

Evento: XXII Jornada de Pesquisa

**BUSCA DE PADRÕES UTILIZANDO O ALGORITMO APRIORI PARA
MINERAÇÃO DE DADOS NAS SUBESTAÇÕES SUBTERRÂNEAS DA CEEE.¹
PATTERN SEARCH USING THE APRIORI ALGORITHM FOR DATA
MINING IN CEEE UNDERGROUND SUBSTATIONS**

**Luciano Antônio Destefani², Leonardo Bressan Motyczka³, Paulo Sausen⁴,
Airam Sausen⁵**

¹ Este artigo está inserido na linha de pesquisa Modelagem Computacional, Otimização e Controle de Sistemas do Programa de Pós-Graduação Stricto Sensu em Modelagem Matemática, UNIJUI, Ijuí, RS.

² Aluno do Curso de Mestrado em Modelagem Matemática da UNIJUI, bolsista UNIJUI, lucianodestefani@gmail.com

³ Aluno do Curso de Mestrado em Modelagem Matemática da UNIJUI, bolsista CAPES, leomoty@gmail.com

⁴ Professor Doutor do Departamento de Ciências Exatas em Engenharias, Orientador, paulosausen@gmail.com

⁵ Professora Doutora do Departamento de Ciências Exatas em Engenharias, Co-orientadora, airamsausen@gmail.com

Resumo: Inicialmente é apresentada uma revisão bibliográfica sobre as técnicas de MD encontrados na literatura, com ênfase no algoritmo Apriori e na ferramenta WEKA utilizados na MD para busca dos padrões em estudo. Destaca-se que a partir da instrumentação das redes de distribuição das concessionárias de energia tem sido disponibilizado uma base de dados com um volume considerável, tanto em tamanho, quanto em informações, que podem propiciar além da caracterização dos perfis de consumo, uma nova gama de aplicações relacionadas a descoberta de conhecimento que possam auxiliar na tomada de decisões administrativas vinculadas, por exemplo, a manutenção preventiva de sua estrutura de distribuição. A partir da parceria existente entre o Grupo de Automação Industrial e Controle (GAIC), da Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUI), com a Companhia Estadual de Distribuição de Energia Elétrica (CEEE), concessionária de serviços de distribuição de energia elétrica na região sul-sudeste do Estado do Rio Grande do Sul, localizada na cidade de Porto Alegre, possibilitou o acesso aos dados compostos das aquisições de 30 subestações de energia do mês de janeiro de 2017, aquela com maior número de pontos foi escolhida para rodar o pré-processamento e o processo de Mineração de Dados. Neste contexto este trabalho tem por objetivo a busca de padrões através dos conceitos de Smart Grids, Descoberta de Conhecimento e da Mineração de Dados (MD) utilizando o algoritmo Apriori e a ferramenta WEKA nas subestações subterrâneas da CEEE.

Abstract: It is initially shown a bibliographic review regarding the data mining techniques found in literature, tailored towards the Apriori algorithm and the WEKA toolkit for pattern search. By leveraging the instrumentalization of the power distribution networks, the power companies have been gathering a considerable amount of data, that could allow besides the characterization of the consumption profiles, a new plethora of applications related to knowledge discovery that can aid

Evento: XXII Jornada de Pesquisa

the administrative decision making, such as, preventive maintenance of the distribution system. A partnership between the Industrial Automation and Control Group (GAIC), from the Regional University of the Northeast of the State of Rio Grande do Sul (UNIJUI), and the State Power Distribution Company (CEEE), based in Porto Alegre/RS, made possible the access to data comprised of the acquisitions for 30 power substations from the month of January of 2017, the one with the most amount of data points was chosen to run through the preprocess step and data mined. The goal of this paper is to find patterns by leveraging the Smart Grid, Knowledge Discovery and Data Mining concepts utilizing the Apriori algorithm and the WEKA toolkit in the CEEE's power substations.

Palavras-chave: Descoberta de Conhecimento, Mineração de Dados, Energia Elétrica, Apriori, WEKA.

Keywords: Knowledge Discovery in Databases, Data Mining, Electric Energy, Apriori, WEKA.

1 INTRODUÇÃO

O crescente avanço tecnológico atualmente proporciona a utilização de ferramentas e técnicas para extrair informações específicas para resolução de problemas nas mais variadas áreas de atividade. Se em outros tempos havia uma limitação na capacidade física de processamento, hoje a evolução da tecnologia da informação possibilita que grandes volumes de dados possam ser armazenados e processados mais rapidamente e com uma maior precisão (SCHUCH, 2010).

Com este volume de dados armazenados crescendo diariamente, responder uma questão tornou-se crucial: O que fazer com os dados armazenados? As técnicas tradicionais de exploração de dados não são mais adequadas para tratar a grande maioria dos repositórios (CAMILO, 2009). Com a finalidade de responder a esta questão, foi proposta, no final da década de 80, a Mineração de Dados (MD), do inglês Data Mining.

Historicamente já se passaram mais de um século do início da utilização da energia elétrica por unidades consumidoras residenciais, comerciais e industriais, porém, a maneira como é feita a distribuição de energia elétrica evoluiu muito pouco, pois os sistemas atuais dependem excessivamente de fontes geradoras únicas, com redes suscetíveis a falhas pontuais, porém críticas, e alto índice de furtos e prejuízos nas unidades consumidoras (DUARTE, 2014).

Neste contexto as Redes Inteligentes (Smart Grids) vêm evoluindo gradativamente com o objetivo de integrar o comportamento e ações de todos os usuários conectados a ela, sejam eles geradores, consumidores ou ambos. Devido a sua importância tanto para as unidades geradoras quanto para os consumidores finais, faz-se necessário a busca de novas técnicas e algoritmos para minimizar as perdas, prever falhas e dar sustentabilidade aos serviços oferecidos. A descoberta de conhecimento vem ao encontro desta necessidade e pode auxiliar na identificação e resolução de diversos problemas na geração e distribuição da energia elétrica com a aplicação de MD e das técnicas corretas.

Evento: XXII Jornada de Pesquisa

Destaca-se que as Smart Grids tradicionalmente vem utilizando as técnicas de MD para tratar problemas relacionados ao consumidor final (CHOI, 2015; JIA, 2012; KITAYAMA, 2002; ROCHA, 2016; ZHANG, 2016), em contrapartida os modelos computacionais estudados/aplicados neste trabalho têm como objetivo auxiliar as empresas concessionárias de energia elétrica na resolução de diversos problemas enfrentados diariamente que vão desde a prevenção de falhas, identificação de fraudes, classificação dos clientes e análise de consumo anormal que são realizados por algoritmos de agrupamento, de análise de correlação e pontos fora do padrão (outlier), entre outros (JIA, 2012).

A acurácia dos dados extraídos vai depender de diversos fatores, dentre eles os algoritmos aplicados, sendo que, para cada tipo de dado em estudo é possível aplicar diferentes tipos de algoritmos e técnicas de mineração, encontrando assim, a relação entre seus elementos correlativos. Desta forma, os resultados da análise podem auxiliar tanto os consumidores de energia, por exemplo, na utilização de eletricidade de forma racional, e também fornecer apoio a decisões estratégicas para as concessionárias de energia elétrica que é o foco principal desta pesquisa (JIA, 2012).

Neste contexto o principal objetivo deste trabalho é utilizar os conceitos das Smart Grids, descoberta de conhecimento e mineração de dados para busca de padrões utilizando as regras de associação e o algoritmo Apriori juntamente da ferramenta WEKA visando a melhoria permanente dos serviços disponibilizados pelas concessionárias de energia elétrica.

2 REVISÃO BIBLIOGRAFICA

Com o intuito de facilitar a leitura e compreensão deste trabalho, nesta seção são apresentadas informações elementares sobre a Descoberta de Conhecimento e MD, suas características relevantes e principais propriedades. É realizada uma revisão bibliográfica do estado da arte sobre as técnicas de MD encontrados na literatura, com ênfase no algoritmo Apriori e na ferramenta WEKA utilizada na MD para busca dos padrões em estudo.

2.1 MINERAÇÃO DE DADOS

O conceito de Mineração de Dados (Data Mining) está se tornando cada vez mais popular como uma ferramenta de descoberta de informações com objetivo de revelar estruturas de conhecimento visando a tomada de decisões em condições de certeza limitada. Recentemente tem havido um interesse crescente em desenvolver novas técnicas de análise de dados, especialmente projetadas para tratar questões relativas a mineração de dados que ainda está baseada em princípios conceituais de Análise de Dados Exploratórios (Exploratory Data Analysis - EDA) e de modelagem (CORTES, 2012).

Um conceito muito difundido e equivocado sobre mineração de dados define os sistemas de MD como fontes para extração automática de todos os conceitos mais valiosos que estão escondidos em um grande banco de dados sem intervenção ou direcionamento humano (CORTES, 2012; KAMBER, 2012).

Evento: XXII Jornada de Pesquisa

Para confrontar esta ideia os principais conceitos aceito por muitos pesquisadores referem-se a **descoberta de conhecimento** em bases de dados como um processo global de descoberta que envolve seleção, pré-processamento, mineração dos dados, interpretação dos resultados e a transformação do conhecimento. Sendo assim, a **MD é uma das etapas deste processo** onde são aplicados algoritmos específicos para extração de padrões a partir dos dados ou até mesmo revelar comportamento de um banco de dados (SCHUCH, 2010).

Em termos gerais a mineração de dados também pode ser descrita como uma área de pesquisa multidisciplinar que engloba diversas outras áreas como: Inteligência Artificial; Aprendizado de Máquina; Redes Neurais; Estatística; dentre outras, sendo que esta só se tornou possível a partir do amadurecimento destas diferentes áreas (KAMBER, 2012; SCHUCH, 2010).

Para melhor entendimento a MD será explicada e dividida em cinco subseções relacionadas respectivamente aos dados, as tarefas, aos métodos, ferramentas para MD e por fim o algoritmo Apriori.

2.1.1 OS DADOS

Conhecer o tipo dos dados com o qual se irá trabalhar, também é fundamental para a escolha do(s) método(s) mais adequado(s). Pode-se categorizar os dados em dois tipos: quantitativos e qualitativos. Os dados quantitativos são representados por valores numéricos. Eles ainda podem ser discretos e contínuos. Já os dados qualitativos contêm os valores nominais e ordinais (categóricos). Em geral, antes de se aplicar os algoritmos de mineração é necessário explorar, conhecer e preparar os dados (CAMILO, 2009).

O processo de preparação dos dados para a mineração, também chamado de pré-processamento (CAMILO, 2009; KAMBER, 2012), consiste principalmente em quatro etapas (limpeza dos dados, integração dos dados, transformação dos dados e redução dos dados). Nesse sentido, uma das primeiras atividades é obter uma visualização dos dados, de forma que se possa ter uma visão geral, para depois decidir-se quais as técnicas mais indicadas.

Na primeira fase é realizada a limpeza dos dados, pois frequentemente os dados são encontrados com diversas inconsistências: registros incompletos, valores errados e dados inconsistentes. A etapa de limpeza dos dados visa eliminar estes problemas de modo que eles não influam no resultado dos algoritmos usados. As técnicas usadas nesta etapa vão desde a remoção do registro com problemas, passando pela atribuição de valores padrões, até a aplicação de técnicas de agrupamento para auxiliar na descoberta dos melhores valores.

Na segunda fase é realizada a integração dos dados e nesta etapa é comum obter os dados a serem minerados de diversas fontes: banco de dados, arquivos textos, planilhas, data warehouses, vídeos, imagens, entre outras. Surge então, a necessidade da integração destes dados de forma a termos um repositório único e consistente. Para isto, é necessária uma análise aprofundada dos dados observando redundâncias, dependências entre as variáveis e valores conflitantes (categorias diferentes para os mesmos valores, chaves divergentes, regras diferentes para os

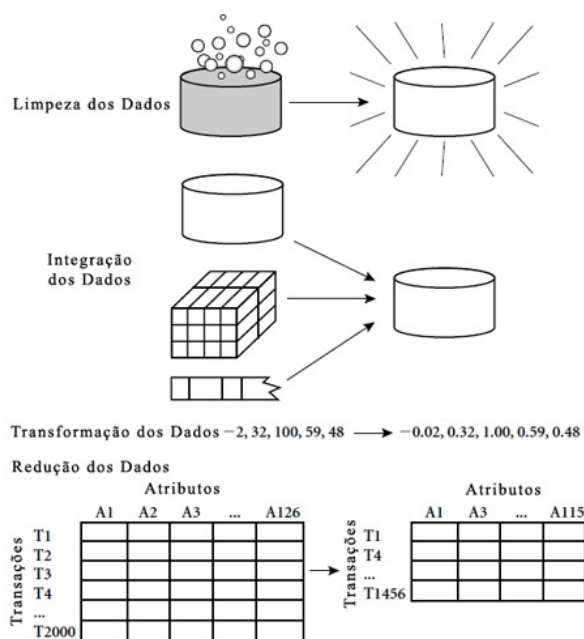
Evento: XXII Jornada de Pesquisa

mesmos dados, entre outros).

Já a terceira fase de transformação dos dados merece destaque e é importante ressaltar que alguns algoritmos trabalham apenas com valores numéricos e outros somente com os valores categóricos, sendo nestes casos necessário transformar os valores numéricos em categóricos ou os categóricos em valores numéricos. Não existe um critério único para transformação dos dados e diversas técnicas podem ser usadas de acordo com os objetivos pretendidos. Algumas das técnicas empregadas nesta etapa são: suavização (remove valores errados dos dados), agrupamento (agrupa valores em faixas sumarizadas), generalização (converte valores muito específicos para valores mais genéricos), normalização (colocar as variáveis em uma mesma escala) e a criação de novos atributos (gerados a partir de outros já existentes).

A última fase é responsável pela redução dos dados pois o volume de dados usado na mineração costuma ser alto. Em alguns casos, este volume é tão grande que torna o processo de análise dos dados e da própria mineração impraticável. Nestes casos, as técnicas de redução de dados podem ser aplicadas para que a massa de dados original seja convertida em uma massa de dados menor, porém, sem perder a representatividade dos dados originais. Isto permite que os algoritmos de mineração sejam executados com mais eficiência, mantendo a qualidade do resultado. As estratégias adotadas nesta etapa são: criação de estruturas otimizadas para os dados (cubos de dados), seleção de um subconjunto dos atributos, redução da dimensionalidade e discretização. A figura 1 ilustra todas as etapas deste processo.

Figura 1: Atividade do Pré-Processamento



Fonte: Kamber (2012).

Evento: XXII Jornada de Pesquisa

Geralmente os repositórios usados possuem milhares de registros. Nestes casos o uso de todos estes registros para a construção do modelo de Mineração de Dados é inviável (CAMILO, 2009). Assim, utiliza-se uma amostra (mais representativa possível) que é dividida em três conjuntos:

- 1) Conjunto de Treinamento (Training Set): conjunto de registros usados no qual o modelo é desenvolvido;
- 2) Conjunto de Testes (Test Set): conjunto de registros usados para testar o modelo construído;
- 3) Conjunto de Validação (Validation Set): conjunto de registros usados para validar o modelo construído;

Essa divisão em grupos é necessária para que o modelo não fique dependente de um conjunto de dados específico e, ao ser submetido a outros conjuntos (com valores diferentes dos usados na construção e validação do modelo), apresente resultados insatisfatórios. A medida que se aumenta a precisão do modelo para um conjunto de dados específico, perde-se a precisão para outros conjuntos (CAMILO, 2009).

2.2.2 TAREFAS

A MD é comumente classificada pela sua capacidade em realizar determinadas tarefas, sendo as mais comuns (CAMILO, 2009; LAROSE, 2014):

Descrição (Description): geralmente oferece uma possível interpretação para os resultados obtidos, sendo muito utilizada em conjunto com as técnicas de análise exploratória de dados, para comprovar a influência de certas variáveis no resultado obtido.

Classificação (Classification): é uma das tarefas mais comuns e visa identificar a qual classe um determinado registro pertence. Nesta tarefa, o modelo analisa o conjunto de registros fornecidos, com cada registro já contendo a indicação à qual classe pertence, a fim de 'aprender' como classificar um novo registro (aprendizado supervisionado).

Estimação (Estimation) ou Regressão (Regression): a estimação é similar à classificação, porém é usada quando o registro é identificado por um valor numérico e não um categórico. Sendo assim, pode-se estimar o valor de uma determinada variável analisando-se os valores das demais.

Predição (Prediction): esta tarefa é similar às tarefas de classificação e estimação, porém ela visa descobrir o valor futuro de um determinado atributo.

Agrupamento (Clustering): visa identificar e aproximar os registros similares. Um agrupamento (ou cluster) é uma coleção de registros similares entre si, porém diferentes dos outros registros nos demais agrupamentos. As aplicações das tarefas de agrupamento são as mais variadas possíveis: pesquisa de mercado, reconhecimento de padrões, processamento de imagens, análise

Evento: XXII Jornada de Pesquisa

de dados, segmentação de mercado, taxonomia de plantas e animais, pesquisas geográficas, classificação de documentos da Web, detecção de comportamentos atípicos (fraudes), entre outras.

Associação (Association): a tarefa de associação consiste em identificar quais atributos estão relacionados. Apresentam a forma: SE atributo X ENTÃO atributo Y. É uma das tarefas mais conhecidas devido aos bons resultados obtidos, principalmente nas análises da "Cestas de Compras" (Market Basket), onde identificamos quais produtos são levados juntos pelos consumidores.

2.2.3 MÉTODOS OU TÉCNICAS

A escolha das técnicas de MD dependerá da tarefa específica a ser executada e dos dados disponíveis para análise, devendo-se levar em conta a natureza dos dados disponíveis em termos de conteúdo, os tipos de campos de dados e a estrutura das relações entre os registros (SCHUCH, 2010).

Tradicionalmente, os métodos de mineração de dados são divididos em aprendizado supervisionado (preditivo) e não-supervisionado (descritivo) (CAMILO, 2009).

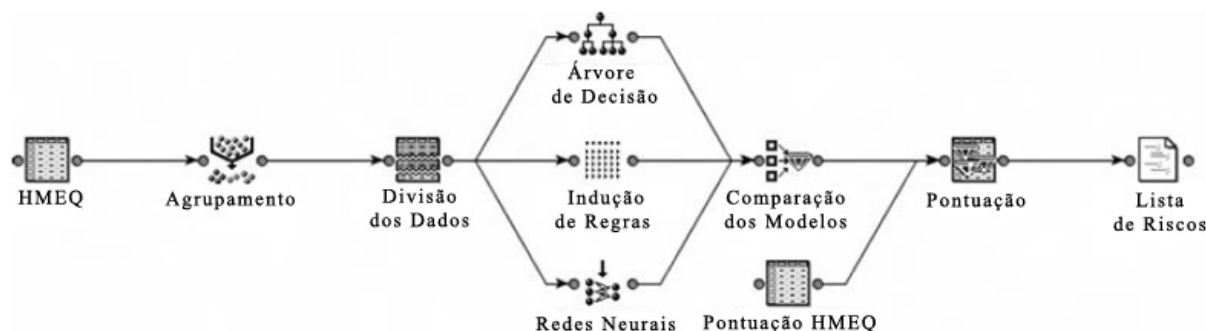
A diferença entre os métodos de aprendizado supervisionados e não-supervisionados reside no fato de que os métodos não-supervisionados não precisam de uma pré-categorização para os registros, ou seja, não é necessário um atributo alvo (CAMILO, 2009; CIOU, 2007; FAYYAD, 2006; KAMBER, 2012). Tais métodos geralmente usam alguma medida de similaridade entre os atributos (CAMILO, 2009; MCCUE, 2014).

As tarefas de agrupamento e associação são consideradas como não-supervisionadas. Já no aprendizado supervisionado, os métodos são providos com um conjunto de dados que possuem uma variável alvo pré-definida e os registros são categorizados em relação a ela. As tarefas mais comuns de aprendizado supervisionado são a classificação (que também pode ser não-supervisionado) e a regressão (CAMILO, 2009; MCCUE, 2014).

Durante o processo de mineração, diversas técnicas devem ser testadas e combinadas a fim de que comparações possam ser feitas e então a melhor técnica (ou combinação de técnicas) seja utilizada (CAMILO, 2009; MCCUE, 2014). Na figura 2 podemos ver um exemplo de combinação dessas técnicas.

Evento: XXII Jornada de Pesquisa

Figura 2: Processo de comparação com algumas técnicas.



Fonte: Mccue (2014).

2.2.4 FERRAMENTAS PARA MINERAÇÃO DE DADOS

Dentre as ferramentas pesquisadas para a descoberta de conhecimento e MD, existem opções gratuitas e outras pagas, sendo que dentre as ferramentas *Free* o WEKA é amplamente referenciado como um software reconhecido pela comunidade científica em aprendizado de máquina e MD (CAMILO, 2009; MAIA, 2015; WITTEN, 2005).

O WEKA (Waikato Environment for Knowledge Analysis) desenvolvido na Universidade de Waikato, Nova Zelândia, trata-se de um pacote implementado na linguagem Java, seguindo o paradigma de orientação a objetos, composto de uma série de algoritmos para solucionar problemas de MD. Os algoritmos podem ser aplicados diretamente a uma série de dados pela ferramenta ou utilizados por programas Java. Tem como principal característica ser portátil, permitindo trabalhar nas mais variadas plataformas e aproveitando todos os benefícios da orientação a objetos. Por ser uma ferramenta de domínio público (software livre) encontra-se disponível em <http://www.cs.waikato.ac.nz/ml/weka/> (SCHUCH, 2010).

Algumas destas ferramentas possuem algoritmos de regressão, classificação, agrupamento, regras de associação e seleção de parâmetros (atributos) (CAMILO, 2009; MAIA, 2015). Sendo que, a escolha das técnicas de MD dependerá da tarefa específica a ser executada e dos dados disponíveis para análise, devendo-se levar em conta a natureza dos dados disponíveis em termos de conteúdo, os tipos de campos de dados e a estrutura das relações entre os registros (CASTANHEIRA, 2008; GOES, 2010; MAIA, 2015; SCHUCH, 2010).

2.2.5 ALGORITMO APRIORI

O algoritmo Apriori realiza a mineração de dados em dois passos: geração e poda. No primeiro passo, o algoritmo executa uma varredura no arquivo, com o propósito de criar os conjuntos de

Evento: XXII Jornada de Pesquisa

combinações entre valores dos atributos existentes e editados no arquivo. No segundo passo, o algoritmo leva em consideração os conjuntos de dados existentes no arquivo com uma menor frequência pré-fixada no valor mínimo, conhecido como grandes conjuntos. Neste caso este valor é pré-fixado na ferramenta de MD (CARVALHO, 2000; MINUSSI, 2008).

Este algoritmo utiliza o princípio de que cada subconjunto de um conjunto de itens frequentes também deve ser frequente. Esta regra é utilizada para reduzir o número de causas a serem comparadas com cada transação do banco de dados, sendo que, todas as causas geradas que contenham algum subconjunto que não seja frequente são eliminadas (LANGE, 2007).

3 METODOLOGIA

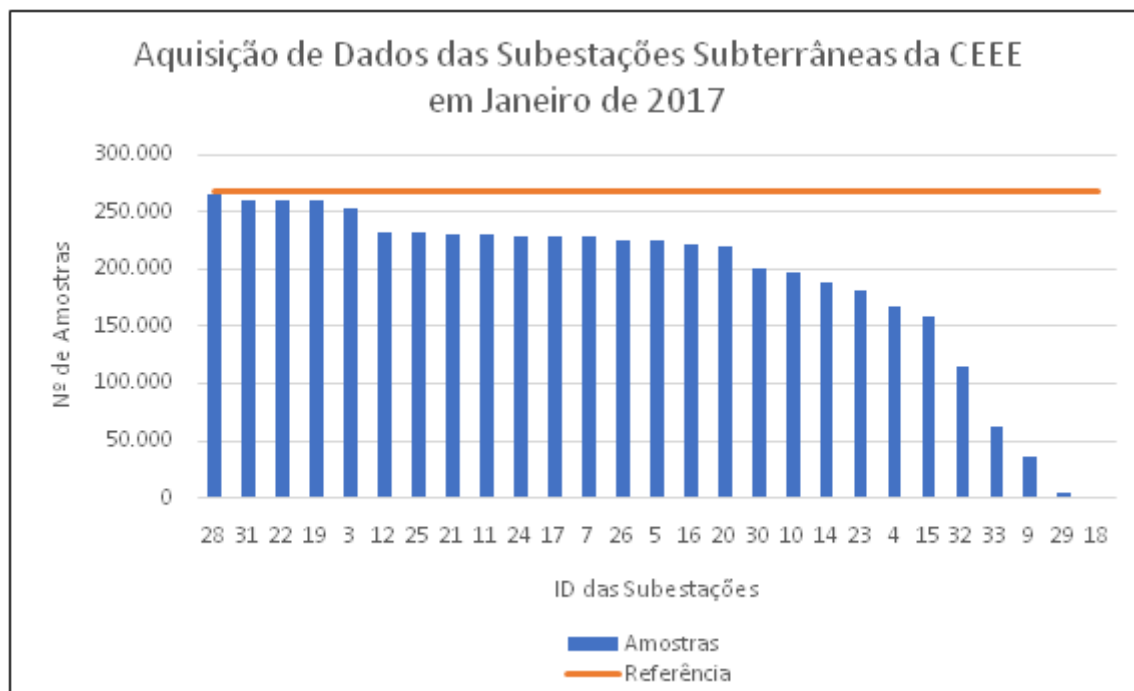
Após o estudo do estado da arte no que se refere a descoberta de conhecimento em bases de dados, focando, em especial, os algoritmos tradicionalmente utilizado no Setor Elétrico, sendo que, em um primeiro momento os dados serão otimizados e organizados (pré-processamento) para realização da MD, posteriormente será realizada a busca de padrões na base de dados de uma concessionária de energia elétrica que neste trabalho está relacionada as subestações subterrâneas da CEEE-D, localizadas no centro da cidade de Porto Alegre/RS. Pretende-se utilizar modelos computacionais e técnicas/algoritmos de descoberta de conhecimento que em conjunto com a análise de um especialista da área possa gerar informações úteis para a tomada de decisões relativa a manutenção da rede de distribuição de energia da CEEE-D. A partir desta análise será possível encontrar padrões de erros ou mesmo ocorrências de falhas/faltas que irão auxiliar na manutenção preventiva na rede. Para os testes preliminares deste trabalho serão exploradas as regras de associação utilizando o algoritmo Apriori e como ferramenta para MD será utilizado o software WEKA descrito na subseção 2.2.4.

4 RESULTADOS E DISCUSSÃO

A partir da parceria existente entre o GAIC e a CEEE foi realizado o pré-processamento para MD da base de dados do mês de janeiro de 2017 contendo 30 subestações subterrâneas de energia elétrica, sendo que a periodicidade na coleta das grandezas que compõem a base de dados em estudo é de uma aquisição a cada 10 segundos o que mesmo em um curto período de tempo pode se transformar em uma rica e grande base de conhecimento para a concessionaria de energia elétrica. Na tabela 1 é possível visualizar a distribuição dos dados no período para cada uma das subestações da CEEE.

Evento: XXII Jornada de Pesquisa

Tabela 1. Aquisição de Dados das Subestações Subterrâneas da CEEE em Janeiro de 2017



Fonte: Próprio Autor

Para melhor entendimento dos resultados apresentados na subseção 4.1 os atributos utilizados na MD são respectivamente (relé, nível de água 1, nível de água 2, bomba, protetor da bomba, ventilador, protetor do ventilador, intruso, protetor), sendo que para os testes iniciais foi escolhida a subestação 28 que é a que possui o maior número de amostras dentro deste período conforme a Tabela 1.

4.1 RESULTADOS

Os resultados obtidos a partir das regras de associação do algoritmo Apriori estão representados na Tabela 2 e foram minerados a partir de 265.474 aquisições referente aos dados extraídos da subestação 28 no mês de janeiro de 2017.

Evento: XXII Jornada de Pesquisa

Tabela 2. Resultados obtidos na ferramenta WEKA após a MD utilizando o algoritmo Apriori.

```
=== Run information ===
Scheme:   weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: CEEE_Subestacao_28.symbolic
Instances: 265474
Attributes: 9
    relay
    water_barrier1
    water_barrier2
    pump
    pump_protector
    fan
    fan_protector
    manhole
    protector

=== Associator model (full training set) ===
Apriori
=====
Minimum support: 0.95 (252200 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 1

Best rules found:
1. water_barrier2=False 265474 ==> water_barrier1=False 265474 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. water_barrier1=False 265474 ==> water_barrier2=False 265474 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. pump_protector=False 265474 ==> water_barrier1=False 265474 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. water_barrier1=False 265474 ==> pump_protector=False 265474 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. fan=True 265474 ==> water_barrier1=False 265474 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. water_barrier1=False 265474 ==> fan=True 265474 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. fan_protector=False 265474 ==> water_barrier1=False 265474 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. water_barrier1=False 265474 ==> fan_protector=False 265474 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. pump_protector=False 265474 ==> water_barrier2=False 265474 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. water_barrier2=False 265474 ==> pump_protector=False 265474 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
```

Fonte: Próprio Autor.

Após análise dos resultados obtidos a partir das regras de associação do algoritmo Apriori na subestação 28 e com base no histórico dos atributos utilizados observou-se que os padrões retornados estão dentro da normalidade e os resultados do algoritmo não são significativos para o sistema de monitoramento da concessionária.

Evento: XXII Jornada de Pesquisa

5 CONSIDERAÇÕES FINAIS

Com base nas regras de associação obtidas através do algoritmo Apriori observa-se que este retornou resultados normais (Tabela 2). Apesar dos resultados obtidos na MD desta subestação não serem significativos para utilização do sistema de monitoramento da concessionária os mesmos não devem ser desprezados pois podem ser importantes principalmente para combinação dos resultados obtidos com outros algoritmos de MD.

Destaca-se que o processo de Descoberta de Conhecimento e a Mineração de dados podem auxiliar de forma satisfatória as concessionárias na busca de padrões para decisões estratégicas não só com relação ao consumidor final e aos seus perfis de consumo, mas também para descobrir quais subestações necessitam de manutenção preventiva, por exemplo, antes de outras, pois só existe orçamento para fazer este procedimento em uma parte das subestações.

Para trabalhos futuros pretende-se ampliar o número grandezas pesquisadas, como por exemplo, corrente, voltagem, temperatura e também correlacionar todas as subestações entre si, para aplicação de outros algoritmos objetivando a busca de padrões na base de dados da CEEE que possam após a análise de um especialista serem utilizados futuramente para o sistema de monitoramento da concessionária.

REFERÊNCIAS

CAMILO, C. O.; SILVA, J. C. da. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. Universidade Federal de Goiás (UFG), p. 1-29, 2009.

CASTANHEIRA, L. G. Aplicação de técnicas de mineração de dados em problemas de classificação de padrões. Dissertação de Mestrado, UFMG, Belo Horizonte, 2008.

CHOI, J.; JEONG, Y. K.; LEE, I. W. Development of the simple building electric power prediction model with local weather forecast based on clustering and silhouette algorithm. In Emerging Technologies & Factory Automation (ETFA), 2015 IEEE 20th Conference on, pages 1-4. IEEE, 2015.

CIOS, K. J.; et. al. Data mining - a knowledge discovery approach. In Data Mining - A Knowledge Discovery Approach. Springer, 2007.

CORTES, S. da C.; PORCARO, R. M.; LIFSCHITZ, S. Mineração de dados - funcionalidades, técnicas e abordagens. PUC, 2002.

CARVALHO, J. V. de; MONGIOVI, M. C. S. G. Utilização de técnicas de data mining para o reconhecimento de caracteres manuscritos. Universidade Federal da Paraíba. Campina Grande, PB, 2000.

DUARTE, M. S. Proposta de um sistema para gerenciamento de energia residencial no contexto das smart grids. Dissertação de Mestrado, Universidade Estadual do Ceará, 2014.

Evento: XXII Jornada de Pesquisa

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. Data mining to knowledge discovery in databases. American Association for Artificial Intelligence, 2006.

GOES, E. A. et. al. Aplicação de redes neurais na identificação de setores em curto circuito nos sistemas de distribuição de energia elétrica. Dincon, 2010.

JIA, N.; WANG, J.; LI, N. Application of data mining in intelligent power consumption. IET, 2012.

KAMBER, M.; HAN, J.; PEI, J. Data mining: Concepts and techniques. Elsevier, 2012.

KITAYAMA, M.; MATSUBARA, R.; IZUI, Y. Application of data mining to customer profile analysis in the power electric industry. In Power Engineering Society Winter Meeting, 2002. IEEE, volume 1, pages 632-634. IEEE, 2002.

LANGE, L. C. Mineração de dados em sistema eficiente de iluminação pública incluindo parâmetros sócio-comportamentais. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, 2007.

LAROSE, D. T. Discovering knowledge in data: an introduction to data mining. John Wiley & Sons, 2014.

MAIA, A. T. et al. Data mining techniques assessment for shutdowns prediction of electric power systems. In Proceedings of the annual conference on Brazilian Symposium on Information Systems: Information Systems: A Computer Socio-Technical Perspective- Volume 1, page 22. Brazilian Computer Society, 2015.

MCCUE, C. Data mining and predictive analysis: Intelligence gathering and crime analysis. Butterworth-Heinemann, 2014.

MINUSSI, M. M. et al. Metodologia de mineração de dados para detecção de desvio de comportamento do uso de energia em concessionária de energia elétrica. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, 2008.

ROCHA, D. V.; SEPÚLVEDA, G. P. L.; FLORES, H. A. R. Mineração de dados aplicada à detecção de fraudes nas redes de distribuição de energia elétrica. VI Congresso Brasileiro de Engenharia da Produção. Paraná, 2016.

SCHUCH, R. et al. Mineração de dados em uma subestação de energia elétrica. Dincon, 2010.

WITTEN, I. H. et al. Data Mining: Practical machine learning tools and techniques with java implementations. Morgan Kaufmann, 2005.

ZHANG, S. et al. The research on smart power consumption technology based on big data. In Smart Grid and Clean Energy Technologies (ICSGCE), 2016 International Conference on, pages 12-18. IEEE, 2016.

Evento: XXII Jornada de Pesquisa