a mecânica. Destaque para o equipamento de TAG 20-3102-360, Picador da Linha 3, que se manteve no topo em todo o período com uma média anual de quase 105 ocorrências;
- Na subárea ou Processo do Cozimento, onde houve as maiores perdas de produção, os 13 equipamentos que provocaram as maiores perdas de produção ou acima de mil TSA no período também estão entre os 18 com maior número de ocorrências de anomalia por defeito ou falhas geradoras de intervenção de manutenção. Destaca-se o equipamento 26-3201-105, Bomba de cavacos no. 1, que é o TOP 1 em perdas no período (6.447 TSA), onde tais perdas se repetem ao longo dos últimos 3 anos e em intensa elevação. Os equipamentos 20-3103-010 e 20-3103-030 também são motivos de preocupação, pois apresentaram importantes perdas em todos os anos do período investigado.
- Quanto às disciplinas de manutenção dos equipamentos responsáveis pelas perdas de produção no processo de Cozimento, a Mecânica foi responsável por 74% contra 26 % das disciplinas agrupadas de Elétrica, Instrumentação e Automação.

Tais descobertas possibilitam alguns desdobramentos ou direcionamentos estratégicos em relação à gestão da manutenção da empresa CEL:

1. Considerando que o enfoque atual da manutenção de equipamentos de uma grande indústria como a CEL é a prevenção das falhas evitando anomalias, indisponibilidade e quebra dos equipamentos, fica evidente através dos resultados das análises de que esta empresa tem muitas oportunidades de melhorias, a começar com aqueles equipamentos TOPs de perdas de produção e TOPs de ocorrências de anomalias, que se repetem ao longo dos anos e que estão concentrados no Macroprocesso da Linha de Fibras, especialmente na subárea do Cozimento devido às grandes perdas. Esta subárea foi responsável por uma perda de 48.122 TSA ao longo do período de 4 anos, uma média anual de mais de 12 mil TSA;
2. Certamente, será necessário rever as estratégias de manutenção para esses equipamentos, desde avaliar os critérios de criticidade dos equipamentos até a revisão dos planos de manutenção preventiva, preditiva e proativa, sobretudo da disciplina de mecânica, objetivando antecipar-se às falhas, diminuído ao máximo as manutenções corretivas, que é o grande desafio da gestão da manutenção.
3. Para garantir a disponibilidade dos equipamentos será necessária a implementação de novas políticas e planos de manutenção fundamentalmente proativos através da intensificação de novas técnicas de inspeções preventivas e preditivas e sempre observando o binômio custo-benefício;
4. Avaliar e revisar as metodologias e a eficácia e eficiência da implementação das análises de falhas ou análises das causas raiz, uma vez que foi constatado que os

equipamentos que mais provocaram perdas de manutenção possuíam várias dessas análises efetuadas, porém as falhas e consequentes perdas de produção continuaram ocorrendo. Da mesma forma, aqueles equipamentos com maior frequência de anomalias e falhas pertencentes à subárea de Preparação de Cavacos e que se repetem com a mesma intensidade ao longo do ano. Nestes casos, é muito importante identificar se o problema está na metodologia de identificação das causas fundamentais ou na implementação das ações para bloquear as causas ou em ambas;

5. Uma boa implementação de técnicas preventivas e preditivas, pressupõe ter planos corretamente elaborados e tecnicamente fundamentados, equipe capacitada, planejamento adequado e uma gestão eficaz. Estes aspectos são críticos e deverão ser devidamente reavaliados;
6. Por último, muito importante verificar o dimensionamento das equipes de manutenção por áreas, disciplinas, tipos de manutenção (corretivas, preventivas e preditivas), etc., se estão em número adequado de profissionais para cada atividade. Por exemplo, na subárea da Linha de Fibras, Preparação de Cavacos, o alto índice de anomalias e falhas de equipamento, mesmo não havendo importantes perdas de produção, demandam uma parte significativa dos recursos humanos e materiais do setor de manutenção impactando inclusive sobre as questões dos custos.

Na análise implícita realizada pelo sistema WEKA de Mineração de Dados em ambiente de aprendizado de máquina, limitado aqui neste trabalho para classificar dados faltantes, identificou-se o seguinte:

- Para o atributo “Origem”, com duas classes, foi possível realizar a classificação utilizando-se dois algoritmos, Naive Bayes e árvore de decisão J48, devidamente validada com uma acuracidade muito satisfatória de 93% para ambos classificadores;
- Para o atributo “Processo”, com 25 classes, utilizando os mesmos algoritmos do atributo “Origem” e com os mesmos métodos de validação, obteve-se um resultado bastante razoável de 78% de acuracidade utilizando-se o classificador Naive Bayes e um resultado ainda melhor de 87% de acuracidade com o classificador J48.

Tais resultados permitem demonstrar que os recursos de mineração de dados como o sistema WEKA, no contexto do escopo deste trabalho, são bastante efetivos em termos de classificação de dados faltantes e, por isso, deveriam ser incorporados no pré-processamento e processamento de dados da manutenção.

Importante observar que assim como os dados registrados erroneamente no sistema da manutenção, os dados faltantes também são bastante comuns, pois em muitos campos do módulo PM do sistema SAP, por exemplo, os preenchimentos não são obrigatórios e no universo da manutenção de uma indústria essa situação pode prejudicar bastante as análises dos dados. Ou seja, tais recursos podem contribuir de maneira decisiva para melhorar a qualidade das informações.

Além da aplicação utilizada nesta investigação, de classificação, as técnicas ou tarefas de mineração de dados podem contribuir significativamente com a gestão do setor de manutenção, em demandas que exijam predição, identificação de padrão por agrupamento, por associação e outros.

É relevante também considerar que a empresa CEL, apesar de contar com os principais sistemas de ERP e Controle de Processo, bem como ativos ou equipamentos de última geração, tem ainda grandes oportunidades para melhorar o aproveitamento dos seus dados relacionados

à manutenção e, com isso, extrair conhecimentos importantes para melhorar a eficiência da gestão deste setor, uma vez que ainda não possui uma política de utilização sistemática de ferramentas para a realização de análises explícitas e, tampouco, implícitas.

6. Considerações Finais

Com a evolução da tecnologia digital, torna-se cada vez mais imprescindível o desenvolvimento ou a implementação de recursos ou ferramentas computacionais que possibilitem a análise e a interpretação das informações de forma sistemática e automática transformando-as em conhecimentos que possam ser úteis à empresa. Por outro lado, sem lançar mão desses recursos de interpretação correta dos dados e com extração do conhecimento, a empresa passa a analisar e tomar decisões com bases empíricas e até intuitivas, o que pode resultar em equívocos em nível tático e estratégico.

Os resultados obtidos nesta pesquisa demonstram que as metodologias de descobertas de conhecimentos em base de dados relacionados à manutenção podem contribuir de forma decisiva nos direcionamentos estratégicos da gestão da manutenção como, por exemplo, a necessidade de reavaliar políticas e planos de manutenção, adoção de novas técnicas proativas, capacitação de equipe, etc. Por isso, a adoção de forma sistemática da metodologia de descoberta de conhecimento é um desafio importante para a empresa CEL.

Outro desdobramento relevante deste trabalho, a partir das análises, é a ênfase dada a necessidade de uma investigação mais criteriosa e com metodologias adequadas, incluindo a da descoberta de conhecimentos com base nos dados, para aqueles equipamentos que apresentaram falhas recorrentes no período estudado e com alto impacto tanto para produção como para os custos da manutenção.

A pesquisa ficou limitada à exploração e análise dos registros de ocorrências de anomalias de manutenção e perdas de produção e é importante que se estenda a metodologia para outros dados fundamentais como: as análises de falhas para identificação de causas raiz, dados de monitoramento da condição dos equipamentos, detalhes das manutenções preventivas, preditivas e corretivas, modos de falhas, etc. Com grandes volumes de dados e sua diversificação (estruturados e não estruturados), provavelmente serão demandadas outras técnicas de mineração de dados e algoritmos.

A adequada exploração e extração de conhecimentos desses dados podem trazer importantes contribuições para o principal objetivo estratégico atual da manutenção que é se antecipar a falha do equipamento, evitando assim sua indisponibilidade e conseqüente perda de produção.

7. Referências

- Amaral, Fernando (2016). Aprenda Mineração de Dados: teoria e prática. Rio de Janeiro: Alta Books.
- Amaral, Fernando (2017). Introdução à Ciência de Dados: Mineração de Dados e Big Data. Rio de Janeiro: Alta Books.
- Bachmann & Associados (2013). Glossário – Indicadores Bachmann para Celulose e Papel. Disponível em: < http://www.bachmann.com.br/website/documents/ACPGlossario1_000.pdf >. Acesso em: 22 ago. 2019.
- Cabral, J. (2006). Organização e Gestão da Manutenção, dos conceitos à prática. Editora Lidel, 6ª edição.

- Castro, H.F (2009). Processos Químicos Industriais II – Papel e Celulose - Apostila 4 – Escola de Engenharia de Lorena – Universidade de São Paulo. São Paulo.
- CEPEA-ESALQ/USP (2017). Informativo Setor Florestal nº 193. Centro de Estudos Avançados em Economia Aplicada – Economia Florestal. São Paulo. Disponível em: < <https://www.cepea.esalq.usp.br/upload/revista/pdf/0019775001519655993.pdf> >. Acesso em: 26 ago. 2019.
- DIN-EN-13306 (2018). *Maintenance – Maintenance terminology - Trilingual version EN 13306:2017*. Deutsches Institut für Normung.
- Fuentes, F. (2006). Metodologia para inovação da gestão de manutenção industrial. Tese, Programa de Pós-Graduação em Engenharia Mecânica, Universidade Federal de Santa Catarina. Florianópolis.
- Fayyad, U., Piatetski-Shapiro, G., Smyth, P (1996). *From Data Mining to Knowledge Discovery*. America Association for Artificial Intelligence.
- Gerhardt, T.E. e Silverira, D.T. (2009). Métodos de Pesquisa. Editora UFRGS, 1ª edição.
- Gil, A.C (2008). Métodos e Técnicas de Pesquisa Social. São Paulo: Editora Atlas S. A., 6ª edição.
- Gil, A.C (2010). Como elaborar projetos de pesquisa. São Paulo: Editora Atlas S. A., 5ª edição.
- ISO-13372 (2012). *Condition monitoring and diagnostics of machines – Vocabulary*. International Organization for Standardization.
- Kardec, A., Nascif, J. (2001). Manutenção: Função Estratégica. Rio de Janeiro: Editora Qualitymark, 2.ª Edição.
- Malhotra, N.R. (2001). Pesquisa de marketing: uma orientação aplicada. Porto Alegre: Bookman.
- Matsumota, L. (2016). Construção de *Business Intelligence* para geração de valor a uma siderúrgica: aplicação na manutenção industrial. Dissertação, programa de pós-graduação em Engenharia de Produção, Universidade Paulista. São Paulo.
- Pinheiro, C. A. R. (2008). Inteligência Analítica – Mineração de Dados e Descoberta de Conhecimento. Rio de Janeiro: Editora Ciência Moderna Ltda.
- Pinto, A. K., & Ribeiro, H. (2002). Gestão Estratégica e Manutenção Autônoma. Rio de Janeiro: Qualimark.
- Ribeiro, M.A (2001). Controle de Processo – Teoria e Aplicações. 7. ed. Salvador. Disponível em: < <https://www.doccity.com/pt/marco-antonio-ribeiro-utomacao-processo-etc-controle-processo-q/4812932/> > Acesso em: 22 jul. 2019.
- Richardson, R.J (1999). Pesquisa social: métodos e técnicas. 3.ed. São Paulo: Atlas.
- Silva, I.A.F (2007). Descoberta de conhecimento em base de dados de monitoramento ambiental para avaliação de qualidade da água. Dissertação, Programa de Pós-Graduação em Física e Meio Ambiente, Universidade Federal de Mato Grosso. Cuiabá.
- Yin, R. K (2015). Estudo de caso: planejamentos e métodos. 4.ed. Porto Alegre: Bookman.