

**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

RILDO DE MENDONÇA NOGUEIRA

**ANÁLISE DOS IMPACTOS HARMÔNICOS EM UMA INDÚSTRIA DE
MANUFATURA DE ELETROELETRÔNICOS UTILIZANDO ÁRVORES DE
DECISÃO**

DM 452/2015

**UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2015**

**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

RILDO DE MENDONÇA NOGUEIRA

**ANÁLISE DOS IMPACTOS HARMÔNICOS EM UMA INDÚSTRIA DE
MANUFATURA DE ELETROELETRÔNICOS UTILIZANDO ÁRVORES DE
DECISÃO**

Dissertação de Mestrado apresentada à coordenação do Programa de Pós-Graduação em Engenharia Elétrica (PPGEE) do Instituto de Tecnologia da Universidade Federal do Pará (UFPA) como parte dos requisitos finais para obtenção do grau de Mestre em Engenharia Elétrica na área de Sistemas de Energia Elétrica.

**UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil
2015**

Nogueira, Rildo de Mendonça , 1969-

Análise dos impactos harmônicos em uma indústria de manufatura de eletroeletrônicos utilizando árvores de decisão / Rildo de Mendonça Nogueira. - 2015.

Orientadora: Maria Emilia de Lima Tostes;

Coorientador: Ádamo Lima de Santana.

Dissertação (Mestrado) - Universidade Federal do Pará, Instituto de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Belém, 2015.

1. Sistemas de energia elétrica -controle de qualidade. 2. Mineração de dados. 3. harmônicos (ondas elétricas) - análise. 4. Indústria eletrônica - modelos matemáticos. I. Título.

CDD 22. ed. 621.3191

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**“ANÁLISE DOS IMPACTOS HARMÔNICOS EM UMA
INDÚSTRIA DE MANUFATURA DE ELETROELETRÔNICOS
UTILIZANDO ÁRVORES DE DECISÃO”**

AUTOR: RILDO DE MENDONÇA NOGUEIRA

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO
COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO
JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA
ELÉTRICA NA ÁREA DE SISTEMAS DE ENERGIA ELÉTRICA.

APROVADA EM: 27/03/2015

BANCA EXAMINADORA:



Prof.ª Dr.ª Maria Emília de Lima Tostes

(Orientadora – PPGEE/UFPA)



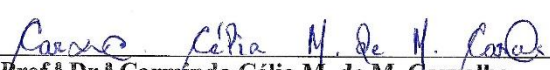
Prof. Dr. Adamo Lima de Santana

(Co- Orientador PPGEE/UFPA)



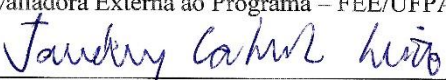
Prof. Dr. Roberto Célio Limão de Oliveira

(Avaliador Interno – PPGEE/UFPA)



Prof.ª Dr.ª Carminda Célia M. de M. Carvalho

(Avaliadora Externa ao Programa – FEE/UFPA)



Prof. Dr. Jandecy Cabral Leite

(Avaliador Externo – ITEGAM)

VISTO:

Prof. Dr. Evaldo Gonçalves Pelaes

(Coordenador do PPGEE/ITEC/UFPA)

AGRADECIMENTOS

À Deus pela vida que me foi confiada.

Aos meus pais por todos os ensinamentos vivenciados desde os primeiros momentos de minha vida.

Aos meus queridos irmãos pelo apoio, compreensão e união.

À Universidade Federal do Pará - UFPA.

Ao Instituto de Tecnologia e Educação Galileo da Amazônia – ITEGAM.

Aos Professores, Dr. Ádamo Lima de Santana e Dra. Maria Emília de Lima Tostes, Dr. Jandecy Cabral Leite, pela condução e orientação com excelência durante o curso.

Ao Instituto CERTI Amazônia, Universidade Paulista UNIP (Campus de Manaus), colega de mestrado e amigo M.Sc. Hindenburgo Elvas Gonçalves de Sá, pelo incentivo nesses meses de estudo e na preparação da dissertação.

Aos colegas e professores do curso.

Ao meu pai Raimundo, mãe Maria de Lourdes, que me incentivam em todos os momentos da minha vida.

Ao professor Wilson Roberto Malavazi e a professora Rosange Menezes pelo apoio e compreensão.

À professora Dra. Antonieta Vieira e o Professor Dr. Katsuhito Takita, que me incentivaram em vários momentos da preparação de minha dissertação.

A todos que direta e indiretamente contribuíram para realização deste trabalho.

Dedicatória

Gostaria de dedicar este trabalho em primeiro lugar a Deus, a minha Mãe e Meu Pai e a duas pessoas muito especiais em minha vida: Meu Filho Emanuel e minha esposa Edinalva.

RESUMO

A Qualidade da Energia Elétrica (QEE) é constantemente alvo de vários estudos, principalmente os que estão relacionados com o setor de produção industrial, onde estão concentradas grandes cargas dos sistemas elétricos. Com a evolução dos processos de produção industrial e a inserção de novas tecnologias no setor industrial, foram adicionadas grandes quantidades de equipamentos eletrônicos que são fontes geradoras de perturbações nos sistemas, e que alteram a qualidade do produto “energia elétrica”. Com o intuito de minimizar os transtornos decorrentes da baixa qualidade da energia e os prejuízos causados para as concessionárias e consumidores (industriais, residenciais e comerciais), foi desenvolvido no Brasil, os procedimentos de distribuição no sistema elétrico nacional (PRODIST), criado e desenvolvido pela Agência Nacional de Energia Elétrica (ANEEL). O PRODIST tem o objetivo de normatizar e padronizar as atividades relacionadas à distribuição de energia, incluindo os padrões de qualidade do produto. O presente trabalho foi concentrado e realizado em uma empresa do polo industrial de Manaus (PIM), que possui um sistema elétrico trifásico de baixa tensão, com objetivo de monitorar a qualidade do produto “energia elétrica” através do conteúdo harmônico gerado pela rede elétrica envolvida na manufatura. Os dados gerados foram submetidos a uma técnica de inteligência computacional (IC), com uso do processo de descoberta de extração de conhecimento em bases de dados ou KDD. O objetivo do trabalho é analisar, identificar e diagnosticar os pontos de acoplamento e processos que apresentam conteúdos harmônicos representativos para o sistema, podendo-se assim verificar o quanto cada processo analisado pode estar afetando a qualidade da energia dentro da própria indústria e no ponto de acoplamento com a concessionária, através das gerações de distorções harmônicas, evitando assim, penalizações e outras sanções regulamentadas.

Palavras-chave: Inteligência Computacional, Mineração de Dados, Distorções harmônicas de corrente, Ponto de acoplamento comum, Qualidade da Energia Elétrica.

ABSTRACT

The Power Quality (PQ) is constantly the subject of many studies, mainly those that are related to the industrial production sector, where are concentrated large loads of the electrical systems. With the evolution of industrial production processes and the introduction of new technologies in the industrial sector, quantities of electronic equipment bars were added that are sources of disturbances in the systems, and affecting the quality of the product "electricity". In order to minimize the inconvenience resulting from low quality of energy and damage to utilities and consumers (industrial, commercial and residential), it was developed in Brazil, distribution procedures in the national grid (PRODIST), created and developed the National Electric Energy Agency (ANEEL). The PRODIST aims to regulate and standardize activities related to energy distribution, including product quality standards. This work was concentrated and held in a company of the industrial pole of Manaus (PIM), which has a three-phase electrical system low voltage, in order to monitor the quality of the product "electricity" through the harmonic content generated by the electrical network involved in manufacturing. The data generated were subjected to computational intelligence technique (HF), using the process of knowledge extraction discovery in databases or KDD. The objective is to analyze, identify and diagnose the coupling points and processes that have representative harmonic content for the system, so being able to check how much each analyzed process may be affecting the power quality within the industry itself and the point of coupling with the concessionaire, through the generations of harmonic distortion, thus avoiding penalties and other sanctions regulated.

Keywords: Computational Intelligence, Data Mining, distortions harmonic current, common coupling point, Power Quality.common coupling point, Power Quality.

LISTA DE FIGURAS

Figura 1	Processo KDD.....	44
Figura 2	Fase de Seleção do KDD.....	45
Figura 3	Fase de Transformação de dados do KDD.....	46
Figura 4	Processo de Aprendizagem.....	47
Figura 5	Exemplo de uma árvore de decisão.....	48
Figura 6	Componentes de uma árvore de decisão.....	49
Figura 7	Dados originais dos arquivos CSV.....	61
Figura 8	Árvore de decisão da Fase A.....	68
Figura 9	Árvore de decisão da Fase B.....	69
Figura 10	Árvore de decisão da Fase C.....	70
Figura 11	Curva ROC da Fase A.....	74
Figura 12	Curva ROC da Fase B.....	74
Figura 13	Curva ROC da Fase C.....	74

LISTA DE TABELAS

Tabela 1	Valores de referência globais das distorções harmônicas totais (em porcentagem da tensão fundamental)	42
Tabela 2	Níveis de referência para distorções harmônicas individuais de tensão.....	42
Tabela 3	Características da Subestação de entrada.....	56
Tabela 4	Centrais de ar-condicionado.....	56
Tabela 5	Linha de Testes de notebooks – Burning test.....	57
Tabela 6	Sala de Compressores e Bombas.....	57
Tabela 7	Montagem de Tablets.....	58
Tabela 8	Características dos analisadores de QEE.....	58
Tabela 9	Resumo dos dados do arquivo CSV	60
Tabela 10	Características dos campos das tabelas.....	60
Tabela 11	Redução e formatação dos dados	62
Tabela 12	Nomenclatura das variáveis nas reduções da fase A.....	63
Tabela 13	Matriz Confusão da Fase A	65
Tabela 14	Acurácia dos dados Fase A	65
Tabela 15	Matriz Confusão da Fase B	65
Tabela 16	Acurácia dos dados Fase B	65
Tabela 17	Matriz Confusão da Fase C	66
Tabela 18	Acurácia dos dados Fase C	66
Tabela 19	Sumário de análises de fase – Mineração de dados.....	66
Tabela 20	Termos utilizado no desempenho do classificador.....	71
Tabela 21	Matriz de Confusão Genérica.....	72

LISTA DE ABREVIATURAS E SIGLAS

ANEEL	Agencia Nacional de Energia Elétrica
KDD	Knowledge Discovery in Databases
PRODIST	Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional
DM	Data mining - Mineração de Dados
IC	Inteligência Computacional
SEP	Sistema Elétrico de Potência
DTS	Técnica de árvores de decisão (DTS)
PAC	Ponto de Acoplamento Comum
IEC	Comissão Eletrotécnica (International Electro technical Commission)
THD	Distorção Total de Harmônicas
QEE	Qualidade da Energia Elétrica
ONS	Operador Nacional do Sistema Elétrico
DIT	Demais Instalações de Transmissão
FP	Fator de Potência
AD	Árvore de Decisão
RNAs	Redes Neurais Artificiais
CSV	Comma separated values
SGBD	Sistemas de Gerenciamento de Bancos de Dados
SOM	Sistemas Operacionais de Máquinas
THDu	Distorção Harmônica Total de Tensão
THDv	Distorção Harmônica Total de Tensão
TDHi	Distorção Harmônica Total de Corrente
WEKA	Waikato Environment for Knowledge Analysis
SQL	Structured Query Language
FP	Falso Positivo
TP	Verdadeiro Positivo
Recall	Sensitividade
ROC	Receiver Operating Characteristic
PQ	Qualidade da Energia Elétrica (Power Quality)

LISTA DE EQUAÇÕES

Equação 1	Cálculo da Distorção harmônica total de tensão.....	38
Equação 2	Cálculo da Distorção harmônica individual de tensão.....	39
Equação 3	Cálculo de índice gini.....	52
Equação 4	Cálculo de entropia.....	52
Equação 5	Cálculo de ganho de informação.....	53
Equação 6	Cálculo de TP RATE (Taxa de Verdadeiros Positivos)	71
Equação 7	Cálculo de FP RATE (Taxa de Falsos Positivos)	71
Equação 8	Cálculo da Precisão do Modelo.....	72
Equação 9	Cálculo do Índice KAPPA.....	72

SUMÁRIO

CAPÍTULO 1	15
1.1 Contextualização.....	15
1.2 Identificação e Justificativa do Problema de Pesquisa.....	17
1.3 Objetivos.....	18
1.3.1 Objetivo Geral	18
1.3.2 Objetivo Específico	18
1.4 Contribuição e Relevância do Tema.....	18
1.5 Delimitação da Pesquisa.....	19
1.6 Estrutura do Trabalho.....	20
CAPÍTULO 2	21
2.1 Estado da Arte.....	21
2.2 Trabalhos.....	21
CAPÍTULO 3	36
3.1 Introdução.....	36
3.2 Conceituação.....	37
3.3 Normas na Qualidade da Energia Elétrica.....	37
3.3.1 A norma IEC 61000-3-6 (IEC/TR,2008-02)	38
3.3.2 A norma Brasileira ANEEL-PRODIST Módulo 8 Revisão 6 (2015)	39
3.3.2.1 Qualidade do Produto.....	39
3.3.3 Distorções Harmônicas	40
3.3.3.1 Procedimentos de Medições de Harmônicos.....	41
3.3.3.2 Os Valores de Referências das Distorções Harmônicas	41
3.4 Considerações Finais.....	42
CAPÍTULO 4	43
Processo do KDD e Os Modelos de Inteligência Computacional.....	43
4.1 Considerações Iniciais.....	43
4.2 Processo do KDD.....	44
4.2.1 Fase de Seleção	44
4.2.2 Fase de Pré-Processamento	45
4.2.3 Fase de Transformação	46
4.2.4 Fase de Mineração de Dados	46
4.2.5 Fase de Interpretação de Dados	47
4.3 Algoritmos de Inteligência Computacional.....	48
4.3.1 Árvore de Decisão	48
4.3.2 Algoritmos Utilizados em Árvores de Decisão	50
4.3.3 Classificação de uma Árvore de Decisão	51
4.3.4 Aplicação do Classificador C4.5 ou J48	51
4.3.5 Processo de Poda	53
4.3.5.1 Pré-Poda.....	53
4.3.5.2 Pós-Poda.....	54
4.4 Considerações Finais.....	54

	CAPÍTULO 5	55
5.1	Introdução.....	55
5.2	Características da Indústria Analisada.....	55
5.2.1	Pontos Analisados da Indústria	56
5.2.1.1	Subestação.....	56
5.2.1.2	Centrais de Arcondicionados.....	56
5.2.1.3	Linhas de Testes de notebooks (Burning).....	57
5.2.1.4	Compressores e Bombas.....	57
5.2.1.5	Produção de Tablets.	57
5.2.2	Equipamentos Utilizados nas Medições	58
5.2.3	Metodologia Computacional Empregada	59
5.2.3.1	Execução da etapa de Seleção.....	59
5.2.3.2	Execução da etapa de Pré-Processamento.....	61
5.2.3.3	Execução da Etapa de Transformação.....	64
5.2.3.4	Execução da Etapa de Mineração de dados.....	64
5.2.3.5	Execução da Etapa de Interpretação de dados.....	66
5.2.3.6	Métricas para avaliação desempenho de um classificador.....	70
5.3	Considerações Finais.....	75
	CAPÍTULO 6	76
6.1	Conclusões.....	76
6.2	Contribuições do Trabalho.....	76
6.3	Recomendações para Trabalhos Futuros.....	77
	REFERÊNCIAS	78

CAPÍTULO I

INTRODUÇÃO

1.1 Contextualização do tema

O tema desenvolvido no presente trabalho possui relacionamento direto com os avanços tecnológicos que vem ocorrendo na indústria de equipamentos eletroeletrônicos e a complexidade com relação à corrente harmônica gerada e injetada no sistema elétrico por estes, dos quais pode-se citar: microcomputadores, televisores digitais (PLASMA, LCD, LED), eletrodomésticos (Forno de Micro Ondas, Máquina de Lavar Louças) e Lâmpadas Fluorescente Compactas. Estes produtos proporcionam uma grande gama de benefícios e praticidades para todos os consumidores, porém com a propagação destes equipamentos dentro dos sistemas elétricos, tem ocasionado um acréscimo significativo das distorções harmônicas da corrente e tensão nas redes elétricas (TOSTES, 2003).

Por outro lado, os constantes acréscimos dos níveis de distorções harmônicas de tensão e corrente nas redes elétricas são prejudiciais aos próprios equipamentos eletrônicos ligados na rede, em decorrência destes possuem grande sensibilidade às distorções presentes no sistema.

Os sistemas elétricos de potência (SEP) deve conduzir a energia elétrica dos núcleos de produção para os locais de consumo, sem paradas e/ou interrupções dos serviços, com gastos mínimos, grau máximo de segurança, a fim de garantir que as pessoas e equipamentos estejam protegidos, com valores de tensão e frequência nos limites exigidos pela agência de regulação.

Com o constante crescimento da demanda e os desperdícios, torna-se necessário o aumento dos requisitos de eficiência e competitividade que hoje se observam nos sistemas de energia, a questão da qualidade da energia elétrica tornou-se um dos pontos de maior interesse econômico e tem ganhado cada vez mais importância, levando em conta que QEE envolve vários aspectos que podem melhorar a eficiência econômica, a competitividade e a qualidade apresentada no sistema.

A eletrônica de potência exerce, atualmente, nos sistemas elétricos uma grande influência na geração de fontes contaminantes ou produtoras de harmônicos destacando-se fundamentalmente cargas não lineares empregadas nas instalações industriais, residenciais e comerciais. Com isto, contribui-se para o aumento de problemas relacionados com a qualidade da energia elétrica fornecida aos consumidores.

Por outro lado, as cargas monofásicas não lineares e de baixa potência que possuem alta representatividade na geração de distorções harmônicas como: fontes de alimentação, notebooks, reatores de lâmpadas fluorescentes, monitores de vídeos, tablets e computadores de mesa, estão presentes em grande quantidade nos sistemas elétricos. A presença destas fontes causam grandes consequências na geração de distorções harmônicas que afetam significativamente as cargas conectadas ao Ponto de Acoplamento Comum (PAC) de um sistema elétrico de potência.

As medições dos conteúdos harmônicos vêm sendo usadas para caracterizar o comportamento de cargas não lineares, para localizar possíveis fontes harmônicas e quantificar os níveis de distorções harmônicas em sistemas elétricos. Mesmo em novos conceitos de distribuição de energia elétrica, há uma preocupação com a distorção harmônica total (THD) e outros indicadores de qualidade de energia. O THD é o índice harmônico mais comum usado para avaliar e medir as condições de variações da Qualidade da Energia em condições não senoidais.

No Brasil, a Agência Nacional de Energia Elétrica (ANEEL) desenvolveu os Procedimentos de Distribuição de Energia (PRODIST). A atual revisão do PRODIST entrou em vigência em primeiro de janeiro de 2010. Dentre os principais objetivos do PRODIST encontra-se o de garantir que os sistemas de distribuição operem com segurança, eficiência,

qualidade e confiabilidade. O Módulo Oito (8) do PRODIST regula a qualidade da energia em redes de distribuição.

A energia elétrica é o componente fundamental de desenvolvimento das sociedades mundiais, conseqüentemente sendo o principal artefato utilizado em todos os seguimentos da sociedade como: na produção de bens duráveis e não duráveis, nos lares, no comércio e nas indústrias, sendo também um dos fatores determinantes na concorrência e competitividade da economia.

Um sistema elétrico de potência é composto por vários componentes tais como: Centrais de geração, Linhas de transmissão e distribuição de energia, Eles são projetados de acordo com as especificações e normas estabelecidas pela ANEEL. Para operarem de forma a atender com segurança e qualidade as cargas neles conectadas.

O estudo deste trabalho analisou a Qualidade da Energia Elétrica - QEE, dentro de um ambiente fabril, com objetivo de detectar e diagnosticar a presença dos conteúdos harmônicos em sua rede elétrica, evitando assim penalizações futuras já que a normatização de QEE para níveis máximos de distorções harmônicas estão sendo elaboradas pela ANEEL. Este trabalho apresenta uma ferramenta para identificar dentro de uma indústria os processos que possam ser os maiores geradores de distorções harmônicas, quantificar os níveis de distorções apresentadas em cada processo, as harmônicas individuais mais significativas na distorção total de tensão - THD do ponto de acoplamento comum - PAC entre indústria e concessionária. Contribuindo assim, com o conhecimento das principais fontes de harmônicos dentro dos processos industriais e com isso possibilitando tomada de ações mitigadoras, contribuindo assim para a melhoria da QEE, presente no interior do ambiente pesquisa e no PAC.

1.2 Identificação e Justificativas da Proposta

A identificação, determinação, eliminação e ou a mitigação da geração do conteúdo harmônico dentro de uma rede elétrica, contribui diretamente na melhoria da QEE. A identificação e determinação servem para avaliar, a eliminação e a mitigação amenizam a degradação da QEE. Com o objetivo de adaptar-se à realidade atual do setor energético, onde problemas ambientais, desperdícios, custos na geração e operação do sistema de energia elétrica

dificultam os investimentos necessário ao setor, é necessário obter o máximo de eficiência com o mínimo de perdas no sistema, ou seja, a melhora da qualidade de energia do sistema deve ser buscada sempre. Uma alternativa de obter estes resultados é realizar o controle dos fluxos de harmônicas na rede.

1.3 Objetivos

1.3.1 Objetivo Geral

Desenvolver uma análise do conteúdo harmônico existente na rede elétrica de uma indústria do PIM, utilizando uma ferramenta de inteligência computacional com técnicas de mineração de dados.

1.3.2 Objetivos Específicos

- a) Expor um estudo sobre a qualidade de energia elétrica, baseada no THD que é um dos indicadores de qualidade de energia;
- b) Analisar os processos envolvidos na produção da indústria e apresentar o conteúdo harmônico presente em cada um dos processos, bem como, os maiores impactos de harmônicos individuais em cada processo;
- c) Criar uma ferramenta de auxílio para à tomada de decisão e controle do conteúdo harmônico gerado na empresa, que podem causar penalizações por geração de harmônicos acima dos limites estabelecidos pelas agências reguladoras, evitando assim, possíveis penalizações com multas.

1.4 Contribuições e Relevância do Tema

O presente trabalho justifica-se pela busca de melhorias da qualidade da energia elétrica (QEE), com foco no controle de distorções harmônicas num sistema de energia elétrica, localizado dentro do ambiente fabril, onde as principais contribuições geradas foram:

- a) Avaliação da Qualidade da Energia Elétrica (Harmônicos) apoiada em técnicas de Mineração de dados e Inteligência computacional;

- b) Permitir conhecer o perfil harmônico individual de cada processo envolvido na manufatura da indústria e suas contribuições no THD no PAC;
- c) Antecipação à legislação, podendo-se evitar assim futuras penalizações em decorrência da poluição da rede elétrica por fontes harmônicas. Conhecendo-se a geração de harmônicos por cada processo dentro da indústria fica mais fácil tomar ações mitigadoras;

O emprego de técnicas de inteligência computacional e Data Mining no estudo de distorções harmônicas propiciam análises mais detalhada das fontes de distorções harmônicas de corrente presentes nos processos de manufatura.

A utilização de técnicas de mineração de dados (árvore de decisão) é capaz de determinar a presença de poluição harmônica, quantificar e identificar quais os processos, horários ocorrências, turnos e fases que são causadores dos distúrbios harmônicos presentes em cada um dos processos analisados, bem como os níveis em cada uma das harmônicas e ainda informa a ordem harmônica presente em cada fase analisada.

1.5 Delimitação da Pesquisa

O estudo é delimitado na investigação das condições de QEE em uma rede de baixa tensão, para diagnosticar os impactos de harmônicos na rede elétrica de uma indústria do PIM, utilizando o KDD e a técnica de mineração de dados de árvore de decisão, com base nas normas presentes no módulo 8 Revisão 6 (PRODIST/ANEEL, 2015), a IEEE Std. 519 – 92 (IEEE, 2004) e IEC 61000-3-6 (IEC/TR, 2008-02).

Os efeitos causados por fenômenos envolvendo harmônicos em redes elétricas são pesquisados a bastante tempo, a pesquisa atual está limitada em uma análise realizada em uma empresa de eletroeletrônicos do PIM, onde foram monitorados os níveis de harmônicos gerados durante uma campanha de medição de uma semana, nos processos de teste de notebooks, montagem de tablets, bombas e compressores, centrais de ar-condicionados e transformadores da entrada da empresa.

1.6 Estrutura da Dissertação

A presente dissertação é composta por seis capítulos, assim estruturada:

O Capítulo 1 – Introdução sobre o tema, os objetivos gerais, a estrutura da dissertação apresentados no decorrer do desenvolvimento deste estudo.

O Capítulo 2 – Neste capítulo apresenta-se o estado da arte com relação ao tema da dissertação.

O Capítulo 3 – São abordados os aspectos que envolvem a Qualidade da Energia Elétrica, como: suas principais definições, a qualidade do produto energia elétrica, os fatores que alteram a qualidade do produto, harmônicos, metodologia de medição, instrumentação, os valores referenciais, entre outros.

O Capítulo 4 – São abordados os assuntos sobre Suporte à tomada de decisão, O processo de KDD, as fases do processo de descoberta de conhecimento em bases de dados, os paradigmas, árvore de decisão, os algoritmos utilizados em árvores de decisão, principalmente o algoritmo o C4.5 ou J48 que foi utilizado nas análises e os modelos de inteligência computacional.

O Capítulo 5 – São abordadas as análises da geração de harmônicos, as características da empresa analisada, os pontos analisados, a metodologia computacional utilizada, aplicação do processo do KDD.

O Capítulo 6 – São apresentadas as conclusões deste trabalho de dissertação bem como é realizada uma referência de como este trabalho pode ser futuramente melhorado e ampliado.

CAPÍTULO II

REVISÃO BIBLIOGRÁFICA

2.1 Estado da arte

As técnicas de mineração de dados são amplamente utilizadas em várias áreas do conhecimento. Dentre as várias aplicações, serão exibidas neste capítulo as aplicações que possuem correspondência entre as áreas de energia elétrica e de potência e inteligência computacional.

2.2 Trabalhos

Houben *et al.* (1995), As árvores de decisão são abordagens de aprendizagem automática bastante singular para avaliação da segurança do sistema de energia, em particular devido à sua interpretabilidade, a sua capacidade de identificar os principais parâmetros de condução e sua eficiência computacional. Neste trabalho é proposto uma técnicas híbridas para tentar combinar as vantagens de árvores de decisão com métodos vizinhos mais próximos, acoplando-os enquanto estiver usando algoritmos genéticos para melhorar ainda mais as suas performances. As abordagens derivados são, em seguida, aplicada a um caso de estudo no mundo real. Demonstrou-se que as abordagens híbridas são realmente superiores aos correspondentes aqueles "puros".

Hatziargyriou *et al.* (1995), Apresentam à aplicação da metodologia de árvore de decisão para a avaliação de um sistema de segurança dinâmico autônomo de um sistema de potência de tamanho médio, com uma grande penetração de fontes de energia renováveis. Os efeito de vários parâmetros é investigado e as perspectivas da integração deste regime no sistema de controle avançado da ilha grega de Lemnos são discutidos e analisados através da metodologia de árvore de decisão.

Karapidakis & Hatziargyriou (2002), desenvolveram um trabalho aplicando a técnica de árvores de decisão (DTS) para linha de segurança dinâmica preventiva de sistemas de energia isolados. As DTs fornecem a velocidade computacional necessária para um desempenho online e a flexibilidade de fornecer um controle preventivo. A ênfase é colocada sobre o uso on-line dos DTs para testar a segurança dinâmica de cada cenário despacho de geração e, assim, Fornecer aconselhamento corretiva via re-despacho de geração. Além disso, o algoritmo implementado fornece a flexibilidade de exibir o custo de cada reexpedição. Desta forma, o método pode ajudar na tomada de decisão objetiva. Os resultados da aplicação do sistema em série de carga real a partir da ilha de Creta, onde o sistema proposto está em operação experimental.

Sannino *et al.* (2003), no trabalho proposto à QEE é definida como Fenômenos de qualidade de energia que incluem todas as situações possíveis em que a forma de onda da tensão de alimentação (tensão de qualidade) ou corrente de carga (qualidade corrente) que se desviam da forma de onda senoidal à frequência nominal com uma amplitude correspondente ao valor rms classificado para todas as três fases. A ampla gama de distúrbios de qualidade de energia abrange diferenças de duração súbita, curta duração, por exemplo, transitórios impulsivos e oscilatórios, quedas de tensão (ou afundamentos), pequenas interrupções, bem como desvios de estado estacionário, como harmônicos e cintilação. Pode-se também distinguir, com base na causa, entre perturbações relacionadas com a qualidade da tensão de alimentação e os relacionados com a qualidade da corrente absorvida pela carga.

De acordo com Dugan *et al.* (2004), qualquer problema manifestado no desvio da corrente, tensão ou de frequência que resulte em falhas ou falhas de operação dos equipamentos dos clientes. Um utilitário pode definir a qualidade de energia com confiabilidade e demonstrando estatísticas de que o seu sistema é 99,98 por cento confiáveis. Os Critérios estabelecidos pela agência reguladora são geralmente nesse sentido. O fabricante de equipamentos de carga pode definir a qualidade de energia como as características do fornecimento de energia que permitem que o equipamento funcione adequadamente. Essas características podem ser muito diferentes para diferentes critérios.

No trabalho desenvolvido por Abdel-Galil *et al.* (2004), apresentam uma nova abordagem para a classificação dos distúrbios de qualidade de energia. A abordagem é baseada no aprendizado indutivo usando árvores de decisão. A transformada wavelet é utilizada para

produzir os vetores representativos de recursos que podem capturar com precisão as características únicas e marcantes de cada distúrbio. A árvore de decisão é obtida com base nos recursos produzidos pela análise wavelet através de inferência indutiva. Durante os testes, o sinal é reconhecido usando as regras extraídas da árvore de decisão. A precisão da classificação da árvore de decisão não só é comparável com a precisão da classificação de redes neurais artificiais, mas também contribui para a explicação da classificação das perturbações detectadas.

Teeuwsen *et al.* (2004), desenvolveram um novo método para a previsão de valores próprios de modos de estabilidade críticas de sistemas de energia baseados em árvores de decisão. Seu interesse Especial está focado em oscilações de sistemas interligados em grande escala inter-área. Os métodos existentes para o cálculo de valores próprios são demorados e requerem o modelo de sistema completo que inclui um grande número de estados. No entanto, a partir de árvores de decisões, a estabilidade oscilatória pode ser prevista com base em poucas entradas selecionadas. Por este meio, as saídas da árvore são atribuídos a taxa de amortecimento dos valores próprios críticos inter-área.

A expansão no uso de alta tecnologia, mas o uso de equipamentos sensíveis levou a um interesse considerável pela qualidade da energia. Além de confiabilidade, é um pré-requisito para o funcionamento contínuo de muitas instalações que utilizam equipamentos de alta tecnologia. Além disso, os usuários finais estão solicitando a obtenção de informações de qualidade de energia, a fim de obter um conhecimento aprofundado do desempenho de seus equipamentos com relação às variações na forma de onda da tensão. No entanto, o acompanhamento de execução em todos os nós da rede de energia com analisadores de QEE pode não ser justificada em termos de custo e complexidade. Devido a isso, tem havido um aumento significativo no interesse em utilizar modelagem computacional e simulações para prever e estimar a qualidade da energia em pontos não monitorados. Nos últimos anos, uma ampla gama de métodos de estimação de estado surgiu, e este documento fornece uma visão geral dos métodos de estimação de qualidade de energia estado comumente usados. Ele descreve as etapas envolvidas na estimativa e discute as várias considerações importantes, bem como os desafios enfrentados por esses empreendimentos (TAN *et al.*, 2005).

Myo & Milanovic (2006), desenvolveram um método abrangente para a previsão estocástica da quantidade e características dos afundamentos de tensão em redes de distribuição

de grandes dimensões. A nova abordagem proposta é em modelagem probabilística da falha do sistema de proteção. A probabilidade da falha do sistema de proteção é determinada utilizando um método simples, com base árvore de decisão. O estudo destaca os efeitos da falha do sistema de proteção sobre o número e as características dos afundamentos de tensão. O método proposto e as suas capacidades são demonstradas no sistema de distribuição genérica de tamanho real. Os resultados apresentados no trabalho mostram que o método proposto é superior ao uso das técnicas tradicionais utilizadas para a avaliação de quedas de tensão, uma vez que resulta em maior precisão da previsão da quantidade e as características de afundamentos.

Mukherjee & Wojszczyk (2007), descrevem a utilização de duas metodologias para encontrar as melhores soluções para os processos de orçamento e planejamento com a utilização de árvores de decisão. A primeira metodologia descreve os padrões para a construção de melhores perfis de carga para os geradores que podem ajudar um gerente de planta otimizar o orçamento dos custos de funcionamento de uma unidade geradora. A segunda metodologia demonstra um processo de triagem de alternativas de transmissão para identificar as melhores soluções com base no custo e segurança de rede. Árvores de Decisão pode ajudar a tomar boas decisões, mas não pode garantir que os bons resultados vão sempre ocorrer como resultado dessas decisões.

Kai *et al.* (2007), descrevem um esquema dinâmico on-line de avaliação de segurança para sistemas de energia interligada em larga escala utilizando medições fasoriais e árvores de decisão. O esquema constrói e atualiza periodicamente as árvores de decisão off-line para decidir atributos críticos como indicadores de segurança. As árvores de decisão fornecem avaliações de segurança on-line e diretrizes de controle de prevenção com base em medições em tempo real dos indicadores de unidades de medição fasorial. O esquema usa um novo método de classificação envolvendo cada caminho inteiro de uma árvore de decisão, em vez de apenas os resultados de classificação em nós terminais para fornecer resultados de avaliação de segurança mais confiáveis para mudanças nas condições do sistema. As abordagens desenvolvidas são testados com 2100 barras, 2600 linhas, 240 gerador modelo operacional do sistema Entergy.

Voumvoulakis & Hatziargyriou (2008), propõe um modelo de aprendizagem de máquina para a avaliação de corte de carga para a segurança dinâmica corretiva do sistema. O método proposto emprega um mapa auto-organizada com árvores de decisão aninhados em alguns de seus nós, a fim de classificar os perfis de carga de um sistema de energia. O método é aplicado em um

modelo realista do sistema de energia Helênico e o seu valor acrescentado é mostrado, comparando os resultados com os obtidos a partir da aplicação de simples mapas auto-organizados e árvores de decisão simples.

Mori *et al.* (2008), apresentam uma proposta utilizando a técnica de mineração de dados para o controle de tensão da rede de distribuição. O método proposto extrai regras eficientes do controle de tensão das três fases minimizando as perdas da rede de distribuição. Do ponto de vista de qualidade de energia, a formulação matemática do controle de tensão da rede de distribuição é difícil de resolução devido à complexidade. Nesse sentido, os operadores de distribuição não entendem o problema. Este trabalho apresenta um método baseado em árvore de regressão (decisão) que esclarece a relação não linear entre variáveis de controle e a perda de rede em uma rede de distribuição. Uma árvore de regressão é construída para descobrir as regras de uma regulação de tensão da rede de distribuição. O método proposto foi aplicado para o sistema IEEE de 34 nós.

No Trabalho proposto por Gunther (2009), apresenta um software de gerenciamento de banco de qualidade de energia elétrica, equipado com análises estatísticas e ferramentas de plotagem possuindo uma interface amigável. Ele consiste de dois componentes principais, a *Power Quality Data Manager* e o *Power Quality Data Analyzer*. O trabalho oferece aos usuários a capacidade de processar e analisar diversos tipos de dados, integrando todas as informações a partir de inúmeros dispositivos de monitoramento em um único banco de dados relacional. Fornece ainda os meios para automatizar o carregamento de novos dados usando o módulo *Data Quality Power Manager*, bem como fornece os meios para criar relatórios automatizados e personalizados através do Módulo *Data Analyzer Power Quality*.

Xiaofeng *et al.* (2009), propõem uma nova abordagem baseada na árvore de decisão (C4.5) para a detecção de ilhamento em geração distribuída de um sistema elétrico (DG). Sem qualquer efeito negativo sobre a qualidade da energia, este novo método reduz consideravelmente os danos causados para resultante do ilhamento, e também aumenta a capacidade de detecção de ilhas do relê de proteção. A construção da árvore de decisão é baseada em dados da operação passadas de um sistema de geração distribuída real. Este método foi testado em um sistema típico de distribuição com vários recursos distribuídos por meio de ferramentas MATLAB / SIMULINK. Os resultados das simulações mostram que a árvore de decisão C4.5 é eficaz e o modo de funcionamento de ilhamento de DG podem ser totalmente previsto por este novo algoritmo.

Zhiyong & Weilin (2009), desenvolveram um estudo de avaliação da segurança do sistema de energia Chinês, utilizando árvore de decisão na avaliação online de um modelo realista da rede elétrica utilizando uma base de dados que abrangem todas as possíveis condições de operação. Além disso, as unidades fasoriais (PMU) são levados em consideração para uma melhor tomada de decisão. Os resultados demonstram que o esquema de mineração de dados proposto é capaz de identificar os indicadores de segurança cruciais e dá previsões de segurança confiáveis. Além disso, a diferença de ângulo de fase da tensão que o obtido pelo PMU é testado para melhorar a precisão de identificação.

Neste trabalho Yen-Tsung & Men-Shen (2009), Propõem uma arquitetura BDI (crença-desejo-intenção) à base de arquitetura multi-agente é proposto. Os agentes podem dar resposta às mudanças do ambiente e tomar decisões com outros agentes para atingir o objetivo final, a prestação de serviços a todos os clientes de tensão. Esses agentes atuam de forma independente, mas de forma cooperativa. Árvore de decisão é usada na arquitetura de agentes com a proposta de fazer planos de restauração. As árvores de decisão são construídas durante o processo de restauração com base na localização de falhas e condições de carga através de comunicações entre os agentes. Utilizando a árvore de decisão, uma decisão mais robusta pode ser alcançado por agentes.

O trabalho desenvolvido por Nantian *et al.* (2009), apresenta uma nova abordagem para o reconhecimento de sinais não estacionários em distúrbios de qualidade de energia, As perturbações identificadas são afundamentos, elevações, interrupções, transitórios e harmônicos. O novo modelo inclui principalmente duas etapas. Em primeiro lugar, S-transformação é usado para analisar sinais de perturbação do sistema de potência, e a segunda são extraídas as duas características mais relevantes. Com base nestas duas características, são agrupados dois vetores de características 2D usando hierarquia fuzzy C-means (FCM). Em segundo lugar, uma árvore de decisão binária é construída a partir dos centros dos clusters FCM de acordo com os padrões das perturbação reconhecidas automaticamente.

Em um mercado de eletricidade liberado, não é de estranhar que os diferentes clientes necessitam de diferentes níveis de qualidade de energia com preço diferente. Então, como conseguir a avaliação de classificação sintética de qualidade de energia, torna-se um grande problema. Neste trabalho, a ideia de avaliação de classificados de qualidade de energia é

ênfatisada, e são analisadas as principais diferenças entre os problemas contínuos e os baseados em eventos de qualidade de energia, tanto nas características e escalas de tempo. Além disso, mínimos quadrados apoiar teoria de máquina de vetor (LS-SVM) é aplicado à avaliação global da qualidade de energia contínua. Enquanto isso, o índice de confiabilidade de fornecimento de energia é introduzida para avaliar o nível de qualidade global da qualidade de energia com base em eventos. Nesta base, o método de ponderação subjetiva é utilizado para alcançar a avaliação abrangente de dois tipos de problemas de qualidade de energia. A avaliação abrangente da qualidade da energia com base na classificação significativa não só destaca os dois tipos de problemas de qualidade de energia, mas também atinge o nível global de qualidade de energia (DING *et al.*, 2010).

No trabalho realizado por Genc *et al.* (2010), propõem um método de controle preventivo e corretivo na melhoria da segurança dinâmica de um sistema de energia. O método avalia os controles corretivos e preventivos como agendamento de geração e o esquema de cortes de cargas, os estudos são desenvolvidos com base nas regiões de segurança e limites que são calculados no espaço de variáveis de decisão mais adequados. As regiões e os limites são determinados pelas regras de árvores de decisão que são desenvolvidas pelo conhecimento gerado em bases de dados. Este trabalho envolve também a melhoria da precisão dos limites de segurança.

Cornoiu *et al.* (2011), definem qualidade de energia elétrica como o conjunto de regulamentos de propriedades elétricas que permite que os sistemas elétricos possam funcionar sem perda significativa de desempenho ou envelhecimento. O termo é usado para descrever a energia elétrica que aciona uma carga elétrica e capacidade do sistema para operar corretamente com a energia elétrica. Sem a alimentação adequada, um aparelho elétrico (ou carga) pode funcionar mal, falhar prematuramente ou não funcionar adequadamente. O objetivo do trabalho é analisar os diferentes impactos em diferentes consumidores, que atuam no mercado aberto, No presente trabalho foi realizado um estudo de caso real dentro de uma subestação do sistema elétrico Romeno. O processo de monitoramento da qualidade da energia foi realizado na respectiva subestação e nas ligações com o resto do sistema de energia.

No estudo de Vaid *et al.* (2011), foi realizado a identificação automática da origem de fontes harmônicas, Através da injeção de correntes harmônicas utilizando um diodo retificador trifásico. Utilizando o método de impedância crítica para detecção de distúrbios causadores de harmônicas, no experimento foi analisado um sistema de cinco barras entre duas empresas

geradoras (GENCO), empregou-se o método de árvore de decisão para identificar automaticamente a fonte harmônica.

O trabalho desenvolvido por Júnior (2011), apresenta um estudo sobre qualidade de energia elétrica (QEE), onde se avaliou a rede elétrica através do monitoramento das distorções harmônicas e fator de potência, com uso de lógica fuzzy, com objetivo de tentar encontrar a melhor solução da QEE aplicando as normas nacionais e internacionais.

Santos & Oliveira (2011), expuseram um estudo que apresentou uma avaliação de desempenho, experimental em laboratório da metodologia fundamentada no método da superposição, para a atribuição das responsabilidades quanto à geração das distorções harmônicas no ponto de acoplamento comum entre a rede da concessionária e os consumidores.

Na proposta de Swearingen (2011), foi apresentado um método para avaliar a rede elétrica totalmente automatizada em tempo real e encontrar potenciais áreas problemáticas ou pontos fracos na rede elétrica, utilizando os princípios da teoria dos jogos de árvores de decisão e "O Dilema do Prisioneiro" para avaliar as melhores ações possíveis de implementar dentro da rede elétrica para evitar possíveis problemas antes que eles aconteçam.

Diao *et al.* (2011), apresentaram um trabalho com um novo método para mitigar as oscilações de um sistema interligado de energia em grande escala inter-área, por meio de geração de reenvio. Os procedimentos ótimos de atenuação são derivados através de pesquisa para a distância mais curta a partir de condições de operação atuais para uma condição operacional de destino com a relação de amortecimento desejado do modo de oscilação. Um método baseado na sensibilidade é usado para selecionar os geradores mais eficientes para a geração de reenvio e uma árvore de decisão é treinada para aproximar o limite de segurança em um espaço caracterizado pelos geradores selecionados. As regras operacionais ideais podem ser encontrados através da resolução de um problema de otimização, onde as restrições de limite são fornecidas pelas regras de árvore de decisão. Este método foi testado em um Conselho de Coordenação de Eletricidade Ocidental (WECC) com um modelo de 179 barras, a simulação dos resultados demonstraram a validade do método baseado em árvore de decisão e mostrou promissora aplicação em operação em tempo real.

Com o desenvolvimento incessantemente do sistema de energia, a proporção de cargas

não-lineares está aumentando rapidamente e fontes harmônicas tornam-se cada vez mais presentes. Elas possuem grandes influências sobre a qualidade da energia. A fim de proteger a segurança da operação de redes de energia e uso seguro da energia elétrica, vários níveis de redes de qualidade de energia são gradualmente formadas, que fornece uma base sólida para a análise e aplicação de dados de qualidade de energia. O trabalho proposto estuda de forma concreta as informações de monitoramento de qualidade de energia, a transmissão de informações e leitura, e faz com que a análise de dados, por outro lado, apresente uma suposição aplicação. Os dados de qualidade de energia e os seus resultados de análise têm um papel importante para as massas de profissionais de energia elétrica e consumidores para garantir o sistema de energia em um estado seguro, confiável e economicamente viável (XIAN-KANG *et al.*, 2011).

Dehkordi *et. al* (2012), apresentaram uma nova abordagem para o diagnóstico de falhas do sistema de energia com base em medições fasoriais sincronizadas durante a falha. As medições sincronizadas são obtidas em tempo real a partir de Medição de Unidades Fasoriais e comparados com os limites off-line determinados por árvores de decisão para diagnosticar a falha. As árvores de decisões já foram treinadas off-line por meio de análise detalhada do sistema de energia para diferentes casos de falha. Enquanto os métodos tradicionais de diagnóstico de falhas usar o status de relés de proteção e disjuntores para inferir a seção de falha no sistema de energia, o método proposto utiliza os sinais disponíveis após a falha e, portanto, pode ser confiável, mesmo em caso de falha do sistema de proteção. Portanto, o diagnóstico de falhas pode ser alcançado com sucesso especialmente em situações de blackout, onde muitos disjuntores e e/ou mau funcionamento dos relés de proteção.

No trabalho proposto por Ozgonenel *et al.* (2012), foi desenvolvido um diagnóstico de falhas em sistema de energia, com principal objetivo de desenvolver um algoritmo de classificação de sinal para diferentes tipos de distúrbios de qualidade de energia(falhas), tendo como base nas mais recentes melhorias nas técnicas de processamento de sinais e reconhecimento de padrões dos sinais analisados.

Ohana *et. al* (2012), apresentaram um estudo sobre algumas experiências realizada com o uso de mineração de dados e árvore de decisão, aplicados a um sistemas elétricos de potência, com o objetivo de extrair conhecimento a partir de dados históricos armazenados em bancos de dados que foram produzidos pelo sistema de uma usina hidrelétrica no Brasil. Nas análises

foram utilizadas técnicas estatísticas em eventos discretos, tais como eventos booleana, alarmes, comandos set-points e quantidades analógicas como frequência elétrica, para apresentar os aspectos relevantes da operação do sistema elétrico. Os resultados deste estudo contribuíram para demonstrar com sucesso a aplicabilidade de mineração de dados para os sistemas de energia, melhorando a gestão das manutenções de usinas hidroelétricas.

Hosseini *et al.* (2012), apresenta um novo método de identificação on-line da área de controle de tensão crítica (CVCA) com base na árvore de decisão. Instabilidade de tensão é uma das principais causas de apagões em sistemas de energia modernos. O objetivo do trabalho é propor uma metodologia on-line para identificação de CVCA, que é função chave para medidas preventivas e corretivas contra instabilidade. O estudo foi realizado em várias condições de operação do sistema e de contingência, as árvores de decisão foram treinadas off-line gerando as medições PMU on-line.

Análise da estabilidade do sistema de energia tradicional com base em modelo computacional completo mostra suas desvantagens em aplicações em tempo real onde as variações rápidas estão presentes em ambos os lados da demanda e da oferta. Este estudo apresenta o uso de Árvores de Decisão para avaliação rápida do sistema de energia e estabilidade oscilatória e estabilidade de tensão com base em medições de tensão e fasoriais. Um ponto de funcionamento são agrupados em uma de várias categorias de estabilidade com base no valor de indicador de estabilidade correspondente. Uma nova metodologia para criação de base de conhecimento foi elaborado para garantir a conjuntos de dados de treinamento prático e suficientes. Os resultados obtidos através do exame de desempenho usando a base de conhecimento gerado são encorajadores. O impacto da DT método de cultivo e configuração do nó na precisão da classificação tem sido explorado. Por fim, as diferenças de desempenho entre a árvore de regressão e várias outras ferramentas de mineração de dados foram comparados (ZHENG *et al.* 2012).

Rodriguez *et al.* (2013), apresentam um trabalho de classificação automática com base em S-transform Como ferramenta de extração de características e árvore de decisão Como algoritmo classificador. Os sinais gerados de acordo com os modelos matemáticos, incluindo distúrbios complexos, têm sido utilizados para conceber e testar esta abordagem, em que o ruído é adicionada aos sinais de 40dB a 20dB.

Na proposta de Tingyan & Milanovic (2013), investiga-se as questões práticas associadas à previsão on-line de estabilidade transitória utilizando o método de árvore de decisão. As questões da qualidade e disponibilidade dos sinais de medição fornecidos por todo o sistema de medição de área (WAMS) são discutidas e seus efeitos sobre a precisão do desempenho da árvore de decisão são avaliados. O método split substituto incluído na árvore de classificação e regressão (CART) algoritmo é utilizado para lidar com a indisponibilidade de sinais de medição, e de ruído presente nos dados on-line é modelado como ruído Gaussiano branco (WGN) com vários sinal-ruído (SNR).

Biswal & Dash (2013), propuseram um novo esquema para a medição, identificação e classificação dos diversos tipos de distúrbios de qualidade de energia. O método proposto utiliza uma variante rápida do algoritmo S-Transform para a extração de características relevantes, que são usadas para distinguir entre os diferentes eventos de qualidade de energia, através de um classificador baseado em árvore de decisão difusa.

Tongyou *et al.* (2013), apresentam um algoritmo usando a técnica de mineração de dados, em especial a modelagem mistura com base no tamanho da mensagem mínima método (MML), para classificar os dados harmônicas em clusters e identificar padrões úteis dentro dos dados. Os clusters que resultaram são aplicados para distinguir as fontes de distúrbios e o calendário das perturbações na rede de distribuição. Além disso, o algoritmo C5.0 foi utilizado com aprendizagem supervisionada para produzir as regras de como os dados medidos são classificadas em vários grupos, usando uma técnica de árvore de decisão. Estas regras geradas podem então ser utilizadas para prever quais aglomerar quaisquer novos dados pertence sem calcular novamente.

Kar & Samantaray (2013), O trabalho apresenta um esquema de detecção de ilhamento passivo em micro redes com geração sincronizada e de inversores, usando árvore de decisão, o esquema envolve a extração de características dos diferentes pontos de localização do ilhamento e constrói a árvore de decisão para a proteção contra o ilhamentos do ponto específico. Após a detecção de cada ilhamento nos vários pontos, então é construída uma árvore de decisão global eficaz para ambas as gerações, desse modo, poderá ser desenvolvido um esquema de anti-ilhamento.

Neste estudo é proposta uma nova abordagem para a predição da dinâmica rápida de

ocorrências de instabilidades de tensão (como um fenômeno de curto prazo e / ou um longo prazo) e a estabilidade da tensão do sistema. A principal contribuição é na introdução de um procedimento para gerar novos recursos a serem aplicados a um classificador padrão, pelo qual o status de estabilidade de tensão dinâmica de um sistema de energia pode ser previsto. O procedimento de geração de recurso proposto só precisa de variáveis pré medidas dos níveis de perturbações e severidades apresentados por unidades de medição fasorial. Como entradas ao passo que um conjunto de variáveis de saída são derivados de um programa de fluxo de potência sem restrições. Como o método proposto não precisa de qualquer perturbação pós dados medidos, a tarefa de predição pode ser realizada logo após a perturbação (KHOSHKHOO *et al*, 2013).

O trabalho de pesquisa apresentado por Rather *et al.* (2013), analisa o impacto da energia eólica, na fase de saída das centrais das usinas e na troca de potência. Árvore de decisão é usada para avaliar a segurança dinâmica no presente e no futuro do Sistema de Alimentação dinamarquês. Os resultados de simulação off-line no domínio do tempo para o grande número de possíveis condições operacionais (OC) e contingências críticas são organizados para criar o banco de dados, que é então usado para prever a segurança do presente e no futuro do sistema de energia.

Beiraghi & Ranjbar (2013), apresentaram um novo método de avaliação de segurança de tensão on-line com base em medições de área ampla e algoritmo de árvore de decisão. Para os cenários de carga e geração de variação previstos (isto é, um dia antes), o banco de dados é obtido usando o fluxo de potência continuado. Além disso, cada ponto de operação é rotulado como "seguro" ou "inseguro" de pontos de vista com base em critérios de segurança de tensão WECC estabilidade de tensão. As árvores de decisão são treinadas sobre o subconjunto de dados existentes através da aplicação de duas regras de decomposição famosos e vários preditores.

De acordo com o trabalho proposto por Govindarajan & Balakrishna (2013), a qualidade da energia elétrica concentra-se sobre os vários aspectos, onde são descritos os recentes avanços na qualidade do sistema de potência e análise do mesmo utilizando software de monitoramento. Com o aumento da adição de cargas não-lineares para sistemas elétricos, tornou-se necessário assegurar que a degradação de tensão não ocorre na fase de transmissão ou de distribuição. Vários problemas práticos estão associados com qualidade da energia como

distorções harmônicas, afundamentos, sobre tensões e transitórios são as principais razões para geração de sinais elétricos de baixa qualidade. Os sistemas de monitoramentos avançados estão agora utilizando estes distúrbios e variações para categorizar. Este trabalho destaca os dados recuperados a partir de tais sistemas de controle, bem como os métodos de análise de vários problemas associados com a qualidade da energia. O trabalho proposto também proporciona medidas corretivas e adequadas que podem ser aplicadas para melhorar a qualidade da energia.

Segundo Leão *et al.* (2014, p.1) a qualidade da energia elétrica (QEE) é a condição do sinal elétrico de tensão e corrente que permite que equipamentos, processos, instalações e sistemas elétricos operem de forma satisfatória, sem prejuízo de desempenho e de vida útil. A expressão é usada para descrever a energia elétrica que aciona uma carga elétrica, fazendo-a operar corretamente. Sem uma alimentação adequada, com padrão de qualidade, a carga pode operar inadequada ou incorretamente, falhar prematuramente ou simplesmente não funcionar.

Chengxi *et al* (2014), propõem uma árvore de decisão com base abordagem sistemática para a avaliação cooperativa sistema de energia on-line segurança dinâmica (DSA) e controle preventivo. Esta abordagem adota uma nova metodologia que treina duas árvores de decisões orientada a contingência geradas diariamente pelos bancos de dados, a partir de simulações do sistema de potência em tempo real com as medições de área ampla, uma árvore de decisão de variáveis mensuráveis é empregado para linha DSA para identificar possíveis problemas de segurança e o outra árvore de decisão de variáveis controláveis fornece suporte a decisões on-line sobre as estratégias de controle preventivo contra essas questões. Um algoritmo de custo-benefício é adotado em sua abordagem proposta para otimizar a trajetória de controle preventivo. O documento propõe um algoritmo de amostragem importância na preparação do banco de dados para o treinamento árvore de decisão.

No trabalho de Kazerooni *et al.* (2014), foi desenvolvido um estudo sobre as aplicações de mineração de dados em sistemas de energia, com uma rede altamente interconectada que fornece energia elétrica para os consumidores. Sustentando a entrega segura e confiável de energia elétrica que requer monitoramento contínuo do sistema e que processe grandes volumes de dados obtidos dos dispositivos de medição. Faz-se necessário investigar se as aplicações de mineração de dados são abrangentes e confiáveis na investigação e monitoramento de sistemas de energia.

Zapata-Tapasco *et al.* (2014), apresenta uma estratégia baseada na aprendizagem que utiliza árvores de decisão para a localização de falhas em sistemas de energia radiais, que visa melhorar a qualidade da energia elétrica, como exigido pelos mercados de energia elétricas desregulamentados. O método proposto subdivide primeiro o sistema de energia em várias regiões, e, em seguida, uma técnica de classificação baseado em árvores de decisão é treinada usando um banco de dados de falha. As árvores de decisão obtidas são usadas para atribuir uma zona de falha para um novo evento de falha, o que reduz o tempo de recuperação e, como consequência ajuda a manter os índices de boa qualidade.

No estudo de Ray *et al.* (2014), foi abordada a influência sofrida pela penetração dos sistemas destruídos nos sistemas convencionais de energia que causam distúrbios de qualidade de energia. Desenvolveram uma classificação melhor dos distúrbios de qualidade de energia elétrica que está associada a vários fatores como as mudanças de cargas e fatores ambientais. Várias formas de distúrbios são levados em consideração incluindo afundamentos (sag), sobre-tensão (swell), cortes de tensão (notch) e harmônicos, foram utilizadas várias técnicas como algoritmos genéticos, Máquina de vetorial de suporte e Árvore de decisão, Para apoiar na classificação de distúrbios. O estudo foi apoiado por três diferentes estudos de caso, um protótipo de montagem experimental a energia elétrica utilizando energia eólica, um sistema fotovoltaico e um sistema Nórdico de 32 barras.

O trabalho desenvolvido por Garg *et al.* (2014), trata-se de uma técnica modificada para o reconhecimento de distúrbios de qualidade de energia de fase única e múltiplas. Utilizando um algoritmo baseado em ST (Stockwell's-Transform), o classificador RNA e uma árvore de decisão. Uma análise e classificação dos distúrbios de fase única, composta por ambos os eventos e variações como sag, ondulação, interrupção, harmônicas, transitórios, notch e flicker são apresentados. Além disso, o algoritmo proposto também é aplicado em vários distúrbios de qualidade de energia elétrica como harmônicos, sag, ondulação, flicker e interrupção. A base de dados destes distúrbios com base no padrão IEEE-1159 é gerado no MATLAB para estudos de simulação. O algoritmo proposto extrai as características significativas de vários distúrbios, usando S-transform como uma entrada para este classificador híbrido para a classificação dos distúrbios.

Swarup (2015), apresenta uma estudo sobre a avaliação da segurança de um sistema elétrico utilizando árvore de decisão para medir e prever a robustez do sistema em decorrência

de distúrbios graves. A tomada de decisão é realizada com a construção da árvore de decisão em off-line, os dados utilizado na construção das árvores e tomada de decisão são oriundos do sistema de controle e simulação tomada de ações de controles adequados para as ocorrências de distúrbios graves. As árvores de decisões são construídas off-line com simulações geradas de grandes bases de dados. Neste trabalho árvores de decisão são utilizadas para classificar os estados operacionais dos sistemas de energia como seguro / inseguro utilizando os fluxos MW e tensões e o nível de gravidade da infração e as ações de controle necessário por encontrar a localização do ponto de operação.

O trabalho desenvolvido por Moutis & Hatziargyriou (2015), incorpora a geração distribuída ao paradigma da central energética virtual (VPP - Virtual Power Plant) que permite um comando central mais coerente e a integração coordenada de diversas fontes de energia elétrica. Dessa forma, VPPs podem participar do controle de frequência, regulando de forma coordenada a sua potência de saída, melhorando a estabilidade do sistema. Este estudo utilizou a metodologia de árvore de decisão para apoiar na mitigação de sobre-frequência do despacho de potência ativa dentro de um VPP.

2.3 Considerações Finais

Neste capítulo, foram apresentados vários trabalhos que possuem correlação com as áreas de QEE, energia elétrica, eletrônica de potência, inteligência computacional e mineração de dados. Demonstrando assim a maleabilidade da aplicação das técnicas de inteligência computacional, mineração de dados e KDD nas mais diversas áreas do conhecimento com a mesma eficácia. Os trabalhos apresentados no presente capítulo utilizam uma gama de técnicas presentes no KDD e inteligência computacional e serão utilizados como base para o desenvolvimento dos demais capítulos.

CAPÍTULO III

QUALIDADE DA ENERGIA ELÉTRICA

3.1 Introdução

O termo qualidade, de forma geral, é um atributo ou grupo de atributos que tem como objetivo distinguir uma coisa de outra em relação ao seu grau de excelência, o item qualidade que tem sido almejado por todos os setores. Este conceito é aplicado no produto energia elétrica com o objetivo de qualificar ou certificar que o produto atende os padrões definidos pela agência reguladora e as necessidades dos consumidores.

As concessionárias de energia elétrica e usuários finais de energia elétrica estão se tornando cada vez mais preocupados com a qualidade da energia elétrica. O termo qualidade de energia tornou-se um dos chavões mais prolíficos da indústria de energia desde o final da década de 1980. É um conceito genérico para uma infinidade de diversos tipos de distúrbios do sistema de potência. As questões que recaem sob esse assunto não são necessariamente novos. O que é novo, é que os engenheiros estão agora tentando lidar com estas questões, utilizando uma abordagem do sistema, em vez de lidar com eles como problemas individuais (DUGAN *et al.*, 2004).

O monitoramento da qualidade da energia é um serviço essencial com muitas utilidades para os clientes comerciais e industriais entre outros. Devido à tecnologia e software que já está disponível, esse monitoramento é altamente eficaz. O sistema de monitoramento não só pode fornecer informações de monitoramento mais também fornece informações sobre a qualidade da alimentação e as causas das perturbações do sistema de energia, mas pode identificar condições de problemas em todo o sistema antes que eles causem queixas generalizadas de clientes, mau funcionamento do equipamento, e até mesmo falhas de equipamento. Problemas de qualidade de energia não são necessariamente limitados ao sistema de energia elétrica. Muitas pesquisas têm mostrado que a maioria dos problemas são localizados dentro das instalações dos clientes.

Dado este fato, o monitoramento é uma oportunidade chave para as concessionárias protegerem sua reputação e melhorar o seu relacionamento com os clientes. Os requisitos essenciais para um sistema de monitoramento de sucesso incluem extensas capacidades de processamento de dados, relatórios de fácil compreensão e compartilhamento universal de informações.

3.2 Conceituação

Não existe uma concordância sobre um conceito abrangente para o termo “Qualidade da Energia Elétrica” ou QEE, nos diversos trabalhos produzidos pela comunidade acadêmica, temos uma variedade muito grande de conceitos, normalmente estes conceitos estão ligados aos parâmetros como: as distorções harmônicas, as flutuações de tensão, as variações de tensão de curta duração, os desequilíbrios de sistemas trifásicos e aos transitórios de curta duração.

Na atualidade existe uma definição que é aceita por grande parte da comunidade científica que é: Energia elétrica de boa qualidade é aquela que garante um funcionamento contínuo, adequado e seguro dos equipamentos elétricos e os processos associados, sem afetar o meio ambiente e o bem estar das pessoas.

3.3 Normas de Qualidade de Energia

A nível mundial existem várias normas e recomendação de diversas agências e organizações que tratam sobre os limites de distúrbios harmônicos nos sistemas de energia. As normas e recomendações são desenvolvidas por diversas entidades como os fabricantes de equipamentos, as concessionárias e institutos de pesquisas e desenvolvimento.

As normas de QEE são composta de um conjunto de princípios e regras que definem os limites de qualidade que o sistema elétrico deve possuir para oferecer uma boa qualidade, nos serviços elétricos prestados aos consumidores.

As concessionárias, os consumidores e o mercado de energia, normalmente são orientados por uma norma que possa definir os limites de emissões de distúrbios aceitáveis nas redes elétricas, No Brasil possuímos a ABNT em conjunto com a ANEEL que são as agências

responsáveis por normatizar o padronizar o mercado Brasileiro, Na Europa existem a IEC, C4, CEN, nos Estados Unidos existem IEEE, NEMA, NEC, ANSI.

De acordo com Leite (2013), existem normas nacionais e internacionais para o controle da distorção das formas de onda de tensão e corrente. Entre as normas internacionais existentes destacam-se: a IEEE Std. 519 – 92 (IEEE, 2004) e a IEC 61000-3-6 (IEC/TR, 2008-02). No Brasil, estão vigentes os Procedimentos de Distribuição de Energia Elétrica do Sistema Elétrico Nacional – PRODIST – Módulo 8 Revisão 6 (ANEEL-PRODIST, 2015).

3.3.1 A norma IEC 61000-3-6 (IEC/TR,2008-02)

A norma IEC 61000 (Comissão Eletrotécnica Internacional) abrange um amplo conjunto de regulações sobre a qualidade de energia. Especificamente, a IEC 61000-3-6 (IEC/TR, 61000-3-6, 2008) estipula os limites de emissão de harmônicos para as instalações conectadas a sistemas elétricos de potência que será utilizada como um das questões básicas a ser analisada.

De acordo com norma IEC 61000-3-6 (IEC/TR, 61000-3-6,2008), é possível calcular as distorções harmônicas total de tensão, utilizando a equação 1. Podemos ainda definir.

- a) Distorção harmônica total de tensão (%THDV) é a raiz quadrada do somatório quadrático das tensões harmônicas de ordens 2 a 50 (NT 0029/2011-SRD/ANEEL). Esse conceito procura quantificar o conteúdo harmônico total existente em um determinado barramento da Rede Básica ou nos barramentos dos transformadores de fronteira, sendo calculada de acordo com equação 1.

$$\%THDV = \frac{\sqrt{\sum_{h=2}^{\infty} V_h^2}}{V_1} 100 \quad (1)$$

Onde:

THDV = Distorção harmônica total de tensão;

V_1 = harmônica de tensão fundamental;

V_h = harmônica de tensão de ordem n

h = harmônica de tensão inicial de ordem 2

- b) Distorção harmônica individual de tensão (%IHDV), Grandeza que expressa o nível individual de uma das componentes que totalizam o espectro de frequências de um sinal distorcido, normalmente referenciada ao valor da componente fundamental. Calculada de acordo com a equação 2.

$$\%IHVD_h = \frac{V_h}{V_1} 100 \quad (2)$$

Onde:

ITHD = Distorção harmônica de tensão individual;

V_1 = harmônica de tensão fundamental;

V_h = harmônica de tensão de ordem n

3.3.2 A norma Brasileira ANEEL-PRODIST Módulo 8 Revisão 6 (2015)

De acordo com o PRODIST elaborado pela ANEEL, A Qualidade da Energia Elétrica possui três aspectos importantes nos padrões de qualidade, estes padrões descrevem os aspectos relevantes dos padrões de Qualidade do Produto, Qualidade do Serviço e Disposições Transitórias, os mesmos estão normatizados e padronizados no Módulo 8 Revisão 6.

No presente trabalho será tratado somente o padrão de qualidade do produto, por ser o principal foco do trabalho.

3.3.2.1 Qualidade do Produto

O conceito sobre qualidade do produto em QEE trata da conformidade do produto “energia elétrica”, que pode ser descrita como a capacidade que o sistema elétrico possui de fornecer energia dentro dos limites aceitáveis de tensões equilibradas e sem deformações na sua forma de onda. Ou seja, o SEP deve possuir disponibilidade de energia elétrica com limites tensões senoidais, equilibradas e com amplitude e frequência constantes sem variações o que quase impossível dada pelas características dos sistemas elétricos que estão sempre sendo afetados por inúmeros fatores que alteram a qualidade do produto.

A definição de qualidade de energia ou, mais especificamente, uma perturbação de qualidade de energia, é geralmente aceito como qualquer mudança na qualidade da energia

(tensão, corrente, ou frequência) que interfere com o funcionamento normal dos equipamentos elétricos. É a susceptibilidade dos equipamentos de uso final que define o nível necessário de qualidade de energia (DOUGHERTY; STEBBINS, 1997).

De acordo com o PRODIST, Módulo 8, estabelecidos pela ANEEL e nos procedimentos de rede definidos pelo Operador Nacional do Sistema Elétrico (ONS), foram definidos que os aspectos mais relevantes em relação à Qualidade do Produto em regime permanente e transitório são:

- a) Tensão em regime permanente;
- b) Fator de potência;
- c) Harmônicos;
- d) Desequilíbrio de tensão;
- e) Flutuação de tensão;
- f) Variações de tensão de curta duração;
- g) Variação de frequência

Mesmo com a existência de vários outros aspectos relevantes na Qualidade do Produto definido e padronizados através do PRODIST Módulo 8 Revisão seis (6) 2015, o aspecto que será evidenciado com maior relevância na atual pesquisa, serão as distorções harmônicas que servirão como as principais variáveis de análise na tomada de decisão.

3.3.3 *Distorções Harmônicas*

As distorções harmônicas são deformações apresentada na forma de onda, onde sua forma não é mais senoidal, ou seja, são ondas múltiplas da onda senoidal de 60hz (Padrão Brasileiro). Quando uma tensão deixa de ser puramente senoidal ocorrem mudanças de comportamento nas instalações elétricas e nos equipamentos. Quanto mais distorcida for a forma de onda, mais acentuados serão os efeitos negativos sobre alguns tipos de equipamentos. Normalmente o nível de distorção harmônica é medido em uma relação percentual, através de um método matemático de decomposição da forma de onda analisada em uma somatória de outras ondas, mensurando a intensidade individual de cada um dos harmônicos e totalizando suas componentes.

De acordo com a norma Brasileira, as distorções harmônicas são fenômenos associados com deformações nas formas de onda das tensões e correntes em relação à onda senoidal da frequência fundamental (PRODIST Módulo 8 Revisão 6, 2015).

A Distorção na forma de onda é definida como o desvio na forma de onda senoidal ideal, considerando a frequência do sistema como referência, principalmente caracterizado pelo conteúdo espectral (nível CC, harmônicos, inter harmônicos, cortes e ruídos) (DUGAN *et al.*, 2003).

3.3.3.1 Procedimentos de Medições de Harmônicos

Os procedimentos de rede estabelecem que o processo de apuração dos valores das distorções harmônicas devem ser realizado, sobretudo em campanhas de medição, coordenados pelo ONS. Em barramentos dos transformadores de fronteira onde existem registros de reclamações ou onde o desempenho de cargas não lineares ou especiais implique a ocorrência de valores expressivos de distorção harmônica de tensão.

O sub-módulo 2.8 (ONS), estabelece que a duração de uma campanha de medição deve ser de sete dias consecutivos, de acordo com protocolo de apuração estabelecido para o indicador em análise, considerando os valores dos indicadores integralizados em intervalos de dez minutos.

O instrumento de medição utilizado no processo de apuração dos indicadores deve ter desempenho compatível com equipamento classe A da IEC, conforme estabelecido na norma IEC 61000-4-30 e IEC 61000-4-7.

3.3.3.2 Os Valores de Referências das Distorções Harmônicas

Os valores utilizados como referência para limites de distorções harmônicas totais de tensão estão descritos na Tabela 1, Já na Tabela 2 estão descritos os valores de referências das distorções individuais de cada harmônica. Estes valores são utilizados como referência do planejamento elétrico em termos de QEE.

Tabela 1 - Valores de referência globais das distorções harmônicas totais.

Tensão nominal do Barramento	Distorção Harmônica Total de Tensão (THDV) [%]
$V_N < 1\text{kV}$	10
$1\text{kV} < V_N < 13,8\text{ kV}$	8
$13,8\text{ kV} < V_N < 69\text{kV}$	6
$69\text{kV} < V_N < 230\text{ kV}$	3

Fonte: ANEEL – Prodíst-Modulo 8.

Tabela 2 - Níveis de referência para distorções harmônicas individuais de tensão.

Ordem Harmônica	Distorção Harmônica Individual de Tensão [%]				
	$V_n < 1\text{ kV}$		$1\text{ kV} < V_n < 13,8\text{ kV}$	$13,8\text{ kV} < V_n < 69$	$69\text{ kV} < V_n < 230$
Ímpares não múltiplas de 3	5	7,5	6	4,5	2,5
	7	6,5	5	4	2
	11	4,5	3,5	3	1,5
	13	4	3	2,5	1,5
	17	2,5	2	1,5	1
	19	2	1,5	1,5	1
	23	2	1,5	1,5	1
	25	2	1,5	1,5	1
	>25	1,5	1	1	0,5
Ímpares múltiplas de 3	3	6,5	5	4	2
	9	2	1,5	1,5	1
	15	1	0,5	0,5	0,5
	21	1	0,5	0,5	0,5
	>21	1	0,5	0,5	0,5
Pares	2	2,5	2	1,5	1
	4	1,5	1	1	0,5
	6	1	0,5	0,5	0,5
	8	1	0,5	0,5	0,5
	10	1	0,5	0,5	0,5
	12	1	0,5	0,5	0,5
	>12	1	0,5	0,5	0,5

Fonte: ANEEL – Prodíst-Módulo 8.

3.4 Considerações Finais

Neste capítulo, foram apresentados os conceitos sobre QEE, as normas e padrões Brasileiros, Norte Americanas e Europeus que são utilizadas como padrões de avaliações da qualidade de energia elétrica no mundo. Bem como as definições sobre os distúrbios harmônicos provocados pela presença de cargas não lineares, o produto energia elétrica, o processo de medição dos harmônicos e os níveis de referência aceitos nos sistemas elétricos

CAPÍTULO IV

PROCESSO DE KDD E OS MODELOS DE INTELIGÊNCIA COMPUTACIONAL

4.1 Considerações Iniciais

Nos dias atuais, onde os mercados apresentam um alto grau de competitividade entre as empresas na busca de conhecimento de mercados e produtos, normalmente todos buscam suas informações através de alguma técnica ou sistema de banco de dados, onde os dados de clientes produtos e mercados estão armazenados. De forma geral as bases de dados são ferramentas indispensáveis na extração de conhecimento.

Atualmente a tarefa de transformar dados em conhecimento tem permanecido extremamente manual, o que torna os procedimentos de extração de padrões em bases de dados muitas vezes ficar mais custosos, lento e altamente subjetivo, muitas vezes também tornando inviável em grandes volumes de dados.

Buscando uma forma de solucionar, ou pelo menos mitigar tais problemas no contexto de aprendizagem em bases de dados, submetem-se então os dados a um processo orientado por computador ou processo de descoberta de conhecimento, este melhora não somente em relação à eficiência, mas também quanto à representação do conhecimento contido na base.

Neste capítulo serão abordadas as características fundamentais do processo de descoberta de conhecimento em base de dados conhecido como KDD, expondo o seu comportamento no transcorrer de cada uma das fases de: seleção, pré-processamento, transformação, Mineração de dados (*Data Mining*) e Interpretação; empregando maior ênfase na etapa de mineração de dados e nos algoritmos computacionais utilizados em árvores de decisão (AD) que é a técnica empregada como alvo no processo de aprendizagem proposto no desenvolvimento do presente capítulo.

4.2 Processo do KDD

O termo *Knowledge Discovery in Databases*, ou KDD, refere-se ao amplo processo de encontrar o conhecimento em dados, e enfatiza a aplicação de "alto nível" de métodos de mineração de dados. Em sua natureza, o KDD engloba as áreas de aprendizagem de máquina, reconhecimento de padrões, bancos de dados, estatísticas, inteligência artificial, e visualização de dados (FAYYAD *et al.*, 1996).

KDD é o processo automatizado que se baseia em métodos de diversas áreas, como reconhecimento de padrões, estatísticas, aprendizado de máquina, redes neurais, etc. a encontrar padrões de dados na etapa de mineração de dados do processo de KDD. A frase "*Knowledge Discovery in Databases*" foi criada na primeira oficina KDD em 1989 por Piatetsky-Shapiro, que enfatizou que o conhecimento é o produto final de uma descoberta data driven (PIATETSKY-SHAPIRO; FRAWLEY, 1991).

De acordo com Fayyad *et al.* (1996), o KDD pode ser dividido em cinco fases: seleção, pré-processamento, transformação, data mining e interpretação, de acordo com a Figura 1. Logo abaixo será exposto um descritivo de cada uma das fases do processo.

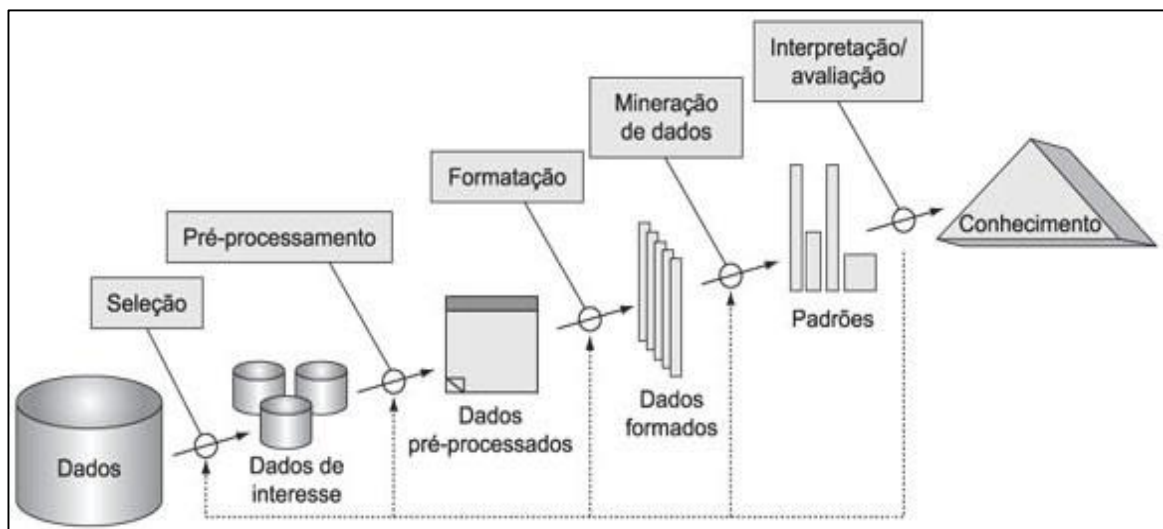


Figura 1 - Processo KDD (FAYYAD et al., 1996).

4.2.1 Fase de Seleção

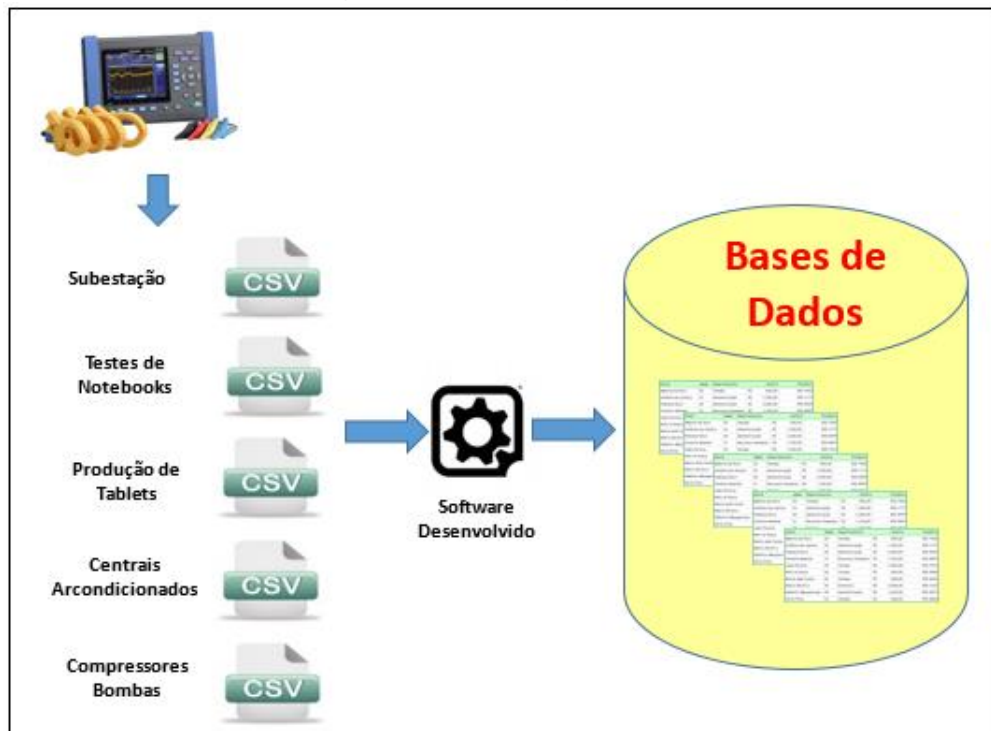
É a fase do processo de descoberta do conhecimento que possui grande importância para o processo de KDD, é nessa fase que os dados de maior importância são selecionados as variáveis que na sequência serão transformadas em características ou atributos e devem ser agrupadas em uma fonte única de dados ou de tabelas.

Normalmente os dados são extraídos de várias fontes distintas como: Sistemas de Gerenciamento de Banco de dados (SGBD), planilhas, arquivos de textos e ou sistemas legados.

Nesta fase se faz necessário o papel de um especialista do domínio, ou seja, alguma pessoa que possua abrangência sobre o assunto para determinar quais os dados que realmente são importantes e quais podem ser descartados ou simplesmente desconsiderados nas análises.

Os dados normalmente são coletados em fontes diversas, algumas vezes, não apresentam um padrão que seja legível pelos softwares de mineração, então, é necessário o desenvolvimento de um software específico para padronizar e adequar os dados coletado, como mostrado na Figura 2.

Figura 2 - Fase de Seleção do KDD.



Fonte: O autor.

4.2.2 Fase de Pré-Processamento

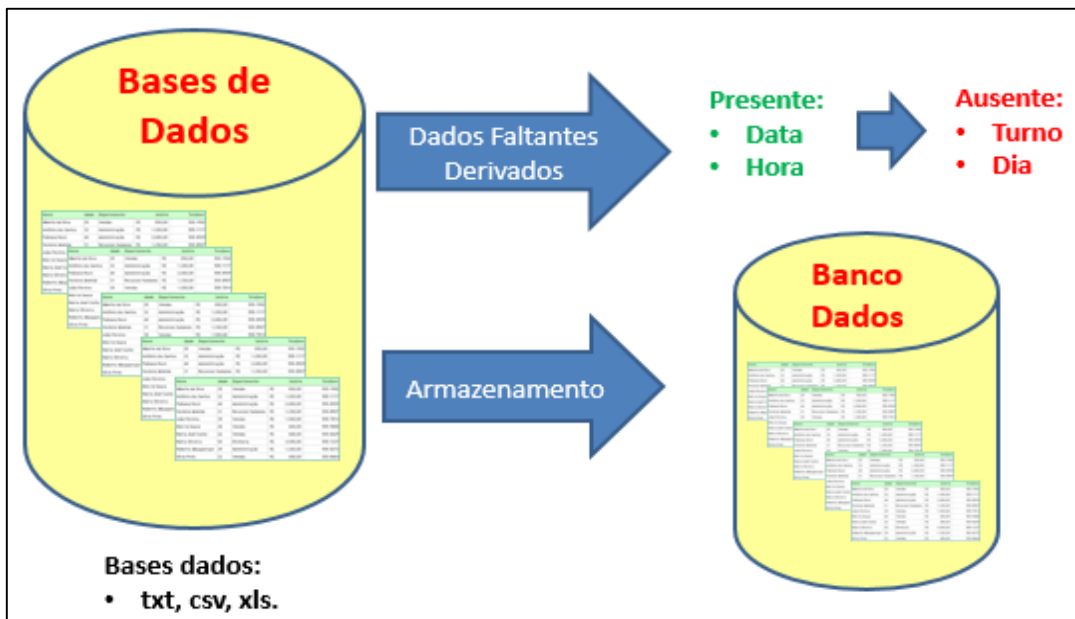
É uma das partes mais determinantes no processo de KDD, uma vez que a qualidade dos dados vai definir a eficácia dos algoritmos de mineração. Nesta etapa são realizadas tarefas que suprimam ou eliminem ruídos, dados redundantes e ou inconsistentes, restauração ou recuperação, de dados truncados ou incompletos e avaliação dos prováveis dados discrepantes.

Ainda nesta fase são utilizados métodos de geração de dados faltantes, redução, armazenamento e ou de transformação de dimensionalidade de quantidade de variáveis, com objetivo de melhorar o desempenho do algoritmo de análise, pois quanto maior for a base dados a ser analisada, mais complexa e demorado será o trabalho de mineração, exigindo um tempo maior no processamento e obtenção dos resultados.

4.2.3 Fase de Transformação

É a fase do KDD que antecede a fase de mineração de dados, após serem selecionados, limpos e pré-processados, os dados devem ser armazenados e formatados adequadamente para que os algoritmos possam atuar, extraíndo conhecimento. Nesta etapa, os dados são organizados de acordo com as particularidades e necessidades das técnicas de extração de conhecimento que serão utilizadas na etapa seguinte como demonstrado na Figura 3.

Figura 3 - Fase de Transformação de dados do KDD.



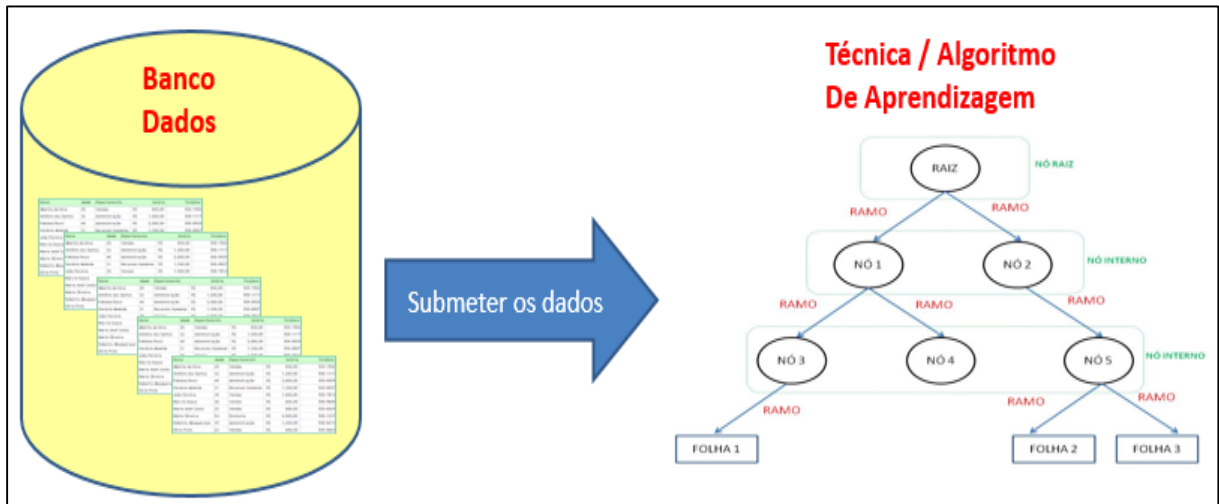
Fonte: O autor.

4.2.4 Fase de Mineração de Dados

É a fase que envolve a busca por informações relevantes, com o desenvolvimento de padrões e modelos de aprendizagem, necessários para representação, associações, mudanças, anomalias, a partir dos dados que foram extraídos e analisados por um profissional especializado. Nesta fase os dados são submetidos a um ou mais algoritmos de aprendizagem

para extrair o conhecimento de aprendizagem e de regras, como demonstrado na Figura 4. A técnica utilizada nas análises dos dados foi à árvore de decisão.

Figura 4 - Processo de Aprendizagem.



Fonte: O autor.

Em mineração de dados podemos utilizar uma grande quantidade de algoritmos de aprendizagem que compõem os paradigmas de aprendizagem disponíveis. Podemos destacar alguns como: simbólico, conexionista, baseados em regras, evolutivos ou genético e o estatístico.

O paradigma utilizado em nosso estudo de caso, foi o simbólico, por representar o conhecimento adquirido com uso de expressões lógicas, nas quais os símbolos são utilizados para representar as relações entre objetos, que justifica seu uso em árvore de decisão, pois as decisões tomadas em AD são facilmente representadas por instruções (se...então) que é a base das comparações utilizadas nas árvores de decisão.

4.2.5 Fase de Interpretação de Dados

Na fase de interpretação os dados que foram minerados são validados, onde o especialista do domínio de aplicação é fundamental para homologação do conhecimento adquirido, pois nesta fase são validados todos os resultados obtidos no projeto.

4.3 Algoritmos de Inteligência Computacional

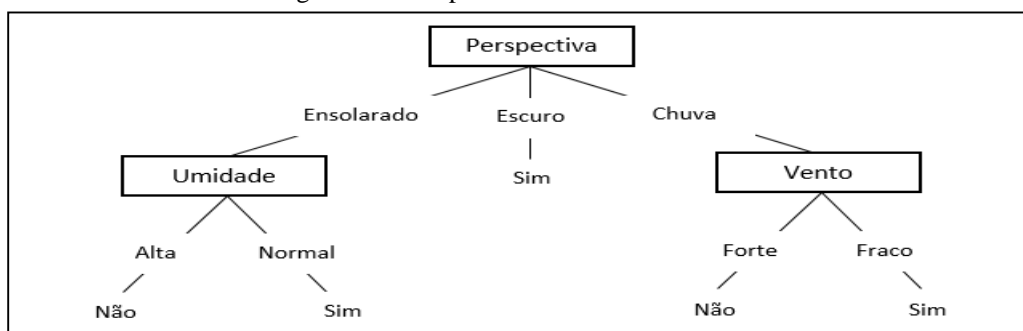
4.3.1 Árvores de Decisão

As árvores de decisões constroem modelos de classificação ou de regressão na forma de uma estrutura de árvore, transforma um conjunto de dados em subconjuntos cada vez menores na forma uma árvore, com o objetivo de transformar um grande problema em problemas menores, facilitando assim a tomada de decisões.

De acordo com Mitchell (1997), As árvores de decisões classificam instâncias, classificando-os da raiz da árvore para baixo até algum nó folha, que prevê a classificação da instância. Cada nó da árvore especifica um ensaio de algum atributo do exemplo, e cada ramo descendente, a partir desse nó corresponde a um dos valores possíveis para este atributo. Um exemplo é classificar, iniciando no nó raiz da árvore, testando o atributo especificado por este nó, em seguida, movendo-se para baixo, o galho de árvore correspondente ao valor do atributo no exemplo dado. Este processo é repetido para a Sub-árvore com raiz no novo nó.

Logo abaixo na Figura 5, será mostrada uma típica apresentação de uma árvore de decisão que classificou os sábados pela manhã como sendo mais adequados para jogar tênis de quadra. Na figura são utilizados vários parâmetros (instâncias) para a descoberta de conhecimento com uso de árvore de decisão, As instancias utilizadas foram: (Perspectiva = ensolarado; Temperatura = Quente; Umidade = Alta; Vento = Forte), de acordo com o algoritmo os ramos mais baixos do lado esquerdo foram classificados como um exemplo negativo (ou seja, a árvore prevê que jogar tênis = não). No entanto, para que houvesse jogo seria necessário que o dia estivesse ensolarado com umidade normal.

Figura 5 - Exemplo de uma Árvore de Decisão.



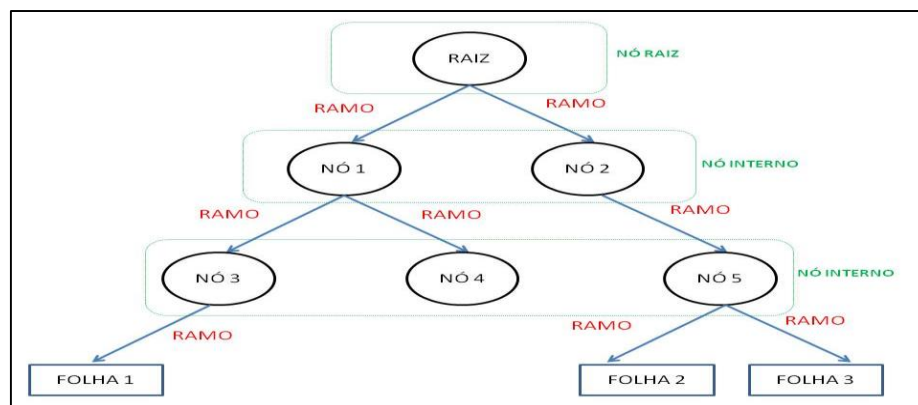
Fonte: o autor.

As árvores de decisão são técnicas bem sucedidas, mas simples para prever e explicar a relação entre algumas medições sobre um item e seu valor-alvo. Além de seu uso em mineração de dados, árvores de decisão, que originalmente derivados de lógica, gestão e estatísticas, são hoje ferramentas altamente eficazes em outras áreas, como a mineração de texto, extração de informações, aprendizado de máquina e reconhecimento de padrões (ROKACH, 2008). Também oferecem muitos benefícios:

- a) Versatilidade para uma ampla variedade de tarefas de mineração de dados, tais como classificação, regressão, clustering e seleção de recursos;
- b) Autoexplicativo e fácil entendimento;
- c) Flexibilidade no tratamento de uma variedade de dados de entrada: nominais, numérica e textuais;
- d) Adaptabilidade em conjuntos de dados de processamento que pode ter erros ou valores ausentes;
- e) Desempenho de previsão de alta para um relativamente pequeno esforço computacional;
- f) Disponível em muitos pacotes de mineração de dados através de uma variedade de plataformas;
- g) Útil para grandes conjuntos de dados;

Uma árvore é formada por vários componentes: a raiz, que inicializa a árvore, os ramos que são os rótulos com os resultados possíveis de cada teste, os internos nós internos da árvore, que corresponde a um teste do valor de uma propriedade, as folhas que especificam os valores a serem retornados se uma folha for alcançada, como mostra a Figura 6.

Figura 6 - Componentes de uma Árvore de decisão.



Fonte: o autor.

4.3.2 Algoritmos Utilizados em Árvores de Decisão

No processo de aprendizagem de máquina, existe uma grande quantidade de algoritmos, este estudo enfatiza os mais relevantes nas literaturas pesquisadas, tendo como objetivo evidenciar tão somente o algoritmo J48 que é uma recodificação em linguagem de programação Java do algoritmo C4.5 escrito em sua origem em linguagem C, esse algoritmo é bem aceito no meio acadêmico e por especialistas, por ser adequado para os procedimentos que envolvem variáveis (dados analisados) quantitativas contínuas e discretas. Os dados analisados nos experimentos deste trabalho possuem características: numéricas, quantitativa e discreta, que podem ser utilizadas com árvore de decisão, o que justifica a sua utilização.

O algoritmo ID3, foi idealizado por Ross Quinlan no ano de 1979 na Universidade de Sydney - Austrália, é um algoritmo que a partir de um conjunto de dados constrói árvores de decisão. O ID3 é o antecessor do algoritmo C4.5, que é muito utilizado na aprendizagem de máquina e em processamento de linguagem natural.

O Algoritmo C4.5, pode ser usado para auxiliar em tarefas onde alguns profissionais necessitam de um classificador, que em alguns momentos necessitam somente de uma resposta de sim ou não. No entanto, este algoritmo de classificação possui a capacidade de aprender, olhando para um conjunto de casos, classifica-los e a partir deles realizar previsões para novos casos, Ou seja, C4.5 gera um classificador que é capaz de atuar como um especialista, classificando inclusive os casos desconhecidos.

Com a necessidade de recodificar e melhorar o C4.5, surge o J48 em Linguagem Java, que é trabalhado independente de sistema operacional de máquina, ou seja, pode ser executado em sistemas operacionais como: Windows, Linux, Unix, Mac e demais sistemas operacionais existentes no mercado.

Um dos aspectos mais relevantes pela comunidade de especialistas em data mining é que esse algoritmo, pode trabalhar com procedimentos que envolvam variáveis qualitativas contínuas e discretas que estão presentes nas bases de dados. O algoritmo que foi proposto por (Quinlan, 1993), é apresentado como o melhor resultado na montagem de árvores de decisão.

4.3.3 *Classificação de uma Árvore de Decisão*

A técnica de classificação é o processo de descoberta de um modelo, utilizando as características disponíveis nos dados de treinamento, para desenvolver modelos para as classes, ou seja, independente da técnica de mineração orientada por classificação, sempre serão utilizados os dados do passado (dados conhecidos ou treinamento) para prever o futuro (modelo de classificação).

A classificação é uma das tarefas mais importantes em mineração de dados e, hoje em dia, tem sido aplicada para resolver problemas relacionados a diferentes áreas, como administração, finanças, educação, saúde e outros. Portanto, a construção de classificadores precisos e computacionalmente eficientes constitui um desafio relevante no campo de mineração de dados (MERSCHMANN; PLASTINO, 2010).

O intenso interesse neste assunto resultou no desenvolvimento de um grande número de técnicas para a construção de classificadores de árvores de decisão, tais como: ID3 (QUINLAN, 1979), CART (BREIMAN et. al, 1984), C4.5 (QUINLAN, 1993), recentemente o Microsoft SQL Server (SEIDMAN, 2001) implementaram árvores de decisão para problemas de classificação e regressão.

As técnicas de classificação são geralmente classificadas em dois tipos: as abordagens ansiosas e preguiçosas. Métodos de classificação ansiosos, como árvores de decisão, constrói um modelo de generalização a partir de um conjunto de dados de treinamento, antes que qualquer instância de consulta for recebida para a classificação. Eles classificam novas instâncias (sem rótulo), usando diretamente o modelo aprendido. De um modo diferente, métodos preguiçosos (Aha, 1997), como k-vizinhos mais próximos, não previamente constrói um modelo de generalização a partir de um conjunto de dados de treinamento para classificar novas instâncias. Para cada instância possa ser classificada, eles processam as amostras dos treinamentos armazenados (MERSCHMANN; PLASTINO, 2010b).

4.3.4 *Aplicação do Classificador C4.5 ou J48*

A tarefa do classificador C4.5 ou J48 é desenvolvida, após o conjunto de dados, ter passado pelos processo de seleção, pré-processamento e formatação, então, os dados são

submetidos a tarefa de aprender utilizando um classificador, utilizando uma função alvo f que mapeie cada um dos atributos de x , para um dos rótulos de classes y predeterminados. A função alvo é conhecida também como modelo de classificação. De forma geral, o classificador aplica recursivamente a função alvo em todo o conjunto de dados, até que alcance a condição de parada ou não existam mais dados para classificar e no final exibe a árvore de decisão com os resultados alcançados.

No contexto de classificação, possuímos algumas métricas que são utilizadas na classificação de AD, de acordo com essas medidas podemos destacar o grau de impureza de cada classe de dados e assim classifica-las adequadamente no processo de construção de uma AD. No presente estudo, serão utilizados o índice gini, medida de entropia, assim como o ganho de informações para classificação.

O índice de gini, foi criado pelo italiano Conrado Gini (1912), é utilizado para medir o grau de concentração de renda em determinado grupo. No contexto de AD, mede a impureza de um atributo em relação a classe, seu cálculo é medido com o uso da equação 3 apresentada abaixo.

$$\text{índice gini} = 1 - \sum_{i=1}^c p_i^2 \quad (3)$$

Onde:

p_i : é igual a frequência relativa de cada classe em cada nó;

c : é igual ao número de classes;

Quando este índice é igual a zero, o nó é puro. No entanto, quando se aproxima do valor um (1), o nó é impuro.

Entropia é a estimativa de ganho de informação é fundamentado em uma medida utilizada na teoria da informação. Caracteriza-se por medir o grau pureza ou impureza de um conjunto de dados, pode ser assinalada pela medida da falta de homogeneidade dos dados em relação a sua classificação, a equação 4 é utilizada nos cálculos de entropia.

A entropia para qualquer s pode ser calculada como:

$$\text{entropia}(s) = \sum p_i \log_2 p_i \quad (4)$$

Pode-se ter:

$$\text{entropia}(S) = - (p_1 \log_2 p_1 + p_2 \log_2 p_2 + \dots + p_n \log_2 p_n)$$

Onde:

(S) é o conjunto de exemplo;

(n) é o número de valores possíveis da classe

(pi) é a proporção de amostras da classe i em relação ao total de amostras

Nos experimentos realizados na presente pesquisa, optou-se por utilizar o cálculo de entropia na formação das ADs, por apresentarem os melhores resultados nas análises de ganhos de informações.

O ganho de informação consiste em um método estatístico, utilizado para medir o grau de impureza de um conjunto de dados, através do cálculo de entropia. O cálculo de ganho de informação pode ser obtido através da equação 5 descrita logo abaixo:

$$\text{Ganho}(S, A) = \text{Entropia}(S) - \sum((|S_v| / |S|) * \text{Entropia}(S_v)) \quad (5)$$

Onde:

Ganho (S, A) = ganho do atributo A sobre o conjunto S

S_v = subconjunto de S para um valor do atributo A

$|S_v|$ = número de elementos de S_v

$|S|$ = número de elementos de S

4.3.5 *Processo de Poda*

A poda é uma técnica de aprendizado de máquina que reduz o tamanho das árvores de decisão, removendo ramos da árvore que fornecem pouca capacidade de classificar instâncias. Possuindo dois objetivos básicos: reduzir a complexidade do classificador final e melhorar precisão preditiva pela redução de “*overfitting*” ou superajuste e remoção de seções de um classificador que pode ser baseado em dados ruidosos ou errôneos. No processo de poda, existem duas abordagens na aplicação da técnica: Pré-Poda e Pós-Poda.

4.3.5.1 *Pré-Poda (realizado durante a criação da árvore)*

A técnica consiste em regras de paradas que previnem a construção de ramos que não poderiam melhorar a predição da AD. A vantagem de uso desta abordagem é a prevenção de criação de sub-árvores muito complexas.

4.3.5.2 Pós-Poda (realizada depois da criação de árvore)

Podar uma árvore, neste contexto, significa que após ser totalmente construída, ela será podada, ou seja, alguns dos ramos que não possuem significados serão retirados da árvore, este tema foi amplamente descrito por (BREIMAN et al., 1984; QUINLAN, 1993), de acordo com Quinlan (1988) “*Construir e podar uma é mais lento, mas mais confiável*”.

4.4 Considerações Finais

Neste capítulo, foi exposto o processo de extração de conhecimento em bases de dados (KDD), que tem como objetivo básico, adquirir conhecimento através de bases de dados, objetivando extrair padrões, regras e relações extraídas da base de conhecimento, fornecendo ao usuário, métodos que auxiliem no processo de avaliações e tomadas de decisões.

Foram descritas as etapas do processo do KDD de forma objetiva, contudo, dando uma maior ênfase à etapa de mineração de dados ou “*data mining*”. Ressaltado também a importância da etapa de pré-processamento, que é a mais demorada do processo KDD, na qual a base de dados é filtrada, retiradas as inconsistências, eliminação de dados redundantes, eliminação de ruídos e por fim, envia os dados para a mineração de dados.

CAPÍTULO V

APLICAÇÃO DO KDD NOS DADOS DA REDE ELÉTRICA DA INDÚSTRIA

5.1 Introdução

Analisar, quantificar e qualificar a qualidade da energia elétrica é um dos novos problemas que surgiram com as melhorias e modernizações nos processos industriais. Além disto, com a normatização proposta pela agência responsável, ANEEL, nos seus Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST no módulo 8, este módulo possui todas as recomendações necessárias para o controle, propagação e recomendações de limites mínimos e máximos dos indicadores de qualidade permitidos de acordo com os procedimentos estabelecidos e revisados pela ANEEL.

Na presente pesquisa foi realizado um estudo sobre os efeitos dos fenômenos associados as deformações de onda, também conhecidos como Harmônicos. Na rede interna de uma indústria do PIM, desenvolveu-se uma análise de alguns processos realizados nessa indústria que possui vários processos de manufatura de equipamentos eletroeletrônicos em suas instalações, no entanto, foram analisados somente cinco (5) pontos do seu processo de manufatura.

O estudo teve apoio de ferramentas computacionais para analisar os dados que foram monitorados e coletados em cada um dos processos, estes por sua vez, foram submetidos ao processo de data mining e uma técnica de IC que deu apoio ao processo de tomada de decisão.

5.2 Características da Indústria Analisada

A empresa analisada é fabricante de produtos de alta tecnologia como desktops, notebooks, netbooks, celulares e TVs dentre outros, seus produtos são verdadeiros sinônimos de inovação tecnologia. Está presente entre os líderes em vendas do segmento no varejo, possui grandes parcerias como INTEL, Microsoft e Qualcomm.

É um dos maiores conglomerados do Polo Industrial em Manaus. Faz parte de um grupo de seis empresas que atuam em diversos segmentos. Desde a montagem de produtos de tecnologia até a logística, passando por televisores LCD/LED, placas de circuito impresso, painéis de LCD, injeção plástica, metalurgia e papel e papelão, a sinergia entre as empresas permite mais flexibilidade e rapidez no processo produtivo.

5.2.1 - Pontos Analisados e suas Características

5.2.1.1 Subestação

É o ponto de entrada de energia elétrica que agrega os equipamentos para transmissão e distribuição, além de equipamentos de proteção e controle de fluxo energético. Na Tabela 3, estão descritas todas as características elétricas presentes na subestação que alimenta a indústria pesquisada e os processos envolvidos.

Tabela 3 - Características da subestação de entrada.

Saída do Transformador de 1000kVA 220/127V, Alimenta a planta da Indústria um Disjuntor Trifásico 3200A.	
Carga conectadas	Cargas de Fonte Chaveada Inversores de frequência UPS Lâmpadas fluorescentes
Comentário sobre o regime de funcionamento	São ligados 24:00 por dia

Fonte: O autor.

5.2.1.2 Centrais de Ar-Condicionados

As centrais de ar-condicionados são responsáveis por aquecer, esfriar ou ventilar os vários ambientes envolvidos nos processos de manufatura da indústria analisada, na Tabela 4, estão descritas todas as características elétricas, bem como os horários de funcionamentos.

Tabela 4 - Centrais de Ar-condicionados.

Cargas	04 Centrais de Ar-condicionado. Cada Central funciona com 2 compressores de 15TR (52,75 kW)
Características Harmônicas	Circuito de Carga Linear com baixo impacto na distorção harmônica
Comentários sobre o regime de funcionamento	São ligados entre 05:30 até 01:30 de segunda a sexta.

Fonte: O autor.

5.2.1.3 Linha de Testes de Notebooks (Burning)

O Processo de testes de notebooks é o ponto, onde todos os notebooks manufacturados na indústria são submetidos aos mais variados tipos de testes de conformidades de hardwares e softwares, instalados durante o processo de produção. A Tabela 5 apresenta todas as características presentes em suas especificações elétricas e de consumo.

Tabela 5 - Linha de Testes de notebooks - Burning Test.

Cargas	Em torno de 500 notebook de 65W cada bifásico Especificação da fonte chaveada de 65W
Regime de Funcionamento	Cada notebook fica ligado por 1 hora e é substituído por outro. No período de 01:30 às 06:00 ficam ligados os notebook do dia anterior. Durante o final de semana os notebooks ficam ligados na rede elétrica. Todos os notebooks são alimentados por um No-Break (UPS) de 80kVA.

Fonte: O autor.

5.2.1.4 Compressores e Bombas

Os compressores e bombas são responsáveis por alimentar os sistemas hidráulicos e pneumáticos como: braços robóticos, esteiras, parafusadeiras envolvidos no processo de manufatura. Na Tabela 6, estão presentes todas as características elétricas, horários de funcionamentos e quantidades de aparelhos envolvidos nos processo da indústria.

Tabela 6 - Sala de Compressores e bombas.

Cargas	1 Compressor de 75 CV (55,93 kW) 1 Bomba de Recalque de 25 CV (18,64 kW) 1 Bomba de Torre de 25 CV (18,64 kW)
Comentários sobre o regime de funcionamento	São ligados entre 05:30 até 01:30 de segunda a sexta.

Fonte: O autor.

5.2.1.5 Produção de Tablets

O processo de produção de tablets envolve uma série de ferramentas, gabinetes, bancadas e esteiras, na Tabela 7, estão descritas todas as características envolvidas no processo, assim como, os hardwares envolvidos e suas quantidades.

Tabela 7 - Montagem de Tablets.

Cargas	10 Impressoras Laser com 57,96W cada 4 Balanças Eletrônicas com 15W cada 10 parafusadeiras elétricas com motor de 40W cada 11 switch de rede com 10,3 W cada 4 Monitores de LED com 45W cada 1 Computador desktop de 352W 1 Ferro de Soldar de 70W 3 Ionizador de Ar com 26,4 W cada 8 impressora de etiqueta com 110W cada 17 notebooks com 65W cada 25 lâmpadas LED com 20W cada 12 dispensador eletrônico de parafusos com 6W cada Em torno de 140 tablets ligados em teste com 10W cada.
Comentários sobre o regime de funcionamento	A Sala funciona de 06:00 às 01:30 do dia seguinte de segunda a sexta, onde os equipamentos da linha são desligados. No período de 01:30 às 06:00 ficam alguns tablets ligados. Todos os equipamentos e tablets são alimentados por um No-Break (UPS) de 80kVA.

Fonte: O autor.

5.2.2 Equipamento(s) Utilizados nas Medições

Na indústria analisada foram utilizados um total de cinco (5) analisadores de QEE, estes possuíam as características descritas na Tabela 8, os mesmos foram instalados em pontos estratégicos para monitorarem a empresa em uma campanha de medição regular de uma semana estabelecida de acordo com os padrões do ONS (Operador Nacional do Sistema Elétrico) no Sub-módulo 2.8 na diretriz 5.5, do procedimento de rede Gerenciamento dos indicadores de desempenho da rede básica e dos barramentos dos transformadores de fronteira, e de seus componentes, onde são descritos os parâmetros de medição que devem ser aplicados em decorrência de campanhas de medições regulares estabelecidas de acordo com Operador Nacional do Sistema Elétrico (ONS).

Tabela 8 - Características dos analisadores de QEE.

Tipo de linha de Medição	Monofásica 2 fios, monofásico 3 fios trifásico 3 fios ou trifásica 4 fios mais um canal de entrada extra (deve ser sincronizado para referenciar canal durante a medição AC / DC)
Faixas de tensão	Medição de tensão: 600.00 V rms
Faixas de corrente	Medida transitória 6,0000 kV pico
Exatidão básica	500,00 mA para 5,0000 kA AC (depende do sensor de corrente em uso)
Itens de medição	Tensão: $\pm 0,1\%$ da tensão nominal
Registro	Atual: $\pm 0,2\%$ leitura. $\pm 0,1\%$ F.S. + Precisão do sensor de corrente
Interfaces	Potência ativa: $\pm 0,2\%$ leitura. $\pm 0,1\%$ F.S. + Precisão do sensor de corrente
Exibição	1. Transitório sobre tensão: MHz amostragem 2
Fonte de energia	Ciclo 2. Frequência: Calculado como um ciclo, de 40 a 70 Hz
Dimensões e massa	3. Tensão (1/2) RMS: um cálculo ciclo atualizado a cada meio ciclo
Acessórios	Atuais (1/2) RMS: cálculo semicírculo

Fonte: O autor.

5.2.3 Metodologia Computacional Empregada

A metodologia computacional utilizada nas análises dos dados coletados pelos analisadores de QEE foram submetidas ao processo de KDD. O processo envolve uma série de etapas que serão descritas abaixo, juntamente com a aplicação do processo e a técnica de árvore de decisão.

5.2.3.1 Execução da etapa de seleção de dados

Nesta etapa os dados são extraídos dos analisadores de QEE, com ajuda de um software de exportação, disponibilizado pelo fabricante do analisador.

O período de coleta foi de uma semana de duração, os arquivos gerados das coletas, possuía o formato de csv, ou seja, os dados foram disponibilizados no formato lido por planilhas de cálculos. Após serem exportados, foram submetidos a uma análise de verificação e reconhecimento, com ajuda de um especialista de domínio, onde buscou-se extrair as características que possuísem representatividade em cada um dos processos envolvidos e em cada fase da rede elétrica.

Da campanha de medição realizada na indústria, foram exportados 5 arquivos CSV, que possuíam os dados das medições, os arquivos possuíam uma grande quantidade de dados e de formatos diferentes, neste momento foi necessário que o especialista de domínio analisa-se os dados para definir quais dados e seus respectivos tipos seriam utilizados nas avaliações do conteúdo harmônico existente em cada um dos pontos analisados, foram definidos a existência de 4 tipos de dados (números reais, caracteres, datas horas) que possuíam mais relevâncias nas análises.

A Tabela 9 possui um resumo dos dados que foram identificados pelo especialista de domínio, onde serão demonstrados os totais de linhas, colunas e tipos de dados de cada um dos processos que foram monitorados. Essa análise é de grande importância para as futuras definições de reduções e de completude dos dados, nesta fase deve ser bem elaborada pois servira de base para as próximas fase, caso a análise não seja bem executada, as demais fases provavelmente não serão bem executadas.

Tabela 9 - Resumo dos dados dos arquivos CSV.

Arquivos de dados CSV	Total de colunas	Total de Linhas	Tipos de dados encontrados arquivos			
Entrada do transformador	164	11.423	Números reais	Caracteres	Datas	Horas
Teste de Notebooks (Burning Test)	164	9.996				
Montagem de Tablets	164	11.350				
Compressores e Bombas	164	9.655				
Centrais de Ar-condicionados	164	11.350				

Fonte: O autor.

Ainda utilizando o conhecimento do especialista de domínio, foram definidas as estruturas iniciais de cada uma das Tabelas que receberiam os dados oriundos dos arquivos CSV, isso se faz necessário em virtude de existirem uma variedade muito grande de colunas e linhas, então, optou-se por criar as tabelas com as seguintes informações (id, data, hora, THDu de cada fase, TDHi de cada fase, harmônicos de tensão de 1° até 49° ordem, harmônicos de corrente de 1° até 49° ordem) de cada um dos processos.

A Tabela 10, exibe os campos que foram identificados nos arquivos CSV (Comma separated values). Nesta primeira análise optou-se por manter os nomes originais de cada coluna que foram exportadas juntamente com os dados, isso foi necessário para evitar que os dados pudessem sofrer mudanças ou imputação de erros nesta fase do processo do KDD.

Tabela 10 - Características dos campos das tabelas.

Campos	Descrição	Tipo de dados
Id	Ordem de cadastro	Numérico
Data	Data da coleta	Data
Hora	Hora da coleta	Hora
Avgurms1, 2 e 3	Tensão das fases A, B e C	Numérico científico
Avgirms1, 2 e 3	Corrente das fases A, B e C	
Avguthd1, 2 e 3	THD de tensão das fases A, B e C	
Avgithd1, 2 e 3	THD de correntes das fases A, B e C	
Avgu1_1, ..., até Avgu1_49	Harmônicas de tensão de 1° até 49° ordem da fase A	
Avgu2_1, ..., até Avgu2_49	Harmônicas de tensão de 1° até 49° ordem da fase B	
Avgu3_1, ..., até Avgu3_49	Harmônicas de tensão de 1° até 49° ordem da fase C	
Avgi1_1, ..., até Avgi1_49	Harmônicas de corrente de 1° até 49° ordem da fase A	
Avgi2_1, ..., até Avgi2_49	Harmônicas de corrente de 1° até 49° ordem da fase B	
Avgi3_1, ..., até Avgi3_49	Harmônicas de corrente de 1° até 49° ordem da fase C	

Fonte: O autor.

Após definido os nomes dos campos, tamanhos e seus respectivos tipos, as tabelas são armazenadas dentro de um SGBD, e em seguida, os dados foram importados dos arquivos csv para as tabelas do banco de dados. Facilitando assim, o processo de análise, visualização, apresentação e disponibilidade dos dados para as próximas fases do KDD.

5.2.3.2 Execução da etapa de pré-processamento

Nessa etapa os dados serão trabalhados com o objetivo de apresentar melhorias em sua qualidade, evitar ou diminuir o máximo possível os: ruídos (oriundos do processo de extração da fase de seleção), inconsistências, duplicações, ausências, redundâncias, balanceamento de dados, limpeza, padronização, formatação, integração e transformação dos dados, para que os mesmos possam ser representados e utilizados na construção de modelos mais fidedignos e confiáveis, tornando assim, mais fáceis e rápidos os possíveis ajustes que podem ocorrer.

Após o processo de seleção de dados da fase anterior, observou-se que os dados tinham uma formatação diferente da suportada pelos próximos processos do KDD, então optou-se por formatar os dados de acordo com as necessidades das próximas fases do processo, garantido que os mesmo não perdessem sua representatividade e valores. Na Figura 7, temos os dados no formato original, de difícil visualização e compreensão, no entanto, com ajuda do profissional de domínio, foi possível identificar e definir novo formato para os dados.

Figura 7 - Dados dos Arquivos CSV.

```
Date,Time,AvgUrms1,AvgUrms2,AvgUrms3,AvgIrms1,AvgIrms2,AvgIrms3,AvgUthd1,AvgUthd2,AvgUthd3,AvgIthd1,
AvgIthd2,AvgIthd3,AvgU1(1),AvgI1(1),AvgU1(3),AvgI1(3),AvgU1(5),AvgI1(5),AvgU1(7),AvgI1(7),AvgU1(9),AvgI1
(9),AvgU1(11),AvgI1(11),AvgU1(13),AvgI1(13),AvgU1(15),AvgI1(15),AvgU1(17),AvgI1(17),AvgU1(19),AvgI1(19),AvgU1
(21),AvgI1(21),AvgU1(23),AvgI1(23),AvgU1(25),AvgI1(25),AvgU1(27),AvgI1(27),AvgU1(29),AvgI1(29),AvgU1(31),AvgI1
(31),AvgU1(33),AvgI1(33),AvgU1(35),AvgI1(35),AvgU1(37),AvgI1(37),AvgU1(39),AvgI1(39),AvgU1(41),AvgI1(41),AvgU1
(43),AvgI1(43),AvgU1(45),AvgI1(45),AvgU1(47),AvgI1(47),AvgU1(49),AvgI1(49),AvgU2(1),AvgI2(1),AvgU2(3),AvgI2
(3),AvgU2(5),AvgI2(5),AvgU2(7),AvgI2(7),AvgU2(9),AvgI2(9),AvgU2(11),AvgI2(11),AvgU2(13),AvgI2(13),AvgU2
(15),AvgI2(15),AvgU2(17),AvgI2(17),AvgU2(19),AvgI2(19),AvgU2(21),AvgI2(21),AvgU2(23),AvgI2(23),AvgU2(25),AvgI2
(25),AvgU2(27),AvgI2(27),AvgU2(29),AvgI2(29),AvgU2(31),AvgI2(31),AvgU2(33),AvgI2(33),AvgU2(35),AvgI2(35),AvgU2
(37),AvgI2(37),AvgU2(39),AvgI2(39),AvgU2(41),AvgI2(41),AvgU2(43),AvgI2(43),AvgU2(45),AvgI2(45),AvgU2(47),AvgI2
(47),AvgU2(49),AvgI2(49),AvgU3(1),AvgI3(1),AvgU3(3),AvgI3(3),AvgU3(5),AvgI3(5),AvgU3(7),AvgI3(7),AvgU3(9),AvgI3
(9),AvgU3(11),AvgI3(11),AvgU3(13),AvgI3(13),AvgU3(15),AvgI3(15),AvgU3(17),AvgI3(17),AvgU3(19),AvgI3(19),AvgU3
(21),AvgI3(21),AvgU3(23),AvgI3(23),AvgU3(25),AvgI3(25),AvgU3(27),AvgI3(27),AvgU3(29),AvgI3(29),AvgU3(31),AvgI3
(31),AvgU3(33),AvgI3(33),AvgU3(35),AvgI3(35),AvgU3(37),AvgI3(37),AvgU3(39),AvgI3(39),AvgU3(41),AvgI3(41),AvgU3
(43),AvgI3(43),AvgU3(45),AvgI3(45),AvgU3(47),AvgI3(47),AvgU3(49),AvgI3(49)
2014/06/03,11:19:00,129.99E+0,127.50E+0,128.04E+0,5.800E+0,1.364E+0,8.467E+0,0.54E+0,0.60E+0,0.72E+0,72.42E
+0,70.45E+0,44.20E+0,129.97E+0,4.579E+0,0.39E+0,1.959E+0,0.19E+0,1.515E+0,0.28E+0,1.422E+0,0.28E+0,1.271E
+0,0.17E+0,0.939E+0,0.11E+0,0.430E+0,0.09E+0,0.107E+0,0.16E+0,0.195E+0,0.07E+0,0.097E+0,0.02E+0,0.116E
+0,0.07E+0,0.137E+0,0.07E+0,0.029E+0,0.10E+0,0.065E+0,0.04E+0,0.055E+0,0.04E+0,0.052E+0,0.02E+0,0.042E
+0,0.01E+0,0.029E+0,0.02E+0,0.016E+0,0.03E+0,0.029E+0,0.03E+0,0.018E+0,0.02E+0,0.025E+0,0.01E+0,0.010E
+0,0.01E+0,0.011E+0,0.01E+0,0.022E+0,127.48E+0,1.069E+0,0.29E+0,0.447E+0,0.16E+0,0.242E+0,0.21E+0,0.220E
+0,0.23E+0,0.267E+0,0.38E+0,0.303E+0,0.09E+0,0.238E+0,0.21E+0,0.123E+0,0.15E+0,0.062E+0,0.24E+0,0.040E
+0,0.17E+0,0.061E+0,0.11E+0,0.076E+0,0.03E+0,0.057E+0,0.10E+0,0.020E+0,0.03E+0,0.022E+0,0.10E+0,0.023E
+0,0.04E+0,0.032E+0,0.03E+0,0.015E+0,0.03E+0,0.008E+0,0.03E+0,0.013E+0,0.03E+0,0.013E+0,0.04E+0,0.005E
+0,0.02E+0,0.011E+0,0.03E+0,0.011E+0,0.04E+0,0.009E+0,128.03E+0,7.645E+0,0.58E+0,2.449E+0,0.18E+0,1.585E
```

Fonte: o Autor.

A formatação de dados e a redução se fazem necessárias, pois os dados encontravam-se com formatos diferentes e unidades diferente, impossibilitando a tarefa de reconhecimento dos dados por parte dos algoritmos do processo de KDD, então foi necessária a execução dos processos de formatações e reduções, na Tabela 11 estão descritos os campos, os formatos originais, bem como os novos formatos e a redução de quantidade de casas decimais que os

dados foram submetidos com o objetivo de reduzir a dimensionalidade e reconhecimento dos dados por parte do algoritmo de mineração de dados. Após a execução destes processos os dados serão dimensionados de forma a reduzir as análises executadas pelos algoritmos, contribuindo assim, para melhor o desempenho da execução das fases futuras.

Tabela 11 - Redução e Formatação dos dados.

Campos identificados	Formatos Originais	Nova Formatação Aplicada	Observações
Avgurms1, 2 e 3	0,12373E+0	0.124	Troca de padrão de notação científica para o padrão decimal e redução para três casas decimais.
Avgirms1, 2 e 3			
Avguthd1, 2 e 3			
Avgithd1, 2 e 3			
Avgu1_1, ..., até Avgu1_,49			
Avgu2_1, ..., até Avgu2_,49			
Avgu3_1, ..., até Avgu3_,49			
Avgi1_1, ..., até Avgi1_,49			
Avgi2_1, ..., até Avgi2_,49			
Avgi3_1, ..., até Avgi3_,49			

Fonte: O autor.

Ainda utilizando o conhecimento da fase anterior, foi necessário realizar reduções de dimensionalidade na quantidade de colunas “*campos*”, nome das colunas, presentes em cada tabela, estas mudanças são justificadas, pois, as tabelas originais de cada processo monitorado possuíam, algo em torno de 164 colunas por 10.000 e com os dados de harmônicas de 1º ordem até 49º ordem, horários, datas entre outros dados. Com ajuda do profissional de domínio decidiu-se criar 5 tabelas com os novos campos descritos na Tabela 12 das fases A, B e C da rede elétrica e considerar as harmônicas de 3º, 5º, 7º e 9º ordens, pois de acordo com as análises do especialista, possuíamos uma grande quantidade de dados com níveis harmônicos muito baixo e de pouca representatividade para as análises, então foi realizada uma nova redução de dados.

Mesmo com todas as reduções que os dados foram submetidos, sempre foram mantidos os valores originais de medições e suas respectivas unidades de medidas, com o objetivo de manter a integridade dos dados que foram coletados pelos analisadores de QEE.

Com a criação das novas tabelas foi possível realizar uma importação de dados mais segura e fidedigna dos dados que agora podem ser utilizados de forma mais prática nas próximas fases do processo de descoberta de aprendizagem. As novas tabelas criadas, importou-se os dados nos formatos adequados para as futuras fases, com isso, foi possível melhorar: a organização, a visualização, e a redução de dimensionalidade de dados.

Os campos representados na Tabela 12 são utilizados para armazenar os dados oriundos dos processos analisados nas fases (A, B e C) do sistema de alimentação elétrica. Nela estão descritos os dados, os tipos de dados e quantidades de casas decimais, assim como, o nome de cada um dos dados que estão sendo armazenados.

Tabela 12 – Nomenclaturas das variáveis utilizadas nas reduções da fase A, B e C.

Nome	Descrição do dados da fase (A, B e C)	Tipos de dados	Decimais
Id	Índice de entrada de dados na tabela	Numero	Não possui
thd_fase_a_ent	Distorção harmônica total	Numero	3 casas
h3_i_burn_fa	Distorção harmônica de corrente 3º ord. burning		
h5_i_burn_fa	Distorção harmônica de corrente 5º ord. burning		
h7_i_burn_fa	Distorção harmônica de corrente 7º ord. burning		
h9_i_burn_fa	Distorção harmônica de corrente 9º ord. burning		
h3_i_compr_fa	Distorção harmônica de corrente 3º ord. (Compr.)		
h5_i_compr_fa	Distorção harmônica de corrente 5º ord. (Compr.)		
h7_i_compr_fa	Distorção harmônica de corrente 7º ord. (Compr.)		
h9_i_compr_fa	Distorção harmônica de corrente 9º ord. (Compr.)		
h3_i_centrais_fa	Distorção harmônica de corrente 3º ord. (ar)		
h5_i_centrais_fa	Distorção harmônica de corrente 5º ord. (ar)		
h7_i_centrais_fa	Distorção harmônica de corrente 7º ord. (ar)		
h9_i_centrais_fa	Distorção harmônica de corrente 9º ord. (ar)		
h3_i_tablets_fa	Distorção harmônica de corrente 3º ord. tablets		
h5_i_tablets_fa	Distorção harmônica de corrente 5º ord. tablets		
h7_i_tablets_fa	Distorção harmônica de corrente 7º ord. tablets		
h9_i_tablets_fa	Distorção harmônica de corrente 9º ord. tablets		
Turnos_fa	Turnos: (Manhã, Tarde, Noite) Intervalo da manhã Intervalo da noite Madrugada	String	Não possui
Horas_fa	Horários que os dados foram coletados	Hora	Não possui
Dias_fa	Dias monitorados (1º dia, 2º dia, ...7º dia).	String	Não possui
Data_fa	Data de coleta dos dados	Data	Não possui

Fonte: O autor.

Após as definições de reduções e formações de dados, executou-se o sincronismo de dados, com o objetivo de garantir que os dados, mesmo estando em tabelas diferentes, seriam sincronizados de forma igual, ou seja, mesmo que as tabelas não possuam números de linhas iguais, será possível executar um processo de filtragem com objetivo de buscar somente os dados que estejam presentes nas 05 tabelas envolvidas nos processos monitorados, o sincronismo foi alcançado, executando comandos de linguagem de programação de banco de dados do padrão “*Structured Query Language*” (SQL). Após a execução da sincronização de tabelas, foi executado o procedimento de discretização, que transforma os atributos numéricos de dados em atributos nominais, ou seja, transforma dados contínuos em categorizado.

A discretização é feita por intervalos de igual frequência, ou seja, os limites de todos os intervalos são selecionados de modo que possuam as mesmas quantidades de valores. A quantidade de intervalos é um parâmetro que pode ser especificado pelo analista de domínio e também pode ser uma necessidade do algoritmo utilizado no processo de descoberta de conhecimento.

5.2.3.3 Execução da etapa de Transformação

Nesta etapa os dados que foram trabalhados nos processos anteriores são armazenados adequadamente em SGBD, para que os mesmos, possam ser utilizados na próxima fase. Na continuação da fase será executada a obtenção dos dados faltantes, que são obtidos através da combinação dos campos de dados já existente.

Nas análises realizadas, foram detectados que estavam faltando dados como os turnos de trabalhos (1-TURNO, 2-TURNO, 3-TURNO e TROCA DE TURNO), os dias de coleta (D1, D2, D3, D4, D5, D6 e D7) e os horários (H00, H01, H02, ..., H23). Os dados faltantes são de grande importância para as análises, pois de acordo com o turno, o dia e hora, serão possíveis executar diversos processos de filtragens e sincronismos de dados para a próxima fase.

5.2.3.4 Execução da etapa de Mineração de dados

Considerada como a fase mais importante, nesta fase será utilizada a técnica de classificação de árvore de decisão com o uso do algoritmo classificador J48 e validação cruzada de dados, após a execução do algoritmo obteve-se o sumário de desempenho das avaliações. Contidos nas tabelas exibidas abaixo.

Nas análises foram geradas seis tabelas, sendo: três com as matrizes de confusão e três com as análises detalhadas de acurácia do algoritmo, nas Tabelas 13, 15 e 17 estão todos os dados referentes as matrizes confusão, nas Tabelas 14, 16, 18 estão descritos os dados sobre a acurácia ou precisão obtidas pelo algoritmo.

A matriz de confusão disponibiliza uma avaliação de medida eficaz do modelo de classificação, ao demonstrar a quantidade de classificações corretas versus as classificações preditas para cada classe, sobre um conjunto de exemplos.

- Dados da Fase A:

Tabela 13 - Matriz confusão da fase A.

a	b	c	d	e	<-- classificado como
485	111	2	1	0	a = '(-inf-2.305]'
83	440	72	0	0	b = '(2.305-2.445]'
1	83	491	43	0	c = '(2.445-2.705]'
0	0	48	541	28	d = '(2.705-3.085]'
0	0	1	38	562	e = '(3.085-inf)'

Fonte: O autor.

Tabela 14 - Acurácia da Fase A.

Classes	TP Taxa	FP Taxa	Precisão	Recall	ROC Área
'(-inf-2.305]'	0,810	0,035	0,852	0,810	0,962
'(2.305-2.445]'	0,739	0,080	0,694	0,739	0,905
'(2.445-2.705]'	0,794	0,051	0,800	0,794	0,943
'(2.705-3.085]'	0,877	0,034	0,868	0,877	0,979
'(3.085-inf)'	0,935	0,012	0,953	0,935	0,991
Média ponderada.	0,831	0,042	0,834	0,831	0,956

Fonte: O autor.

- Dados da Fase B:

Tabela 15 - Matriz confusão da fase B.

a	b	c	d	e	<-- classificado como
577	41	1	0	0	a = '(-inf-1.835]'
64	476	85	0	0	b = '(1.835-2.005]'
0	104	431	49	0	c = '(2.005-2.145]'
0	2	53	518	33	d = '(2.145-2.425]'
0	0	2	54	540	e = '(2.425-inf)'

Fonte: O autor.

Tabela 16 - Acurácia da Fase B.

Classes	TP Taxa	FP Taxa	Precisão	Recall	ROC Área
'(-inf-1.835]'	0,932	0,027	0,900	0,932	0,981
'(1.835-2.005]'	0,762	0,061	0,764	0,762	0,929
'(2.005-2.145]'	0,738	0,058	0,753	0,738	0,922
'(2.145-2.425]'	0,855	0,042	0,834	0,855	0,955
'(2.425-inf)'	0,906	0,014	0,942	0,906	0,987
Média ponderada.	0,839	0,040	0,839	0,839	0,955

Fonte: O autor.

- Dados da fase C:

Tabela 17 - Matriz confusão da fase C.

a	b	c	d	e	<-- classificado como
540	53	2	2	0	a = '(-inf-2.075)'
32	512	84	1	0	b = '(2.075-2.225)'
0	98	481	34	0	c = '(2.225-2.395)'
0	1	54	517	23	d = '(2.395-2.705)'
0	0	0	55	541	e = '(2.705-inf)'

Fonte: O autor.

Tabela 18 - Acurácia da Fase C.

Classes	TP Taxa	FP Taxa	Precisão	Recall	ROC Área
'(-inf-2.075)'	0,905	0,013	0,944	0,905	0,982
'(2.075-2.225)'	0,814	0,063	0,771	0,814	0,943
'(2.225-2.395)'	0,785	0,058	0,775	0,785	0,935
'(2.395-2.705)'	0,869	0,038	0,849	0,869	0,966
'(2.705-inf)'	0,908	0,009	0,959	0,908	0,988
Média ponderada.	0,855	0,037	0,858	0,855	0,962

Fonte: O autor.

Os dados apresentados na tabela 19, são oriundos do resumo das fases A, B e C apresentados nas validações das instâncias de dados geradas para ajudar na comparação e validações de eficiência do algoritmo utilizado para minerar os dados submetidos a técnica de árvore de decisão. Os índices avaliados foram a estatística Kappa, instâncias classificadas corretas e incorretas pelo algoritmo, a cobertura total de casos.

Tabela 19 - Sumário de análises de fases - Mineração de dados.

Verificações	Fase A	Fase B	Fase C
Instâncias corretamente classificados	2519 - 83.13%	2531 83.5314 %	2591 85.5116 %
Instâncias incorretamente classificados	511 - 16.86%	499 16.4686 %	439 14.4884 %
Kappa estatística	0.7892	0.7941	0.8188
Cobertura dos casos (0,95 nível)	96.6997 %	96.4026 %	96.8647 %
Número total de instâncias	3030	3030	3030

Fonte: O autor.

5.2.3.5 Execução da etapa de Interpretação dos dados

No presente estudo, buscou-se avaliar como as correntes harmônica de ordem 3º, 5º, 7º e 9º provenientes dos processos da indústria analisada: burning, tablets, compressor e centrais de ar, influenciam na taxa de distorção harmônica total de tensão (THDv) da subestação de entrada da indústria. Para tal, construiu-se, por meio do algoritmo J48, três árvore decisão que relaciona

as correntes harmônicas de ordem 3º, 5º, 7º e 9º com o THDv do secundário do transformador de entrada da subestação da indústria, cujo nível de tensão é 220 V.

As árvores de decisão criadas apresentaram como nó raiz a corrente do 7º harmônico do processo de tablets, o que indica que essa corrente harmônica apresenta uma correlação linear significativa com o THDv da entrada da indústria.

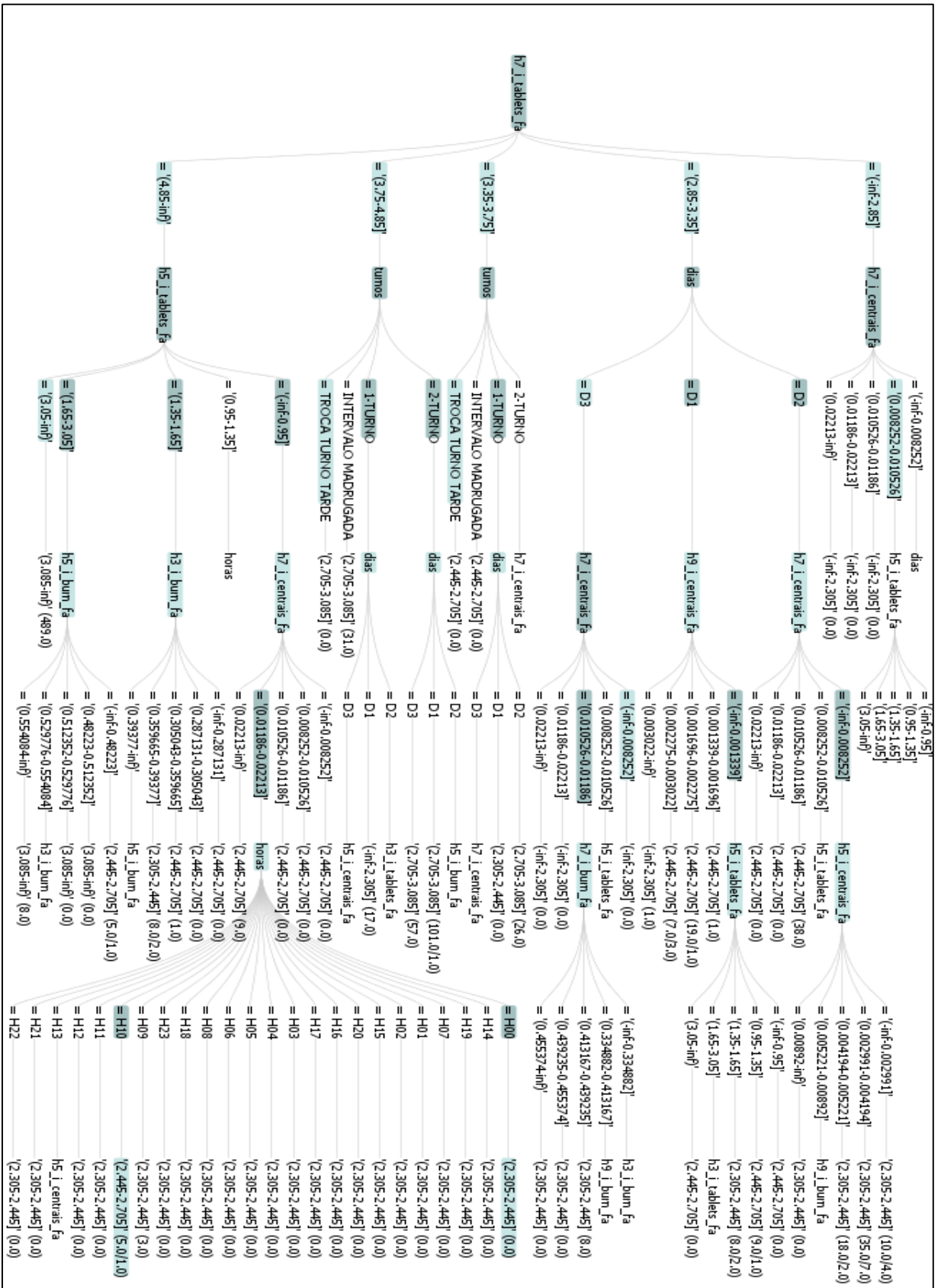
Analisando a árvore gerada na fase A, Figura 8, pode-se observar que a harmônica de sétima ordem aparece como raiz e na sequência possuímos a harmônica de sétima ordem das centrais de arcondicionados como a segunda maior influência na geração de harmônicos, depois temos a presença da harmônica de quinta ordem do processo de tablets também influenciado na geração de harmônicos, outros aspectos importantes, é que dependendo dos horários e dias o conteúdo harmônico presente nos processos podem variar do decorrer destes períodos.

Analisando a árvore gerada na fase B, figura 9, foi observado que a harmônica de sétima ordem do processo de tablets apresentou-se como raiz e na sequência temos a presença da harmônica de sétima ordem das centrais de arcondicionados, bem como a harmônica de quinta ordem dos compressores, assim também como harmônica de quinta ordem das centrais de arcondicionados e harmônica de terceira ordem dos tablets, e dependendo do dia e horários, bem como os turno de trabalho os valores do conteúdo harmônicos poderá apresentar variações.

Analisando a árvore gerada na fase C na Figura 10, foi observado que a harmônica de sétima ordem do processo de tablets apresentou a maior influência, tornando-se a raiz, na sequência foram apresentadas as harmônicas de quinta ordem das centrais, a harmônica de terceira ordem do processo de tablets e harmônica de quinta ordem do processo de tablets como as que mais influenciaram o conteúdo harmônico presente na fase C da indústria analisada.

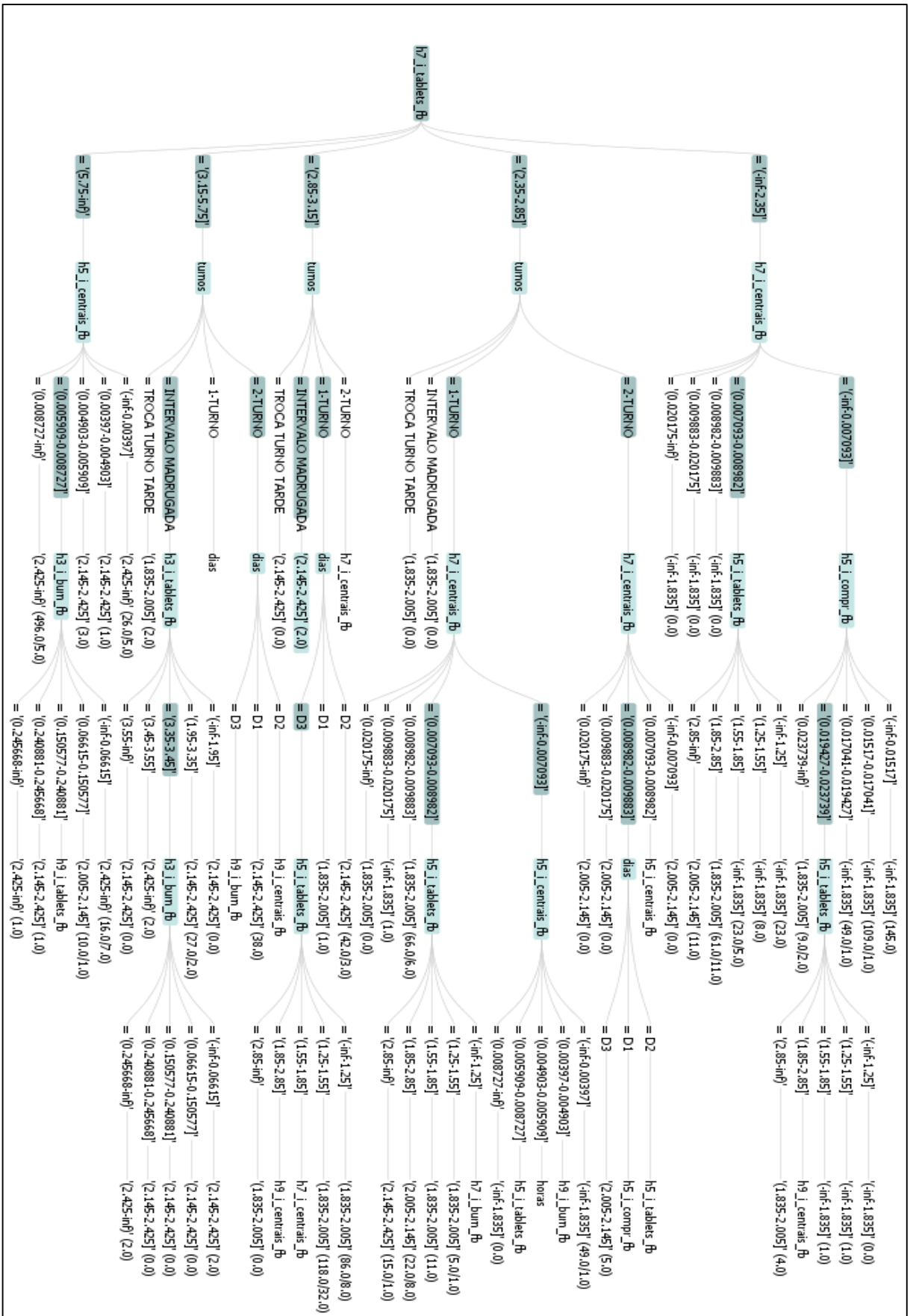
Analisando as Árvores de decisões geradas nas fases A, B e C, que estão representadas nas Figuras 8, 9 e 10, pode-se observar que dependendo dos turnos, dias e horários a harmônica de 7º ordem do processo de tablets manteve-se em níveis significativos nas três (3) fases que foram monitorados, concluindo-se então, que o conteúdo harmônico gerado pela sétima harmônica foi o de maior representatividade entre todas as harmônicas que foram monitoradas através de uma campanha de medição executadas nas três fases da indústria analisada e em seus cinco (5) processos monitorados no estudo de caso da Industria. Gerando assim o seu perfil harmônico.

Figura 8 - Árvore de decisão da Fase A.



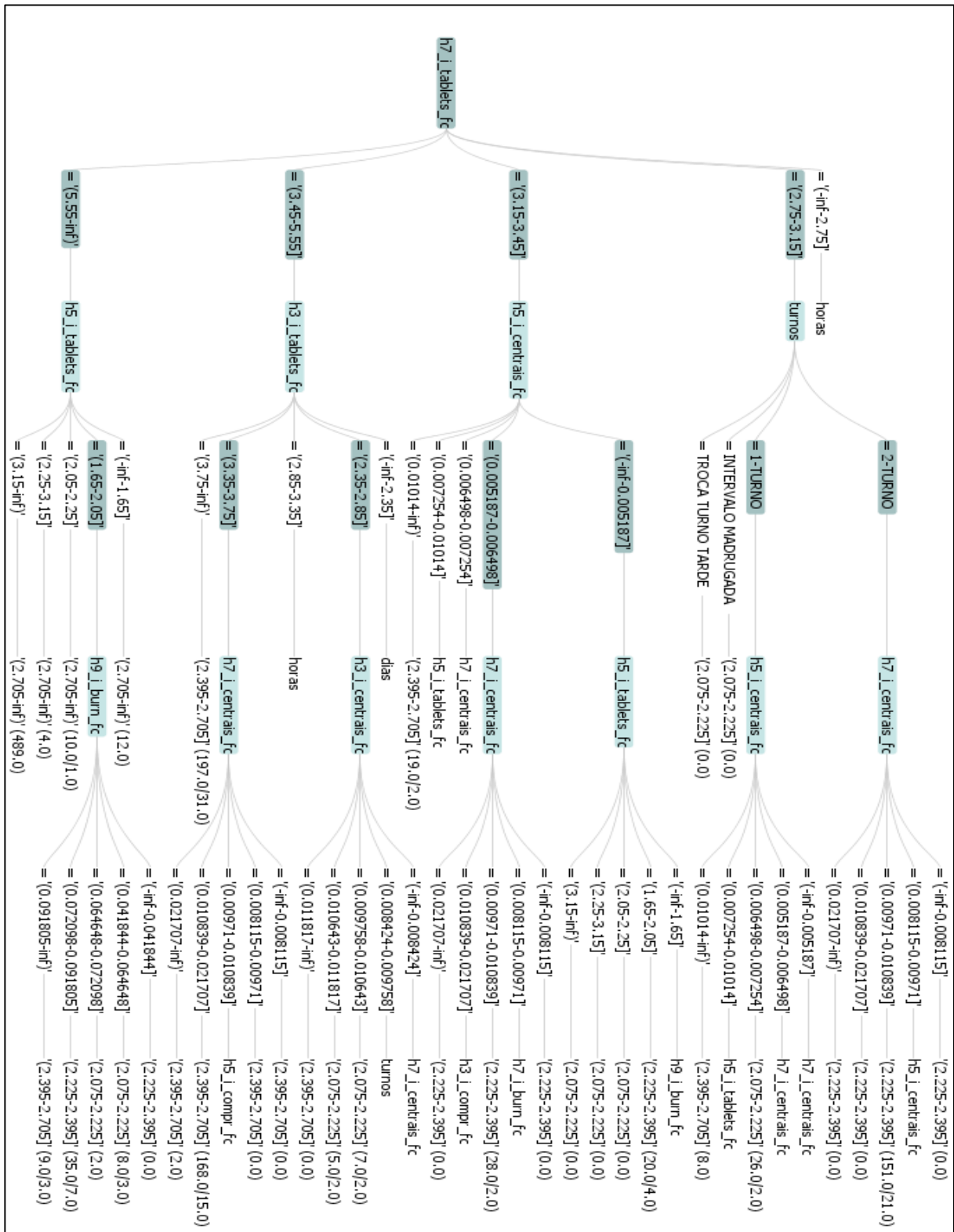
Fonte: O autor.

Figura 9 - Árvore de decisão Fase B.



Fonte: O autor.

Figura 10 – Árvore de decisão da fase C.



Fonte: O autor.

5.2.3.6 Métricas para Analisar o desempenho de um classificador

Um classificador gera uma série de dados que após a sua execução de busca de aprendizado, devem ser analisadas para avaliar o seu desempenho. Dependendo do classificador

os dados são inúmeros, entretanto, os dados que serão analisados pelo algoritmo classificador (Árvore de decisão) estão descritos logo abaixo como: Verdadeiros positivos, Falsos positivos, Falsos negativos, Verdadeiros negativos, Taxa de verdadeiros positivos, Taxas de Falsos positivos, Precisão, Cobertura, Índice Kappa e Área ROC.

No processo de avaliação de desempenho de um classificador é necessário que alguns termos e parâmetros utilizados na avaliação sejam descritos de forma clara e precisa. Na Tabela 20, foram descritos os termos e parâmetros utilizados na avaliação do algoritmo J48 empregado na árvore de decisão para classificação dos dados gerados no monitoramento do estudo.

Tabela 20 - Termos utilizados no desempenho do classificador.

Termos	Descrição
Verdadeiros positivos (TP)	Total de casos verdadeiramente classificados pelo algoritmo como positivos.
Falsos Positivo (FP)	Total de casos classificados pelo algoritmo como positivos que são falsos.
Falsos Negativos (FN)	Total de casos classificados pelo algoritmo como negativos que não são negativos.
Verdadeiros Negativos (VN)	Total de casos classificados pelo algoritmo como verdadeiros negativos.
TP Rate (TVP)	Taxa de verdadeiros positivos.
FP Rate (TFP)	Taxa de Falsos positivos.
Precisão	Número de casos positivos por total de casos cobertos
Índice KAPPA	É o índice que compara o valor encontrado nas observações com aquele que se pode esperar do acaso.
Curva ROC	É um gráfico da verdadeira taxa positiva versus taxa de falso positivo.

Fonte: O autor.

- TP Rate (Taxa Verdadeiro Positivo): Evidenciam a taxa de proporções de casos verdadeiros entre todos os casos com testes positivos. A equação 6, logo, quanto mais próximos o resultado estiver de um melhor será a classificação.

$$TVP = TP / (TP + FP) \quad (6)$$

- FP Rate (Taxa Falso Positivo): Indicam as proporções de casos falsos entre todos os casos com teste falso, portanto, quanto mais próximo de zero estiverem os resultado, melhor será a classificação (Equação 7).

$$TFP = FP / (TP + FP) \quad (7)$$

- **Precisão:** é o valor da predição positiva (número de casos positivos por total de casos cobertos), muito influenciada pela especificidade e pouco pela sensibilidade. Sensibilidade é o número de casos positivos que são verdadeiramente positivos e especificidade é o número de casos negativos que são verdadeiramente negativos (Equação 8).

$$\text{Precisão} = (TP + TN) / (TP + TN + FP + FN) \quad (8)$$

- **Índice KAPPA:** É o valor calculado dos resultados de confiança encontrados nas observações que nos fornece uma ideia do quanto o dados se afastam das observações esperadas. Quanto menor o valor de Kappa, menor a confiança de observação, o valor 1 revela uma correlação perfeita (Equação 9).

$$K = \frac{Po - Pe}{1 - Pe} \quad (9)$$

Onde:

Po: Proporção de concordâncias observadas dividida pelo total de amostras, ou seja, é a soma da diagonal principal da matriz de confusão dividido pelo total de amostras.

Pe: Proporção de concordâncias esperadas, ou seja, é a soma das somas marginais de linha e colunas dividida pelo total de amostras.

A matriz de Confusão é utilizada para quantificar os exemplos presentes em uma base de dados que seriam classificados como bom pelo modelo construído (representado pelos exemplos da diagonal principal) sendo que os demais seriam classificados mal. Na Tabela 21, é representado um modelo genérico de uma matriz de confusão com os devidos calculos que os dados são submetidos.

Tabela 21 - Matriz de Confusão Genérica.

	C0	C1	C2	Totais
C0	a	b	c	(a+b+c)
C1	d	e	f	(d+e+f)
C2	g	h	i	(g+h+i)
Totais	(a+d+g)	(b+e+h)	(c+f+i)	Total diagonal (a+e+i)

Fonte: O autor.

- **ROC:** Descrevem o desempenho de um classificador sem levar em conta a distribuição de classe ou de custos de erros. Ele traça a taxa de verdadeiros positivos no eixo vertical

contra a taxa de verdadeiro negativo no eixo horizontal. O primeiro é o número de positivos incluídos na amostra, expresso como uma percentagem do número total de pontos positivos ($TP \text{ Taxa} = 100 * TP / (TP+FN)$), o último é o número de negativos incluídos na amostra, expresso como uma percentagem do número total de pontos negativos ($FP \text{ Taxa} = 100 * FP / (FP +TN)$).

As Figura 11, Figura 12 e Figura 13 demonstram as curvas ROC (*Receiver Operating Characteristic*) das fases A, B e C com os respectivos resultados alcançados após a execução do algoritmo de árvore de decisão J48. Estas figuras demonstram os valores de sensibilidade e especificidade que foram gerados nas análises.

Os valores demonstrados nas figuras descrevem comportamento de sensibilidade e especificidade que são duas medidas utilizadas nas análises de modelos gerados a partir da execução de um dos algoritmos utilizados em mineração de dados.

Para melhor definição dos resultados alcançados das curvas ROC de cada uma das fases é necessário que seja analisado o posicionamento da curva em relação aos eixos vertical e horizontal e a linha traçada na diagonal, onde os melhores resultados são encontrados próximos ao eixo vertical acima da linha diagonal. Ou seja, os melhores valores alcançados estão localizados acima da linha diagonal e o mais próximo possível do valor um.

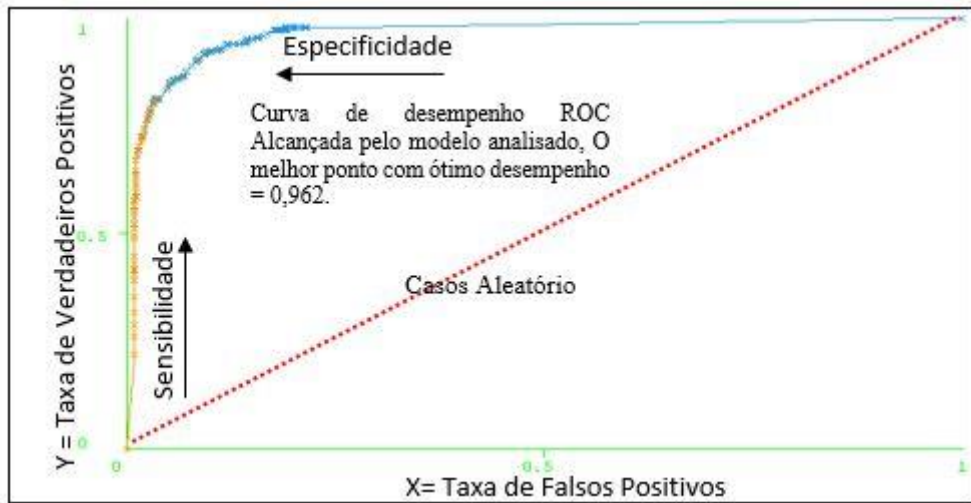
A curva ROC é um gráfico que relaciona a taxa de falsos positivos (FP) localizados no eixo “X” e a taxa de verdadeiros positivos (TP) localizados no eixo “Y” de um classificador.

A curva pode ser medida através das taxas de sensibilidade e especificidade. Estas taxas estão diretamente relacionados na formação da curva.

A Sensibilidade é uma medida onde um classificador prediz corretamente os exemplos da classe positiva (os verdadeiros positivos sobre o total de positivos) e a Especificidade é similar, mas em relação aos exemplos negativos (os verdadeiros negativos sobre o total de instâncias negativas).

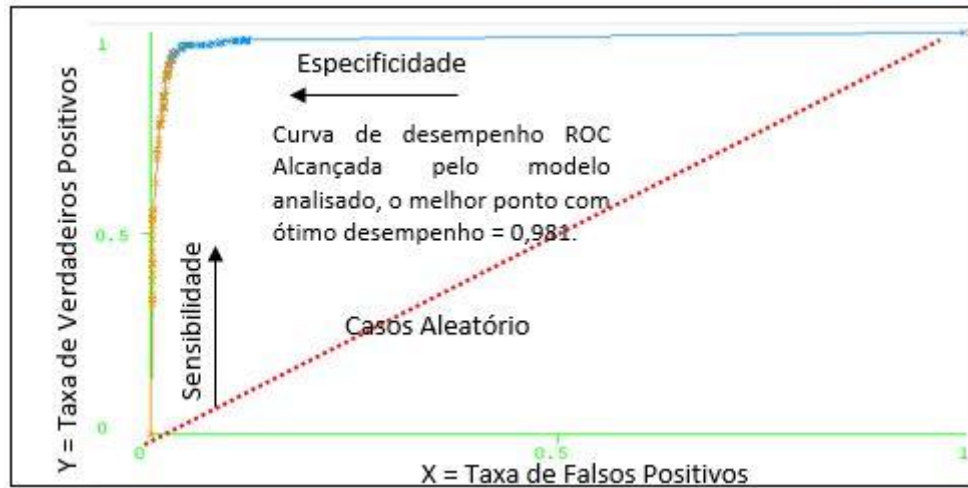
A sensibilidade é equivalente ao eixo “Y” do gráfico ROC e a especificidade é o complemento do eixo “X” representados nas Figuras 11, 12 e 13.

Figura 11 - Curva ROC da Fase A.



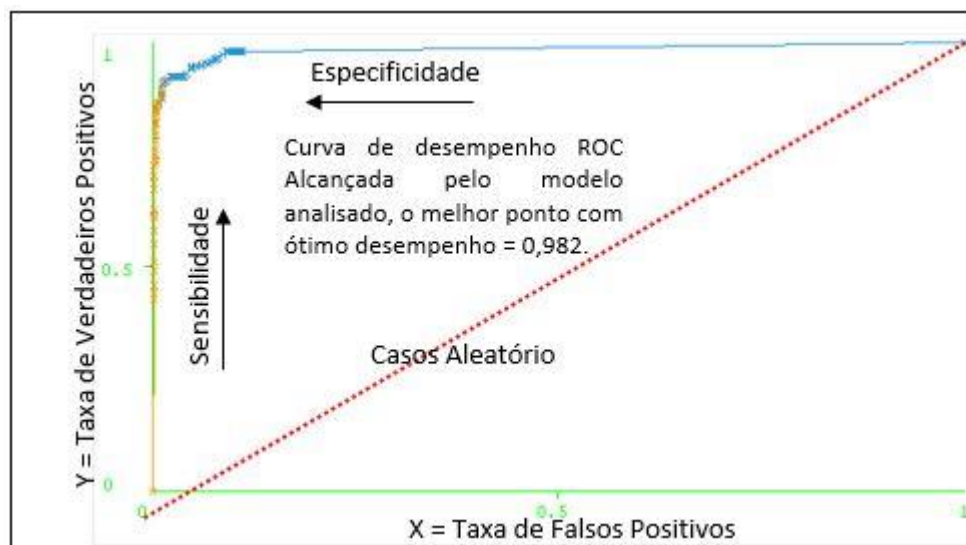
Fonte: O autor.

Figura 12 - Curva ROC da Fase B.



Fonte: O autor.

Figura 13 - Curva ROC da Fase C.



Fonte: O autor.

5.3 Considerações Finais

Neste capítulo foram realizadas as análises dos dados com a aplicação do processo de descoberta de conhecimento em bases de dados ou KDD, onde foram realizadas as fases de seleção, pré-processamento, transformação, mineração, interpretação e avaliação dos resultados obtidos no processo de descoberta, a técnica empregada na mineração, foi árvores de decisão que apresentou excelentes resultados com o aprendizado extraído dos dados coletados na indústria analisada.

O conhecimento extraído foi representado de acordo com cada fase analisada e os processos envolvidos. Na Fase A, foram detectadas que as harmônicas de 5º ordem do processo de tablets, a 7ª harmônica das centrais de ar, a 3ª harmônica do processo de teste de notebooks, a 3ª harmônica dos compressores influenciaram na contaminação da fase A, no entanto, a harmônica de maior contaminação foi a 7ª harmônica do processo de tablets que apresentou os maiores níveis de influência na geração do THD, pois sua influência foi durante a madrugada, quando haviam poucas cargas ligadas e com isso, pouca possibilidade de cancelamento harmônico. Os altos níveis de THD gerados no período da madrugada foram fortemente influenciados pelo processo de queima de tablets.

Na Fase B, foram detectadas que as harmônicas de 7ª ordem das centrais de ar condicionados, a 9ª ordem das centrais de ar condicionados, 3ª ordem dos compressores e 7ª ordem do teste de notebooks apresentavam níveis relevantes no processo da geração do THD, entretanto, a harmônica de maior representatividade foi a 7ª harmônica do processo de tablet que apresentou os maiores níveis de influência na geração do THD, pois sua influência foi durante a madrugada, quando haviam poucas cargas ligadas e com isso, pouca possibilidade de cancelamento harmônico. Os níveis de THD gerados no período da madrugada foram fortemente influenciados pelo processo de queima de tablets.

Na Fase C, Foram detectadas que as harmônicas de 3ª e 5ª ordem dos tablets, a de 3ª ordem das centrais de ar condicionados, a de 9ª ordem do processo de testes de notebooks, apresentaram níveis significativos, no entanto, a harmônica de 7ª ordem do processo de tablets foi a que apresentou os maiores índices de distorção harmônica.

CAPITULO VI

CONCLUSÕES, RECOMENDAÇÕES E CONTRIBUIÇÕES

6.1 Conclusões

De forma geral a aplicação da técnica de Árvore de decisão sobre os dados coletados em uma campanha de sete (7) dias (PRODIST), revelou que nas fases A, B e C foi detectada que em contribuição com o THD no PAC a presença da harmônica de 7º Ordem foi a apontada como mais significativa levando em consideração todo o período medido, gerada pelo processo de tablets. Este resultado se dá pelo fato de que ser o processo de tablets o único a ficar em funcionamento durante a noite e ter pouca carga em funcionamento, o que justifica a significância deste processo na geração de harmônicos em comparação a momentos de maior carga na indústria. Em outros períodos, harmônicos diferentes se apresentam com mais significância, porém nos 7 dias de medições contínuas, o período da madrugada, que tem pouca carga ganha destaque na metodologia utilizada, o que é de se esperar pela falta de cancelamento harmônico.

6.2 Contribuições do Trabalho

No presente estudo contempla-se as características, conceitos, fases e algoritmos utilizados no processo KDD, bem como as etapas de monitoramento e a avaliação da QEE fundamentos de acordo com PRODIST no módulo 8 Revisão 6, Normas IEC e IEEE, através dos indicadores de taxa de distorção harmônica total (THD) e as distorções harmônicas individuais de 3º, 5º, 7º e 9º ordens em um sistema trifásico de baixa tensão; Demonstrando as aplicações da inteligência computacional nas análises de fenômenos de diversas áreas do conhecimento, dando maior ênfase na análise realizada na indústria, onde foi montado um estudo de QEE e inteligência computacional com objetivo de permitir uma tomada de decisão no cenário atual, mediante os dados extraídos dos processos de manufatura da indústria.

Identificando as possíveis causas e mitigações dos impactos causados por distorções harmônicas dentro e fora do PIM.

Dentre as principais contribuições deste trabalho:

- Auxiliar na tomada de decisão e controle do conteúdo harmônico gerado na empresa, que podem causar penalizações por geração de harmônicos acima dos limites estabelecidos pela agência reguladora (ANEEL), evitando assim, possíveis penalizações com multas (Futuro).
- Esta pesquisa adiciona um conhecimento importante sobre a qualidade de energia elétrica e utilização de técnicas de inteligência computacional para quantificar e qualificar a qualidade da energia dentro do ambiente fabril, tendo como base as normas e recomendações nacionais e internacionais sobre os padrões de qualidade de energia elétrica.

6.3 Recomendações para trabalhos futuros

- Realizar novas análise nos dados utilizando outras técnicas de data mining, para verificar e comparar os resultados obtidos nas mais diversas técnicas, identificando as que apresentarem os melhores resultados;
- Realizar um estudo mais abrangente envolvendo outras industrias similares do PIM para comparar e avaliar a eficácia da técnica empregada;
- O trabalho desenvolvido pode servir como base para futuras análises do conteúdo harmônico presente no comércio e residência (Visando atendimento de legislação futura);

REFERÊNCIAS

- AMBRA et al., *Power-electronic solutions to power quality problems - Review Article*. Electric Power Systems Research, 2003.
- ANEEL. Módulo 8 – *Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional* – PRODIST. 2015.
- ANEEL. Nota Técnica, nº0083/2012-SRD/ANEEL, de 12 Junho 2012. Processo: 48500.002798/2012-61.
- ARRILLAGA, J. et al. *Power System Harmonic Analysis*. John Wiley & Sons. 1997.
- ARRILLAGA, J; WATSON, Neville R; *Power System Harmonics*. John Wiley & Sons. 2003.
- BARBOSA F. M.; “*Exploração de Sistema Elétricos de Energia*”, Faculdade de Engenharia da Universidade do Porto, 2006.
- BIKAS, A. K.; Voumvoulakis, E. M.; Hatziargyriou, N.D., *Intelligent System Applications to Power Systems*, ISAP '09. 15th International Conference, 2009.
- BEIRAGHI, M.; RANJBAR, A.M. *Online Voltage Security Assessment Based on Wide-Area Measurements*, Power Delivery IEEE Transactions on, 2013.
- BERNABEU et. al. *Methodology for a Security/Dependability Adaptive Protection Scheme Based on Data Mining*, Power Delivery IEEE Transactions on, 2012.
- BRACHMAN, R.; ANAND, T. *The process of knowledge discovery in databases: a human-centered approach*. In *Advances in Knowledge Discovery and Data Mining* (pp. 37–58). AAAI Press, 1996.
- BREIMAN, Friedman et al. *Classification and Regression Trees*, Califórnia: Wadsworth International, 1984.
- CARVALHO, L, A. V. de. *Datamining: a mineração de dados no Marketing, Medicina, Economia, Engenharia e Administração*. São Paulo: Ciência Moderna, 2006.
- CHENG-LIN et al. *The application of operation optimization decision support system based on data mining in power plant*, CHINA, 2005.
- CORNOIU et al. *Power quality monitoring and analysis. Case study for 220/110 kV substation*, Exploitation of Renewable Energy Sources (EXPRES), IEEE 3rd International Symposium on, 2011.

CUTSEM et al. *Decision trees for detecting emergency voltage conditions Proc. Int. Workshop on Bulk Power System Voltage Phenomena - Voltage Stability and Security* Deep Creek Lake, McHenry, Maryland, 1991.

DÁN, A. M. *Identification of Individual Harmonic Sources and Evaluation their Contribution in the Harmonic Distortion Level*. Power and Energy Society General Meeting 2009. Budapest – Hungary. 2009.

DING et al., **Comprehensive evaluation of power quality based on meaningful classification**. Critical Infrastructure (CRIS), 5th International Conference on, 2010.

DAS, J. C. *Power System Analysis: Short-Circuit Load Flow and Harmonics*. Marcel Dekker Inc. 2002.

DEHKORDI et al., *A decision tree-based method for power system fault diagnosis by synchronized Phasor Measurements Innovative Smart Grid Technologies (ISGT Europe)*, 3rd IEEE PES International Conference and Exhibition on, 2012.

DIAO, et. al., *Deriving optimal operational rules for mitigating inter-area oscillations*, Power Systems Conference and Exposition (PSCE), IEEE/PES, 2011.

DOUGHERTY, J. G.; STEBBINS, Wayne L.; **Power quality: a utility and industry perspective**, Textile Fiber, and Film Industry Technical Conference Annual, IEEE 1997.

DUGAN et al. *Electrical Power Systems Quality*. Second Edition. McGraw- Hill, 2004.

FILHO et. al., *Pricing analysis in the Brazilian energy market: A decision tree approach PowerTech*, IEEE Bucharest, 2009.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. *From data mining to knowledge discovery in databases. AI magazine*, 1996.

FUCHS, Ewald F., MASOUM, Mohammad A. S., *Power Quality in Power Systems and Electrical Machines*. Elsevier, 2008.

GENC et al., *Decision Tree-Based Preventive and Corrective Control Applications for Dynamic Security Enhancement in Power Systems*, IEEE Transactions on, 2010.

GENC et al., *Computation of transient stability related security regions and generation rescheduling based on decision trees*, Power and Energy Society General Meeting, IEEE, 2010.

GONÇALVES, Benevaldo Pereira, “*Metodologia para Diagnosticar a Qualidade de Energia Elétrica Referente à Distorção Harmônica em um sistema Trifásico de Baixa Tensão utilizando Lógica FUZZY*”. Dissertação de Mestrado do Programa de Pós-Graduação em Engenharia Elétrica do Instituto de Tecnologia da Universidade Federal do Pará (ITEC-UFPA). 2010.

GUNTHER, **Power quality monitoring**. Power Engineering Society Summer Meeting, IEEE, 1999.

HAN, J.; KAMBER, M; *Data Mining: Concepts and Techniques, Second Edition*. Morgan Kaufmann. Hoffer, J., Prescott, M., & McFadden, F. *Modern Database Management 6th ed.* Prentice Hall, 2002.

HATZIARGYRIOU et al., **Decision trees for fast security assessment of autonomous power systems with a large penetration from renewables**, Energy Conversion, IEEE Transactions on, 1995.

HIROSE et al., *Diagnosis of electric power apparatus using the decision tree method Dielectrics and Electrical Insulation*, IEEE Transactions on, 2008.

HIYONG LI; WEILIN WU, *Phasor Measurements-Aided Decision Trees for Power System Security Assessment*, Information and Computing Science, ICIC '09. Second International Conference on, 2009.

HOUBEN et al., *Coupling of K-NN with decision trees for power system transient stability assessment*. Control Applications, Proceedings of the 4th IEEE Conference on, 1995.

IMBERMAN, Susan, P. *Effective Use of the KDD Process and Data Mining For Computer Performance Professionals. Proceedings of CMG 2001*.

IEC 61000-3-2. IEC 61000-3-2 **Consolidated edition Electromagnetic compatibility (EMC) - Part 3-2: Limits – Limits for harmonic current emissions (equipment input current \leq 16A per phase)**, 2001.

IEC 61000-4-30. *International Electrotechnical Commission*, IEC 61000-4-30 Ed. 2.0: Electromagnetic Compatibility (EMC) – Part 4 – 30: Testing and Measurements Techniques – Power Quality Measurement Methods, 2008.

IEEE Std. 519-1992 IEEE Recommended Practices and Requirements for Harmonic Control in Electrical Power Systems, 2014.

JESUS, Nelson; OLIVEIRA, H. R. P. M. *Amplificação da Distorção Harmônica em Sistemas de Distribuição*. VII CBQEE – Congresso Brasileiro sobre Qualidade da Energia Elétrica. Santos – São Paulo. 2007.

JUNIOR, Jorge De Almeida Brito, “**Proposta de um Sistema de Diagnóstico da Qualidade de Energia Baseado em Logica Fuzzy: Estudo de Caso de uma Empresa do Polo Industrial de Manaus - PIM**”. Dissertação de Mestrado do Programa de Pós-Graduação em Engenharia Elétrica do Instituto de Tecnologia da Universidade Federal do Pará (ITEC-UFPA). 2011.

JUNLAKARN, S.; ILIC, M., **Toward implementation of the reconfiguration for providing differentiated reliability options in distribution systems**, PES General Meeting | Conference & Exposition, IEEE, 2014.

KAI et al., **An Online Dynamic Security Assessment Scheme Using Phasor Measurements and Decision Trees** Power Systems, IEEE Transactions on, 2007.

KARAPIDAKIS, E. S.; HATZIARGYRIOU, N. D. *Online preventive dynamic security of isolated power systems using decision trees*, Power Systems, IEEE Transactions on, 2002.

KATTI, Faceli et. al. *INTELIGÊNCIA ARTIFICIAL – Uma Abordagem de Aprendizado de Máquina*, Rio de Janeiro: LTC, 2001.

KAZEROONI, M. et al.; **Literature review on the applications of data mining in power systems**, Power and Energy Conference at Illinois (PECI), 2014.

KHOSHKHOO et al, **On-line dynamic voltage instability prediction based on decision tree supported by a wide-area measurement system** Generation, Transmission & Distribution, IET, 2012.

KRISHNAN et. al, **Efficient Database Generation for Decision Tree Based Power System Security Assessment**, Power Systems IEEE Transactions on, 2011.

KUMARSWAMY et. al, **Analysis of Islanding Detection in Distributed Generation Using Fuzzy Logic Technique**. Modelling Symposium (AMS), 7th Asia, 2013.

LAZCORRETA, E. et al., **Towards Personalised Recommendation by Two-Step Modified Apriori Data Mining Algorithm**. Expert Systems with Applications, 35(3), 1422-1429.

LI, Chun et al.; **“Critical Impedance” –Based Method for Method for Identifying Harmonic Sources**. IEEE Transaction on Power Delivery. Vol 19, Nº 2, April. 2004.

LIU et al. *A self learning expert system for voltage control of power systems* Proc. 2nd Symp. on Expert Systems Application to Power Systems, Seattle, 1989.

LEITE, Jandecy Cabral, **“Projeto Multicritério de Filtros Harmônicos Passivos Para Instalações Industriais Utilizando Técnicas de Inteligência Computacional”**. Tese de Doutorado do Programa de Pós-Graduação em Engenharia Elétrica do Instituto de Tecnologia da Universidade Federal do Pará (ITEC-UFPA). 2013.

LEITE et al., *Planning passive filters using NSGA II for industry applications*, Industry Applications (INDUSCON), 11th IEEE/IAS International Conference, 2014.

LEONIDAKI et. al, **Decision trees for determination of optimal location and rate of series compensation to increase power system loading margin** Power Systems, IEEE Transactions on, 2006.

LOBATO et al. *Decision Trees Applied to Spanish Power Systems Applications, Probabilistic Methods Applied to Power Systems*, PMAPS. International Conference on, 2006.

MA et. al, *Building decision trees for characteristic ellipsoid method to monitor power system transient behaviors* Power and Energy Society General Meeting, IEEE, 2010.

MADAN, S. et al.; *Applications of data mining for power systems, Electrical and Computer Engineering*, Engineering Innovation: Voyage of Discovery. IEEE 1997, CANADA, 1997.

MAHDI, I.; NADJI, B. *Application of the binary decision diagram (BDD) in the analysis of the reliability of the inverters Power Engineering*, Energy and Electrical Drives (POWERENG), Fourth International Conference on, 2013.

MAZIN, Hooman et al.; *Determining the Harmonic Impact of Multiple Harmonic-Producing Loads*. IEEE Transaction on Power Delivery. Vol 26, NO. 2, April. 2011.

M.E.L. Tostes, “**Avaliação dos impactos causados pela geração de harmônicos na rede de distribuição em Consumidores em baixa tensão**”. Tese de Doutorado do Programa de Pós-Graduação em Engenharia Elétrica do Instituto de Tecnologia da Universidade Federal do Pará (ITEC-UFGPA). 2003.

MIAO at. al, *A data mining framework for online dynamic security assessment: Decision trees, boosting, and complexity analysis*, Innovative Smart Grid Technologies (ISGT), IEEE PES, 2012.

MYO T. A.; MILANOVIC, J. V., *Stochastic prediction of voltage sags by considering the probability of the failure of the protection system*, Power Delivery, IEEE Transactions on, 2006.

MOUTIS, P.; HATZIARGYRIOU, N. D., *Decision Trees-Aided Active Power Reduction of a VirtualPower Plant for Power System Over-Frequency Mitigation Industrial Informatics*, IEEE Transactions on, 2015.

MORI et al., *A data mining technique for three-phase distribution network voltage control* Transmission and Distribution Conference and Exposition, T&D. IEEE/PES, 2008.

MUKHERJEE, S.; WOJSZCZYK, B., *Screening of load patterns and transmission planning alternatives using decision trees*, Bulk Power System Dynamics and Control - VII. Revitalizing Operational Reliability, iREP Symposium, 2007.

MCNABB et. al., *Classification of mode damping and amplitude in power systems using synchrophasor measurements and classification trees Power Systems*, IEEE Transactions on, 2013.

NASCIMENTO, et al. *Harmonic identification using parallel neural networks in single-phase systems*, Applied Soft Computing 11 (2) 2178–2185, 2011.

NANTIAN et al., *Power quality disturbance recognition using transforms and FCM-based decision tree*, Intelligent Computing and Intelligent Systems, ICIS 2009. IEEE International Conference on, 2009.

NEVES, R. de C. D. das. *Pré-processamento no processo de descoberta de conhecimento em banco de dados*. 2003. Dissertação (Programa de Pós-graduação em Computação) - Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.

NOGUEIRA et al. *Análisis de Los Impactos Armónicos en la Industria de La Electrónica Utilizando Árboles de Decisión*, 17 Convención Científica de Ingeniería y Arquitectura, La Habana-CUBA, 2014.

NUNES Santos, I.; Oliveira, J.C., *Critical Analysis of the Current and Voltage Superposition Approaches at Sharing Harmonic Distortion Responsibility*, Latin America Transactions, IEEE (Revista IEEE America Latina), 2011.

OZGONENEL, O. et al.; *Superiority of decision tree classifier on complicated cases for power system protection*, Developments in Power Systems Protection, 2012. DPSP. 11th International Conference, 2012.

PAENSUWAN, N.; YOKOYAMA, A. *Probabilistic Dynamic TTC Calculation with Decision Tree Classification Intelligent System Applications to Power Systems*, ISAP '09. 15th International Conference on, 2009.

PIATETSKY-SHAPIRO, G.; FRAWLEY, W. *Knowledge Discovery in Databases*. Menlo Park, California: AAAI Press, 1991.

PERERA, N.; RAJAPAKSE, A.D., *Power system transient classification for protection relaying, Harmonics and Quality of Power, ICHQP 2008*. 13th International Conference, 2008.

PIRTTIMÄKI, V. *The Roles of Internal and External Information in Business Intelligence*. Frontiers of E-Business Research, 2004.

POSTGRESQL: *The world's most advanced open source The PostgreSQL Global Development Group is pleased to announce the availability of PostgreSQL 9.4 Beta 2*, the second beta release of the upcoming 9.4. Disponível em: < www.postgresql.org/ > Acesso em: 07 Julho 2014.

QUINLAN, J R “*Learning efficient classification procedures and their application to chess end games, in Machine Learning: An Artificial Intelligence Approach*”, R S Michalski, J G Carbonell and T M Mitchel (Editors) Springer, Berlin, 1984.

QUINLAN, J. R. C4.5: *Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

RATHER et al. *Dynamic security assessment of Danish power system based on decision trees: Today and tomorrow*, PowerTech (POWERTECH), 2013 IEEE Grenoble, 2013.

RAPIDMINER *provides software, solutions, and services in the field of advanced analytics, including predictive analytics, data mining, and text mining*. URL: <http://http://rapidminer.com/>, 2015.

RODRIGUEZ, A. *Decision Tree and S-transform based approach for power quality disturbances classification, Power Engineering*, Energy and Electrical Drives (POWERENG), Fourth International Conference, 2013.

RAY, P. K.; *Optimal Feature and Decision Tree-Based Classification of Power Quality Disturbances in Distributed Generation Systems*, Sustainable Energy, IEEE Transactions, 2014.

SANTANA, Ádamo Lima de, “*Projeto de Implementação de um Sistema de Suporte à Tomada de Decisão para o Observatório de Saúde da Amazônia*”. Dissertação de Mestrado do Programa de Pós-Graduação em Engenharia Elétrica – Instituto de Tecnologia da Universidade Federal do Pará (ITEC-UFPA). 2005.

SANTOS, Ivan Nunes; *Uma Proposta de Metodologia para Atribuição de Responsabilidade sobre as Distorções Harmônicas Utilizando o Princípio da Superposição*. XXI SNPTEE – Seminário Nacional de Produção e Transmissão de Energia Elétrica. Florianópolis – Santa Catarina. 2011.

SANTOS, I. N., OLIVEIRA, J. C.; *Avaliação Laboratorial de uma Nova Proposta Metodológica para Atribuição de Responsabilidades sobre as Distorções Harmônicas*. The 9th Latin-American on Electricity Generation and Transmission – CLAGTEE 2011. Mar Del Plata – Argentina. 2011.

SHU-HSIEN et al. *Data mining techniques and applications* – A decade review from 2000 to 2011 Review Article Expert Systems with Applications, Volume 39, Issue 12, 15 September 2012.

SABOYA et. al. *Start-Up Decision of a Rapid-Start Unit for AGC Based on Machine Learning*, Power Systems, IEEE Transactions on, 2013.

SAI et. al., *Towards intelligent decision making in power plant operation*, India Conference (INDICON) Annual IEEE, 2014.

SHAHRTASH, S. M.; SARLAK, M., *High Impedance Fault Detection Using Harmonics Energy Decision Tree Algorithm*, Power System Technology PowerCon. International Conference, 2006.

SAMANTARAY, S.R., *Decision tree-initialised fuzzy rule-based approach for power quality events classification*, Generation, Transmission & Distribution, IET, 2010.

SAEH, I. S; KHAIRUDDIN, A. *Decision Tree for Static Security Assessment Classification*, Future Computer and Communication, 2009. ICFCC. International Conference on, 2009.

SHENGYONG et. al. *Power System Transient Stability Assessment Based on Adaboost and Support Vector Machines*, Power and Energy Engineering Conference (APPEEC), Asia-Pacific, 2012.

SILVA et al., *Naive Bayes Aplicados na Análise de Impactos Harmônicos em Sistemas Elétricos Industriais*, XI IEEE/IAS International Conference on Industry Applications – INDUSCON, Brasil, 2014.

SWARUP et. al. *Decision tree for steady state security assessment and evaluation of power systems*, Intelligent Sensing and Information Processing, Proceedings of International Conference on, 2005.

TAN et al., **An overview of power quality state estimation**, Power Engineering Conference, IPEC. The 7th International, 2005.

TARASIUK, T; *Estimator-analyzer of power quality. Part I. Methods and algorithms, Measurement* 44 (1) (2011) 238–247. UIE, Types of Disturbances and Relevant Standards, Guide to Quality of Electrical Supply for Industrial Installations, Power Quality Working Group WG2, 1996.

TEEUWSEN et al., *Decision tree based oscillatory stability assessment for large interconnected power systems*, Power Systems Conference and Exposition, IEEE PES, 2004.

TIAKO et. al., *A class of intelligent algorithms for on-line dynamic security assessment of power systems* Universities Power Engineering Conference (AUPEC), 20th Australasian, 2010.

TIAKO et. al. *An artificial intelligence approach for on-line dynamic security assessment of wind farm connected active*, Power Systems Innovative Smart Grid Technologies Asia (ISGT), IEEE PES, 2011.

TSO, S. K.; *Data mining for detection of sensitive buses and influential buses in a power system subjected to disturbances*, Power Systems, IEEE Transactions on, CHINA, 2004.

UNIVERSITY OF WAIKATO. Weka 3 – *Machine Learning Software in Java*. Disponível no site da University of Waikato. URL: <http://www.cs.waikato.ac.nz/ml/weka>, 2010.

VAID, K. et al. *Critical impedance based automatic identification of harmonic sources in deregulated power industry*, *Signal Processing, Communication, Computing and Networking Technologies (ICSCCN)*, International Conference, 2011.

VOUMVOULAKIS, E.M.; HATZIARGYRIOU, N.D. *Decision Trees-Aided Self-Organized Maps for Corrective Dynamic Security* Power Systems, IEEE Transactions on, 2008.

VLAHINIC S., et al, *Indices for harmonic distortion monitoring of power distribution systems*, IEEE Transactions on Instrumentation and Measurement 58 (5) 1771–1777, 2009.

WEHENKEH et al. *Decision trees applied to on-line transient stability assessment of electric power systems*, *Procs. IEEE Int. Symp. on Circuits and Systems* Vol 2 Helsinki, Finland, 1988.

WITTEN, I.; FRANK, E.; *Data Mining Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann, 2005.

XIA at. al, *Application of intelligent algorithm in island detection of distributed generation*, Transmission and Distribution Conference and Exposition, IEEE PES, 2010.

XIAOFENG et. al, *A C4.5-based research on islanding detection in distributed generation, system Advances in Power System Control, Operation and Management* (APSCOM 2009), 8th International Conference on, 2009.

XU, Wilsun.; LIU, Y; *A method for determining customer and utility harmonic contributions at the point of common coupling*, Power Delivery, IEEE Transactions on, CANADA, 2000.

XU, Wilsun et al. *A method to determine the harmonic contributions of multiple loads*, Power & Energy Society General Meeting, PES '09. IEEE, 2009.

XU, Wilsun, LIU, Yilu. *A Method to Determine Customer Harmonic Contributions for Incentive-Based Harmonic Control Applications*. 1999 IEEE PES Summer Meeting: Panel on Harmonic Measurements and Allocation. 1999.

XU, Wilsun et al, *An Investigation on the Validity of Power-Direction Method for Harmonic Source Determination*. IEEE Transaction on Power Delivery. Vol 18, N° 1, April. 2003.

YEN-TSUNG Pan; MEN-SHEN Tsai, *Development a BDI-Based Intelligent Agent Architecture for Distribution Systems Restoration Planning* Intelligent System Applications to Power Systems, ISAP '09. 15th International Conference on, 2009.

YIQIANG et. al. *Power-efficient access-point selection for indoor location estimation Knowledge and Data Engineering*, IEEE Transactions on, 2006.