

Detecção de Fraudes nas Redes de Distribuição de Energia Elétrica Utilizando Técnicas de Inteligência Computacional

Altamira de Souza Queiroz
Universidade E. do Oeste do Paraná
Foz do Iguaçu, Paraná, BR
Email: altamirafz@gmail.com

Edgar Manuel Carreno Franco
Universidade E. do Oeste do Paraná
Foz do Iguaçu, Paraná, BR
Email: emfra.unioeste@gmail.com

Gloria Patricia López
Universidade Tecnológica F. do Paraná
Medianeira, Paraná, BR
Email: pattylla@gmail.com

Resumo—Um dos principais problemas que enfrentam as empresas concessionárias de energia elétrica é a ocorrência de perdas de energia nas redes de distribuição, causados por fraudes e furtos de energia elétrica. O presente trabalho apresenta uma análise entre algoritmos de Redes Neurais Artificiais, Máquinas de Vetores de Suporte e Árvore de Decisão, que são algoritmos de Inteligência Computacional, para extrair conhecimento de bancos de dados que contém registros de consumo de energia elétrica, a fim de identificar padrões de consumo com anomalias que levam a detecção de possíveis fraudes nas redes de distribuição de energia elétrica. Os resultados obtidos são classificatórios, e além de classificar, tais técnicas detectam novos padrões de possíveis fraudes nas redes de distribuição de energia elétrica.

I. INTRODUÇÃO

O crescimento da concorrência entre as concessionárias distribuidoras de energia elétrica e o aumento dos prejuízos financeiros causados pelas perdas de energia elétrica consequentes das anomalias nas redes de distribuição, têm incrementado a busca por novas tecnologias de detecção de fraudes no mercado da energia elétrica [1].

As anomalias presentes nas redes de distribuição causam perdas que podem ser classificadas como perdas técnicas e perdas comerciais. As perdas técnicas ocorrem naturalmente por características próprias do sistema e ações internas de materiais que compõe o sistema elétrico. Enquanto as perdas comerciais ocorrem por ação do homem nos equipamentos de medição e nas redes de distribuição de energia [2].

De acordo com um estudo realizado pelo Departamento de Energia Elétrica da Universidade Estadual de Campinas, estima-se que 17% da energia elétrica inserida no sistema de distribuição no Brasil são desviados através de perdas comerciais, o que corresponde à cerca de 7.428.000MWh [3]. Um número notavelmente alto e que gera prejuízo às empresas concessionárias e ao consumidor, pois parte do custo da energia perdida é repassada ao consumidor, de acordo com as normas da Agência Nacional de Energia Elétrica.

Levando em consideração que as perdas comerciais além de representarem grandes perdas econômicas, elas também colocam em risco a segurança pública e criam um grande impacto social, as empresas concessionárias de energia elétrica veem fazendo utilização de uma metodologia conhecida como

Inteligência de Negócios com o objetivo de diminuir essas perdas e reduzir os impactos socioeconômicos causados pelas perdas comerciais [4]. De acordo com Rezende [5], a Inteligência de Negócios é um conjunto de métodos que podem auxiliar na definição de estratégias com o objetivo de aumentar a competitividade de uma empresa no mercado e nos negócios.

Neste contexto, a utilização de Inteligência Computacional para detecção de anomalias na rede de distribuição de energia elétrica com o objetivo de diminuir as perdas comerciais é uma alternativa de Inteligência de Negócios atrativa. Sua atratividade é devido a rapidez de análise da curva de consumo mensal dos consumidores de energia elétrica através de técnicas de Mineração de Dados (MD), e o grande número de opções de ferramentas que a MD oferece para a classificação de usuários.

De acordo com Guizzo [6], a MD é a extração de informações previamente desconhecidas e potencialmente úteis de grandes bases de dados, com o objetivo de descobrir perfis de consumidores e outros comportamentos que não poderiam ser identificados somente com análises manuais de especialistas. Ainda é possível complementar que a MD é uma ferramenta composta por diversos algoritmos que podem ser utilizados para classificação, regressão, agrupamento entre outras tarefas. A escolha de tarefa à ser utilizada e do algoritmo que chegará no melhor resultado depende do tipo de dados utilizados e também do que se espera do processo de mineração.

Com o objetivo de detectar consumidores com anomalias na curva de consumo de energia elétrica, e com uma base de dados compostos por dados de consumo mensais em kWh de 3.840 usuários, o presente trabalho realizou um conjunto de experimentos para definir qual dos algoritmos de MD, utilizados para classificação, entre Redes Neurais Artificiais (RNA), Máquinas de Vetores de Suporte e Árvore de Decisão, realiza a melhor classificação e descobre quais são os usuários com consumos anormais.

II. PADRÕES DE CONSUMO DE ENERGIA ELÉTRICA

Os valores de consumo de energia elétrica mensal dos usuários são armazenados pelas concessionárias após a leitura dos valores de consumo nos medidores. Através da análise

dos valores de energia consumida é possível classificar os consumidores por padrões de consumo. Esses padrões de consumo variam de acordo com as tipologias de edificações, situações sociais e financeiras de cada consumidor, situação econômica do País, entre outros fatores que inclui clima e determinados acontecimentos em períodos do ano, como férias escolares, viagem, entre outros [7].

É importante ressaltar que independente do tipo do consumidor, o consumo de energia elétrica possui um comportamento sazonal e cíclico que pode ser observado quando é feito análise da curva de consumo anual dos usuários, e tal curva mantém um comportamento de modo regular, chamado de padrão de consumo. A curva de consumo de energia elétrica anual é um gráfico de linhas onde os valores de consumo mensal são colocados em uma tabela e em sequência é gerado um gráfico onde pode ser visualizado o consumo mensal de energia e analisado os padrões da curva de consumo. Na Figura 1 pode-se observar no gráfico a curva de consumo de um determinado usuário de energia elétrica durante três anos onde é possível observar o comportamento sazonal e cíclico neste período.



Figura 1. Gráfico das curvas de consumo de energia elétrica sem anomalias em um período de três anos

Através da análise das curvas de consumo de energia elétrica e levando em consideração o comportamento de tais curvas, é possível detectar padrões de consumo que podem ser utilizados para classificar consumidores e detectar anomalias nas redes de distribuição de energia elétrica.

A. Principais Padrões de Curva de Consumo com Anomalias

Analisando as principais causas de anomalias nas redes de distribuição de energia elétrica é fácil observar o padrão de possíveis fraudes. Na Figura 2 é possível observar um caso de desvio na rede elétrica por outro usuário próximo a rede de energia deste usuário representado no gráfico. Nota-se que o consumo aumenta de forma constante durante todo o ano. Este é um caso de fraude que não causa prejuízo

financeiro para a concessionária de energia elétrica, pois a energia está sendo passada no medidor de um usuário e consumida por outro, ou seja, de qualquer forma, um usuário está pagando pela energia consumida, mas este tipo de fraude causa transtorno à sociedade, além de colocar em risco a segurança pública.

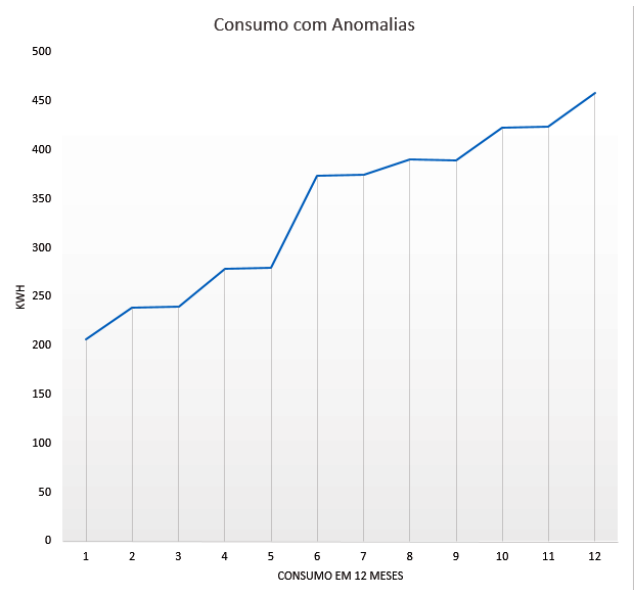


Figura 2. Gráfico das curvas de consumo de usuário com anomalia causada por ligação direta por outro usuário após a medição

Outro caso comum de fraudes nas redes de distribuição, é o caso de medidores fraudados com freios (furos, e/ou agulhas) no disco de medição. Esse tipo de fraude, representado na Figura 3, faz com que o disco medidor de energia fique travado em um determinado valor, e não mede toda a energia consumida. Este padrão de curva de consumo também acontece quando é realizada ligação direta antes do medidor de energia e o usuário utiliza a energia passada pelo medidor apenas para determinados fins, diminuindo assim, o consumo registrado pelo medidor.

Nota-se que o padrão das curvas quando existe algum tipo de fraude nas redes de distribuição de energia elétrica é notável quando a fraude é mais evidente, se comparado com curvas de consumo normais. Vale lembrar que os tipos de padrões de consumo, com anomalias, ou não, variam de consumidor para consumidor, mas se mantém sempre um padrão de consumo ao longo do ano. Levando isso em consideração, com ferramentas adequadas de análises de dados é possível detectar padrões de consumo e detectar anomalias nas curvas de consumo que podem representar possíveis fraudes e furtos nas redes de distribuição de energia elétricas através do histórico de consumo de cada consumidor.

III. ESTUDO DE CASOS

Como estudos de casos, foram selecionados os três algoritmos de MD mais utilizados para a tarefa de classificação e estes foram testados com 10 configurações diferentes para

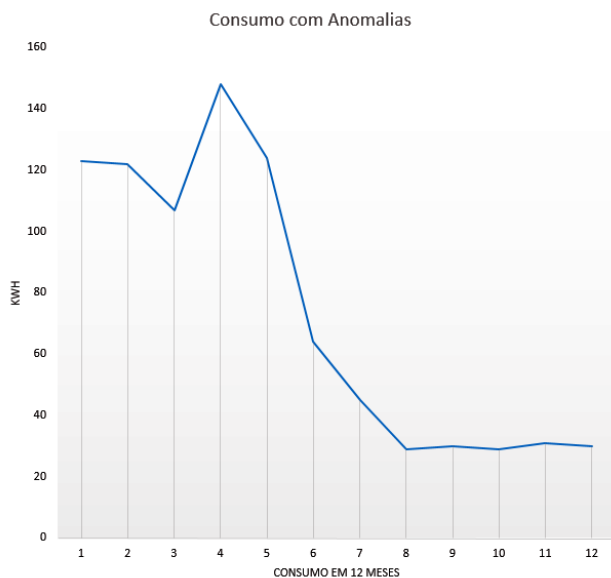


Figura 3. Gráfico das curvas de consumo de energia elétrica de usuário com medidor fraudado

verificar qual obteria melhor resultado para um banco de dados com os registros de consumo em kWh mensal de energia elétrica de mais de 31400 usuários. Os algoritmos selecionados para os testes foram: árvores de decisão através da aplicação do algoritmo *J48*; as máquinas de vetor de suporte através da aplicação do algoritmo *libSVM* e as RNA através do algoritmo *MultilayerPerceptron*.

Para a execução dos testes, foi utilizado o software open source chamado WEKA, na versão 3.7.12. O WEKA é uma ferramenta de descoberta do conhecimento em base de dados que possui uma série de algoritmos de aprendizado de máquina aplicados na área de inteligência computacional [10]. Possui acoplado em seu código fonte os algoritmos selecionados para os testes realizados nesta pesquisa. Além disso, reconhece ficheiro em varios formato entre eles podem ser destacados o *Attribute Relation File Format (.arff)*, e o *comma separated values (.csv)*.

A. Base de Dados

A base de dados utilizada no presente trabalho foi fornecida por uma concessionária de energia elétrica colombiana e contém informações de consumo mensal durante os anos de 2011, 2012 e 2013, informações do tipo do consumidor, código de usuário, tipos de fases e alimentador, entre outras informações. Para complementar, a concessionária também forneceu uma base de dados com as informações de revisões manuais realizadas em todos os clientes cadastrados e as anomalias encontradas durante tais revisões.

B. Algoritmos de Mineração de Dados

O *J48* é um algoritmo de Árvore de Decisão que permite realizar a classificação dos usuários de acordo com o tipo de anomalias no consumo. O algoritmo tem o objetivo de

interpretar dados a fim de produzir uma quantidade de padrões de consumos anômalos para posteriormente classificar novos dados [1].

O algoritmo *libSVM* é um algoritmo de Máquinas de Vetor de Suporte, de aprendizagem supervisionada com o objetivo de classificar padrões [8]. A ideia principal desse tipo de algoritmo é construir um hiperplano como superfície de decisão de uma maneira que a margem de separação entre exemplos positivos e negativos seja máxima [8].

O algoritmo *MultilayerPerceptron* é um algoritmo de Rede Neural composto por redes *perceptron* de múltiplas camadas. Esse tipo de rede possui todas as entradas conectadas diretamente com as saídas e podem ter diversas camadas ocultas [9].

Estes três algoritmos foram utilizados no processo de MD fazendo uso da ferramenta computacional WEKA, esta ferramenta permitiu realizar o treinamento e validação dos algoritmos mediante a utilização da plataforma *Explorer*. Nesta plataforma foi possível executar o pre-processamento do banco de dados através da aplicação de filtros, permitindo que os algoritmos de classificação possam realizar o processo de aprendizado de forma eficiente.

C. Modelagem dos Dados

Além das informações de consumo mensal de energia elétrica dos usuários presentes na base de dados utilizada neste trabalho, informações referentes a registros de clientes, tipos de consumidores, tipo de fase e alimentadores também compõem a base de dados. Como alguns dessas informações não são necessárias para definir a curva de consumo dos usuários, foram retiradas, mantendo apenas os *ids* de usuários e consumos mensais de energia elétrica. Após a realização do pre-processamento, os dados foram convertidos para arquivos com extensão *csv*, com o objetivo de facilitar a interpretação dos mesmos pela ferramenta WEKA.

D. Criação dos Modelos

Com o objetivo de extrair o conhecimento da base de dados e encontrar padrões de consumos através da análise das curvas de consumo de usuários de energia elétrica, foi realizado o processo de MD utilizando algoritmos de Árvore de Decisão, RNAs e Máquinas de Vetores de Suporte e a base de dados dos consumos do ano de 2010. Cada algoritmo foi testado 10 vezes, alterando as configurações de trabalho de cada um deles, e cada teste criou um modelo, com uma porcentagem de acerto.

O ferramenta computacional permite alterações dos valores estabelecidos como padrão de cada parâmetro que é utilizado para configurar cada algoritmo. No *J48* foi possível editar os seguintes parâmetros: valores do número mínimo de casos por folha (*minNumObj*), fator de confiança utilizado para a poda (*confidenceFactor*), quantidade de dados utilizados para a redução de erros de poda (*numFolds*) e, semente usada para randomizar os dados (*seed*). Fazendo as alterações permitidas desses parâmetros dentro da ferramenta WEKA, pode ser

verificado se a alteração para maior ou menor valor melhora o resultado obtido.

Na Tabela I é possível observar a mudança na porcentagem de acerto a cada alteração realizada no processo de criação do modelo. É importante ressaltar que para essa base de dados, a alteração nos valores de *seed* e *numFolds* não fizeram nenhuma alteração na porcentagem de acerto da classificação.

Tabela I
ALTERAÇÕES NAS CONFIGURAÇÕES DO J48

Teste	conf.Factor	minNumObj	n.Folds	seed	Acerto
01	0.25	2	3	1	85,3339%
02	0.10*	2	3	1	85,6670%
03	0.50*	2	3	1	85,1336%
04	0.10*	1*	3	1	82,6277%
05	0.10*	4*	3	1	85,7609%
06	0.10*	5*	3	1	85,6997%
07	0.10*	4*	1*	1	85,7609%
08	0.10*	4*	5*	1	85,7609%
09	0.10*	4*	3	3*	85,7609%
10	0.10*	4*	3	0.5*	85,7609%

*Valores alterados

O algoritmo *libSVM* permite alterações em diversos parâmetros, entre eles: os valores de coeficiente, degraus e o Kernel. Mas nenhuma alteração, para a base de dados utilizadas e/ou similares, altera a porcentagem de acerto na classificação dos dados, que é de 76,6740% para a base de dados utilizada.

O algoritmo *MultilayerPerceptron* também permite configurações que podem provocar alteração na porcentagem de acerto da classificação. Como pode ser observado na Tabela II, a alteração que provocou alteração nos resultados foram os valores dos números de épocas para treinar completamente a rede (*TrainingTime*).

Tabela II
ALTERAÇÕES NAS CONFIGURAÇÕES DO *MultilayerPerceptron*

Teste	TrainingTime	Acerto
01	2000	80,5422%
02	5000	84,3011%
03	10000	86,4733%
04	50000	91,0249%
05	100000	92,7324%
06	200000	92,6829%

É importante observar que o aumento do número de épocas melhora o resultado na classificação até determinado momento, a partir de então, a porcentagem de acerto começa a sofrer uma queda, como é ilustrado no teste 06 da Tabela II. Nas tabelas só foram apresentados os dados que com as alterações dos parâmetros de configuração levaram a obter uma porcentagem de acerto maior nos resultados da classificação.

E. Testes com os Modelos Criados

Após a criação dos modelos utilizando os dados do ano de 2010, os quais já possuíam classificação de usuarios fraudulentos ou não, o melhor modelo de cada algoritmo de classificação, é validado com os registros de consumo dos anos de 2011 e 2012, cabe salientar que estes dados não foram utilizados no treinamento dos modelos, tendo como objetivo estudar os resultados que o modelo criado obteve ao classificar os novos dados inseridos.

1) *Árvore de Decisão*: o modelo de *Árvore de Decisão* que alcançou 85,7609% de acerto durante a criação do modelo, classificou os dados dos anos de 2011 e 2012 com uma porcentagem melhor do que a obtida na criação do modelo. A Tabela III apresenta a porcentagem de acerto durante o treinamento com novos dados, e pode-se observar que houve uma pequena melhoria na classificação.

Tabela III
TESTES COM O MODELO J48

	acerto	erro
Criação do Modelo	85,7609%	14,2391%
Classificação de 2011	86,739%	13,261%
Classificação de 2012	86,782%	13,218%

2) *Máquina de Vetor de Suporte*: O modelo MVS que alcançou 76,6740% durante a criação do modelo obteve um resultado com uma pequena variação entre a porcentagem de acerto nos anos de 2011 e 2012. A Tabela IV ilustra a porcentagem de acerto durante o treinamento com os dados de 2010, os testes de classificação com os dados de 2011 e 2012.

Tabela IV
TESTES COM O MODELO MVS

	acerto	erro
Criação do Modelo	76,674%	23,326%
Classificação de 2011	76,739%	23,261%
Classificação de 2012	81,239	17,761%

3) *Rede Neural Artificial*: O modelo RNA que alcançou 92,7324% de acerto durante a criação do modelo, classificou os 314023 usuários de 2011 e os 313972 usuários de 2012 com um tempo de execução para classificação de 4 segundos, e a porcentagem de acerto na classificação foi de 92,7324 em ambas as classificações. Ou seja, dos usuários inseridos, 92,7324% foram classificados entre usuários que possuem anomalias ou não. A Tabela V ilustra a porcentagem de acerto na classificação durante a criação do modelo e os dois anos classificados. Vale lembrar que a porcentagem de acerto se manteve em ambos os anos com 98,2909%.

IV. ANÁLISES DOS RESULTADOS OBTIDOS

Com as simulações realizadas durante o desenvolvimento do presente trabalho, tendo em mãos um banco de dados

Tabela V
TESTES COM O MODELO RNA

	acerto	erro
Criação do Modelo	92,7324%	7,2676%
Classificação de 2011	92,7324%	7,2676%
Classificação de 2012	92,7324%	7,2676%

de consumo mensal de consumidores de energia elétrica é possível comparar e analisar os resultados obtidos durante o treinamento e validação com os modelos criados, afirmar qual dos algoritmos teve um melhor resultado e pode-se contribuir de forma eficiente nas tomadas de decisões das empresas distribuidoras de energia elétrica.

Dado que os resultados obtidos durante o treinamento para a criação dos modelos usando o algoritmo de MVS não apresentou nenhuma alteração na porcentagem de acerto, não foi possível fazer uma análise comparativa do motivo da mudança dos resultados.

Nos resultados obtidos pelo algoritmo J48 foi possível observar que alterações no fator de confiança, no número mínimo de casos por folhas e na quantidade de dados utilizados para a poda alteraram a porcentagem de acerto.

No caso da aplicação do algoritmo de RNA, nota-se que o aumento do número de épocas para treinamento obteve o aumento da porcentagem de acerto até um certo tempo, quando o valor do *TrainingTime* foi aumentado para 200.000, a porcentagem de acerto caiu. Isso acontece devido ao treinamento excessivo da RNA, que pode causar uma má classificação dos dados.

Mesmo nas configurações iniciais nos três algoritmos testados, os resultados do treinamento para criação do modelo apresentaram uma porcentagem de acerto positiva, com resultados entre 75% e 85% de acerto na classificação. Essa positiva porcentagem de acerto mostra que mesmo em configurações padrões dos algoritmos, para o banco de dados utilizado, a MD pode ser utilizada com um grau de confiança notável na colaboração das tomadas de decisões.

A Figura 4 ilustra a melhor porcentagem de acerto de cada algoritmo durante o treinamento para criação dos modelos. É possível observar que, para esse tipo de dados, o algoritmo de RNA foi notavelmente melhor no processo de treinamento para classificação dos dados.

É importante ressaltar que o processo de treinamento é, em alguns casos, lento. Alguns treinamentos demoram cerca de 15 dias para se obter um modelo com boa quantidade de acerto. Mas o tempo de treinamento, como é realizado apenas para a criação do modelo, não prejudica na utilização de MD para classificações, pois depois do modelo criado, a execução dos modelos para classificar novos dados, variam de 3 a 10 segundos, isto para os dados e modelos utilizados no presente trabalho.

Na execução dos modelos para classificar novos dados, utilizando o banco de dados dos anos de 2011 e 2012, o algoritmo de RNA também obteve resultado superior ao

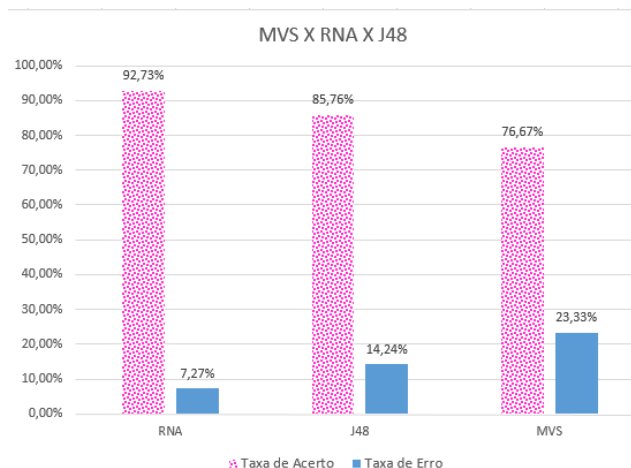


Figura 4. Comparação dos resultados obtidos no treinamento e/ou criação dos modelos

resultado obtido com o algoritmo de MVS e J48, como pode ser observado na Figura 5, onde é realizada a comparação entre os três algoritmos.

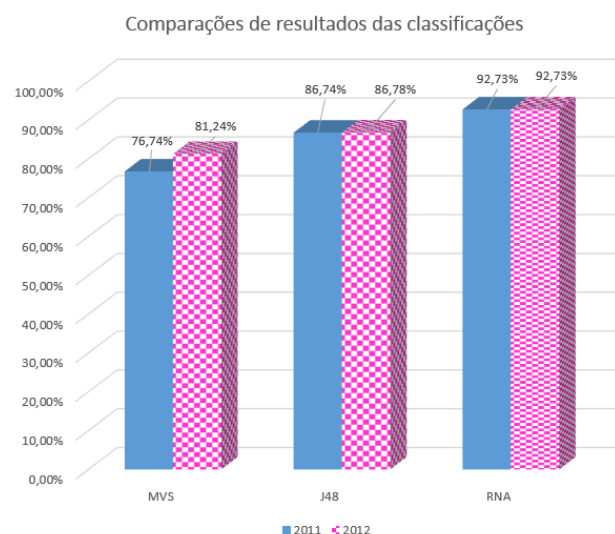


Figura 5. Comparação dos resultados obtidos nas classificações utilizando os modelos criados

Neste contexto, e analisando os gráficos, é possível afirmar que, apesar de ambos os algoritmos apresentarem um resultado positivo na classificação dos dados, o algoritmo RNA alcançou melhor porcentagem de acerto durante o processo de classificação.

Durante o processo de classificação, novos usuários foram detectados com padrões de consumo que levam a considerá-los como possíveis usuários fraudulentos, sendo que estes, na tabela de classificação oferecida pela concessionária, não estavam classificados como usuários fraudulentos.

Em alguns casos, a empresa concessionária classificou um usuário como não fraudulento e o processo de MD encontrou anomalias no padrão de consumo deste usuário, sendo que

este comportamento anormal pode ser observado na curva de consumo gerada a partir do histórico de consumo deste usuário.

O gráfico da Figura 6 ilustra um caso no qual o consumo foi reduzido cerca de 50% se comparado aos quatro primeiros meses de consumo, e na tabela de classificação feita pela concessionária consta-se como sendo um usuário sem anomalias. Este é um possível caso de freio no medidor, onde o usuário, através de técnicas de fraudes, freia o medidor para marcar menor quantidade do que a energia realmente consumida.

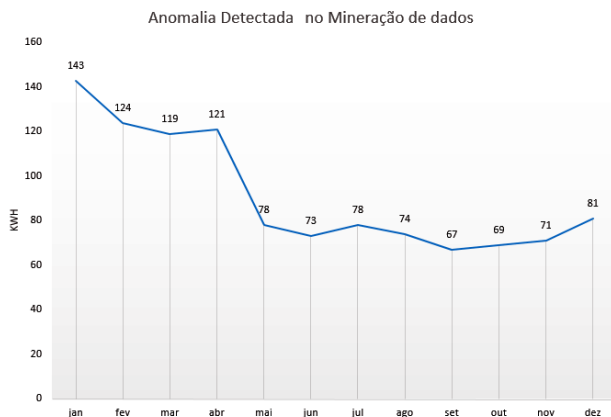


Figura 6. Usuário com anomalias a partir do mês de maio

Ligações diretas feitas nas redes de distribuição, são detectadas com facilidades quando todo o consumo do usuário fraudador é feito através dessa ligação. No caso de usuários que fazem ligação direta somente para utilizar determinados aparelhos elétricos, e parte do consumo continua sendo através da energia passada pelo medidor são mais difíceis de serem percebidos pelas concessionárias. Na Figura 7 pode-se visualizar um possível caso de ligação direta na rede de distribuição, o qual não foi detectado pelo processo de classificação da concessionária, e no processo de MD, foi classificado como usuário com possíveis anomalias. É possível observar que o consumo de energia foi reduzindo mensalmente até chegar no mínimo permitido pela concessionária.

V. CONCLUSÃO

Levando em consideração a quantidade de dados armazenados diariamente com os registros das leituras mensais nos medidores dos usuários de energia elétrica, é possível afirmar que a utilização de técnicas de inteligência computacional para transformar tais dados em conhecimento podem auxiliar nas tomadas das decisões das empresas.

Sendo que o principal objetivo deste trabalho é detectar padrões de possíveis fraudes através de análises das curvas de consumo dos consumidores de energia elétrica, pode-se afirmar que com a utilização de RNA, o resultado das classificações é melhor do que quando utilizado algoritmos de Árvore de Decisão e Máquinas de Vetores de Suporte.

É importante salientar que o tempo de treinamento pode ser lento durante a criação dos modelos, mas para a classificação

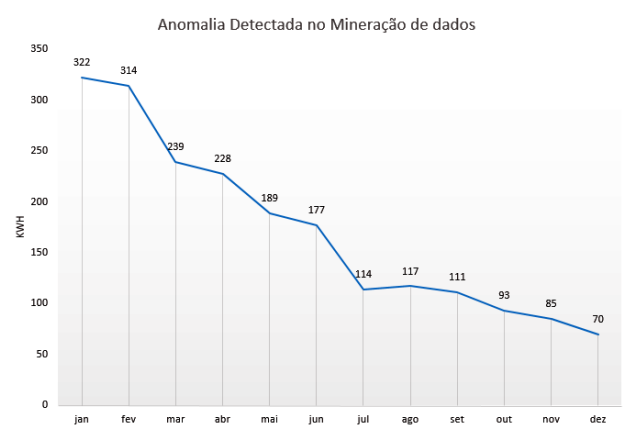


Figura 7. Caso de possível fraude não detectado no processo de classificação feito pela concessionária

de novos dados com o modelo pronto, todos os algoritmos testados alcançaram cerca de 100.000 classificações por segundo.

Os resultados obtidos com os diferentes algoritmos encorajam suas utilizações como ferramenta de auxílio nas tomadas de decisões. Pois, com mais de 90% de acerto, a MD torna-se viável, e essa porcentagem de acerto ainda pode ser melhorada com novos treinamentos.

Essa alta porcentagem de acerto na classificação dos dados reduz uma grande quantidade de fiscalizações manuais, que antes eram realizadas pelas concessionárias a fim de encontrar possíveis fraudes nas redes de distribuição de energia elétrica.

REFERÊNCIAS

- [1] G. P. López, M. J. Rider and C. J. Zapata-Grisales, *Deteção de Fraudes nas Redes de Distribuição de Energia Elétrica Utilizando Árvore de Decisão*, Simpósio Brasileiro de Sistemas Elétricos. Anais 2014. Foz do Iguaçu-PR.
- [2] C. A. S. Penin, *Combate, Prevenção e Otimização das perdas Comerciais de Energia Elétrica*, Tese de Doutorado. Universidade de São Paulo, 2008.
- [3] S. Anunciação, *Cerco às Fraudes de Energia: Grupo Desenvolve Equipamentos e Metodologias de Baixo Custo e Alta Produtividade*, Jornal da Unicamp, 2015.
- [4] W. H. Inmon, *Building the Data Warehouse: Getting Started*, 4ª Edição. Editora Wiley Publishing, 2005.
- [5] S. O. Rezende, J. B. Pugliese and F. M. Varejão, *Sistemas Baseados em Conhecimentos, Sistemas Inteligentes*, Cap.2, Editora Manole, 2003.
- [6] E. Guizzo, *Quem Procura Acha*, Revista de Negócios Exame, Acesso realizado em maio/2014. Disponível em: <http://portalexame.abril.uol.com.br/complementos/revista0002.html>
- [7] A. M. D. Hansen, *Padrões de Consumo de Energia Elétrica em Diferentes Tipologias de Edificações Residenciais*, em Porto Alegre. Universidade Federal do Rio Grande do Sul - UFRS. Dissertação de Mestrado. Porto Alegre, 2000.
- [8] D. A. Marreto, *Aplicação de Máquinas de Vetores de Suporte para Desenvolvimento de Modelos de Classificação e Calibração Multivariada em Espectroscopia no Infravermelho*, Tese de Doutorado. Universidade Estadual de Campinas, 2011.
- [9] P. Norvig and S. Russell, *Inteligência Artificial*, 3ª Edição. Editora Campus - Elsevier, 2013.
- [10] U. Waikato, *Weka Machine Learning Software in Java*, The University of Waikato. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka>. Acesso realizado em agosto/2015.
- [11] A.J.R. Cruz, *Data Mining via Redes Neurais Artificiais e Máquinas de Vetores de Suporte*, Dissertação de Mestrado. Universidade do Minho, Braga-Portugal, 2008.