# Aprendizagem de Máquina Aplicada a Consumidores Comerciais Buscando Identificar Padrões Atípicos de Consumo de Energia Elétrica Utilizando o Software R

Lucas Evangelista de Souza
Universidade do Vale do Itajaí
UNIVALI - SC
lucas.won@hotmail.com

Raimundo C. Ghizoni Teive†
Universidade do Vale do Itajaí
UNIVALI
rteive@univali.br

## ABSTRACT

The electricity distribution network is responsible for supplying energy to consumers in the National Interconnected System, serving 99% of consumers in Brazil. There are two types of losses in this network: technical losses and non-technical losses or commercial losses. In the case of non-technical losses, the focus of this work, the existence of these results in a higher tariff for all consumers, so that the concessionaire can compensate for such reduction in revenue. Non-technical losses are usually associated with fraud (meter tampering or deviations). The main objective of this work is the application of machine learning techniques, using software R, to identify possible fraudulent behaviors of commercial consumers in the state of Santa Catarina. Considering data from typical consumer load curves and functional information from the company. Preliminary results, using real data from consumers, indicate that the SVM classifier used performed well in the cases studied, achieving precision and accuracy greater than 90%. The input variables selected for the classifier, based mainly on data and information from typical load curves, are the differential of this work, as well as the main reason for the success in the initial tests.

## PALAVRAS-CHAVE

Perdas não técnicas, Aprendizagem de máquina, Classificador SVM, Distribuição de energia elétrica.

### 1 Introdução

Novas tecnologias estão sendo estudadas e desenvolvidas para melhorar o desempenho e eficiência das redes de distribuição de energia, buscando como meta, atingir a implantação de conceitos existentes nas redes elétricas inteligentes (*Smart Grids*). Neste contexto, a eficiência energética é um dos principais aspectos a serem melhorados, a qual pode ser obtida por meio da redução de perdas técnicas e comerciais, pela melhoria da qualidade energética ofertada ao consumidor e pela gestão do horário do consumo de energia pelo consumidor [1].

As perdas totais representaram 14% do mercado consumidor em 2018, segundo estudo recente da Aneel [1]. Este percentual, envolvendo tanto perdas técnicas, quanto não técnicas (comerciais), é bastante significativo, sendo equivalente ao consumo da região norte e centro-oeste no ano de 2016. Assim, a Aneel incentiva as empresas distribuidoras de energia, a cada vez mais buscarem a redução destas perdas. Em especial, em relação às perdas não técnicas, a ANEEL recomenda a implantação de técnicas para o combate dessas perdas de energia.

As perdas não técnicas, em particular àquelas referentes às fraudes de energia, são perdas indesejáveis e devem ser combatidas, utilizando-se técnicas computacionais eficientes. Neste sentido, torna-se cada vez mais necessária a pesquisa de novos métodos que obtenham mais flexibilidade e facilidade de adaptação a este problema, como os modelos baseados em técnicas da inteligência artificial, conforme observado por [2].

Dentro deste contexto, este trabalho visa a aplicação da técnica de aprendizagem de máquina para identificação de possíveis fraudes no sistema de distribuição de energia elétrica, particularmente em consumidores comerciais. Desta forma, foi desenvolvido um classificador SVM no software R, para identificar possíveis comportamentos fraudulentos (ou atípicos) no consumo de energia, considerando para isto dados de curvas de carga típicas dos consumidores escolhidos (CNAE escolhido) e demais dados funcionais da empresa.

Este artigo está subdividido da seguinte forma. A Seção 2 analisa o contexto do problema de perdas não técnicas de um sistema de distribuição. A Seção 3 apresenta alguns conceitos relativos a aprendizagem de máquina, classificador SVM e software R. Na seção 4 são apresentados alguns aspectos técnicos do sistema desenvolvido. Os resultados preliminares, envolvendo a aplicação do classificar SVM nos consumidores selecionados, estão apresentados na seção 5, enquanto que na seção 6 são apresentadas as considerações finais.

### 2. Perdas não Técnicas em Sistemas de Distribuição

O sistema distribuição é responsável por distribuir a energia recebida do sistema de transmissão para os consumidores da rede elétrica brasileira. Este sistema é composto de equipamento que operam em baixa tensão (igual ou inferior a 1kV), média tensão (superior a 1kV e inferior a 69kV) e alta tensão(superior a 69kV e inferior a 230kV). Segundo a ANEEL, o Brasil possuía 105 distribuidoras de energia elétrica, sendo 54 concessionárias e 38 permissionárias, além de 13 cooperativas de eletrização rural; em

2019. Em Santa Catarina, existe uma distribuidora (CELESC) e 20 permissionárias de energia.

As perdas de energia elétrica podem ser divididas em dois grupos: perdas técnicas e perdas não técnicas (perdas comerciais). As perdas técnicas estão relacionadas as perdas por efeito Joule, decorrentes das características físicas dos equipamentos e fios. Por outro lado, as perdas não técnicas, estão diretamente ligadas as fraudes e furtos de energia elétrica, a falta de manutenção nos medidores e a falta de manutenção nos equipamentos, estando diretamente associadas à gestão das concessionárias e as características sócio-econômicas das áreas de concessão. O foco deste trabalho é na detecção de perdas não técnicas, relacionada a fraudes nos medidores de energia elétrica. As perdas globais podem ser definidas como a diferença entre a energia elétrica despachadas nas subestações pelas distribuidoras e a faturada pelo seus consumidores.

No Brasil, de acordo com [1] as perdas globais (técnicas e não técnicas) representam aproximadamente 14% da energia total injetada no sistema elétrico, sendo um montante extremamente significativo e com grande potencial de redução. Minimizar as perdas, sob a ótica da concessionária de distribuição, significa não só dispor de uma parcela maior da energia para ser faturada, mas também, melhorar a qualidade do produto oferecido aos consumidores. Para o agente regulador e para a sociedade em geral, a redução das perdas representa a garantia de investimento na qualidade do produto, na manutenção do patrimônio da concessão e modicidade tarifária [3].

Segundo [4] o crime mais comum de fraude de energia se refere à adulteração do medidor, ocorrendo principalmente em instalações comerciais em geral, mas especialmente em instalações de maior consumo de energia tais como: padarias, postos de combustíveis, restaurantes, entre outros. Assim, por esta razão esta classe de consumidor foi escolhida para ser analisada neste trabalho.

## 3. Técnica de IA Aplicada

### 3.1 Aprendizagem de Máquina

A aprendizagem de máquina é uma área da Inteligência Artificial (IA) que tem como objetivo principal o desenvolvimento de sistemas inteligentes que melhorem os seus desempenhos, de forma automática, a partir de suas experiências [5]. A aprendizagem de Máquina é uma subárea da IA que busca estudar métodos computacionais para adquirir novos conhecimentos, novas habilidades e novos meios de organizar o conhecimento já existente.

De forma geral, as técnicas de aprendizagem de máquina podem ser classificadas em dois tipos: aprendizagem supervisionada e aprendizagem não supervisionada. A aprendizagem supervisionada, segundo [2], é baseada em uma estratégia que busca inferir uma função a partir de dados de treinamento rotulados. O algoritmo analisa os dados de treinamento e produz uma função inferida, podendo ser usada para mapeamento de novos exemplos. Alguns exemplos de algoritmos de aprendizagem supervisionada são: Support Vector Machine, Classificador Neive – Bayes, e *K-Nearest Neighbour*.

Por outro lado, a aprendizagem não supervisionada é baseada em uma estratégia para encontrar uma estrutura oculta em dados não rotulados. Considerando que os exemplos dados para o algoritmo não são rotulados, não há nenhum sinal de erro ou recompensa para avaliar uma solução em potencial [2].

### 3.2 Classificador SVM

Esta técnica de classificação, apresenta uma abordagem estatística de aprendizagem, trazendo resultados com base na observação e experiências realizadas com os dados, melhorando seu desempenho a medida que a quantidade de dados de entrada aumenta. Ressalta-se que o algoritmo, detecta o hiperplano de margem máxima, que é o limite de decisão com maior margem de separação dos dados, visando ter erros de generalização melhores do que aqueles com pequenas margens. [6].

A aplicabilidade do SVM consiste em, de acordo com [7], resolver problemas, tanto de classificação como de estimação. O tempo de treinamento do modelo pode vir a ser bastante lento, porém, esta técnica possui uma alta acurácia, sendo ela utilizada em aplicações como o reconhecimento de objetos e de fala. Estes autores ressaltam que no caso da possibilidade de separação linear dos dados, o SVM procura neste espaço o hiperplano de margem máxima. No caso em que os objetos não podem ser separados de forma linear, deve-se fazer uma transformação para um espaço de maior dimensionalidade, realizando um mapeamento não linear dos dados para o novo espaço. Em seguida, neste novo espaço é encontrado um hiperplano, o que leva a resolução do problema utilizando a abordagem inicial.

### 3.3 Software R

O software R possui duas grandes vantagens que o tornam bem popular, segundo [8]: a primeira é que ele é um software de código aberto, livre, grátis, ou seja, os usuários podem executar o programa como quiserem e adaptá-lo às necessidades. A segunda é que ele é altamente extensível, ou seja, pode ser utilizado para realizar qualquer atividade computacional, desde que compatível com suas capacidades. Isto é feito através da criação de funções próprias e dos pacotes, conjuntos de códigos/comandos relacionados a determinados temas.

Este software de programação contém diversos pacotes que auxiliam na criação de nossa *machine learning*, além do pacote do classificador *Support Vector Machine* (SVM), que foi o classificador selecionado para a classificação dos dados obtidos pela distribuidora. Entende-se pacote como uma coleção de funções e dados que ampliam o potencial do software R. [9].

## 4. Desenvolvimento

### 4.1 Base de dados e seleção dos consumidores

Os dados de potência ativa e reativa dos consumidores comerciais selecionados, medidos a cada intervalo de 15 minutos foram disponibilizados por uma concessionária de energia elétrica, para os doze meses do ano, gerando um total de 8760 dados para cada consumidor comercial. Entretanto, utilizou-se apenas a potência ativa. Estes dados também foram utilizados em [10], o qual fez uma aplicação em hardware (FPGA).

Nesta base de dados existiam 21 consumidores comerciais, localizados em uma mesma região do estado de Santa Catarina, envolvendo os seguintes ramos da economia: pesqueiras, madeireiras, têxteis, entre outras, conforme apresentado na Figura 1 abaixo. Para o estudo de caso foram selecionadas as empresas de comércio em varejo de madeiras, pois estas apresentavam a maior quantidade de consumidores (seis).

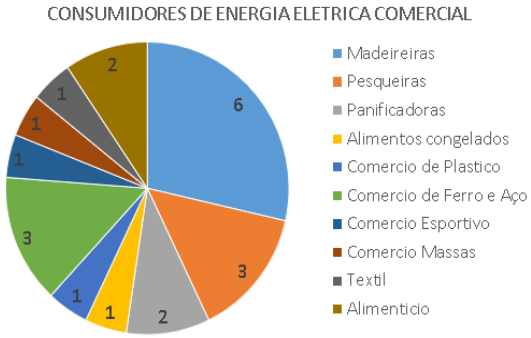


Figura 1 - Consumidores comerciais da base de dados

A partir dos dados de potência ativa diários dos consumidores selecionados, foram levantadas as curvas de carga diária no período de uma semana, buscando identificar padrões de consumo para os dias de semana destes consumidores. Este procedimento, que será explicado em detalhes na seção seguinte, se repetiu para todas as semanas do mês.

### 4.2 Modelagem das Curvas de Carga

Cada curva de carga dos consumidores selecionados possui 96 pontos de potência em kW (quatro medições por hora). Para que fosse possível encontrar padrões de consumo nas curvas de carga dos seis consumidores escolhidos (setor de venda de madeiras), todas as curvas de carga diárias foram comparadas, primeiramente para cada empresa, e posteriormente entre as madeireiras. Como se observou uma diferença entre o consumo nos dias úteis e fins de semana, optou-se por utilizar os dados apenas dos dias úteis.

Desta análise inicial foi possível inferir que a maioria das empresas apresentavam padrões de consumo similares em apenas sete meses do ano. Assim, foram retirados da análise os meses de janeiro, fevereiro, março, novembro e dezembro, os quais apresentaram consumo diferenciados, considerados atípicos.

A Figura 2 apresenta as curvas de carga das empresas selecionadas, considerando o primeiro mês de consumo típico. Nesta figura estão representadas uma curva de carga para cada dia do mês. Esta figura tem como objetivo apenas observar o comportamento das curvas de carga, ou seja, suas tipologias, não se preocupando com a magnitude de consumo de potência (eixo y). Neste caso, quando se analisa a tipologia das curvas de carga, o mais importante é a identificação dos vales, picos e rampas de aumento e diminuição de consumo. A questão da magnitude das curvas de carga não é importante neste momento, podendo ser considerada por meio de características funcionais das empresas.

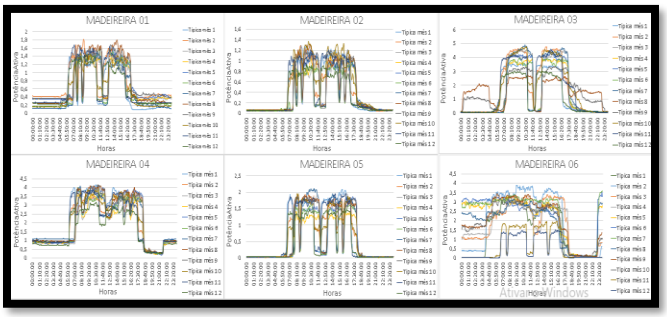


Figura 2 - Curvas de carga típicas consumidores selecionados

Como pode ser observado na Figura 2, duas empresas (madereiras 3 e 6) apresentaram um padrão de consumo diferenciado em relação às demais, ou seja, sua tipologia de curva de carga é diferente das demais empresas. Sendo assim, estas duas empresas foram retiradas da análise. Por outro lado, Empresas de diferentes CNAEs podem apresentar a mesma tipologia de curva de carga, o que não representa um problema para a identificação. O importante é que o classificador consiga identificar o padrão de cada consumidor, a partir dos dados de entrada apresentados na Figura 4.

Desta forma, obteve-se as curvas de carga típicas para o ano, considerando as quatro empresas e os sete meses do ano. Como pode-se observar na Figura 3, a tipologia da curva se mantém para os quatro consumidores, ou seja, o padrão de consumo é similar, variando apenas a magnitude do consumo, o que não vai interessar para o fim de classificação. Em futuros trabalhos, pretende-se considerar a magnitude de consumo, a partir da utilização de dados funcionais das empresas.

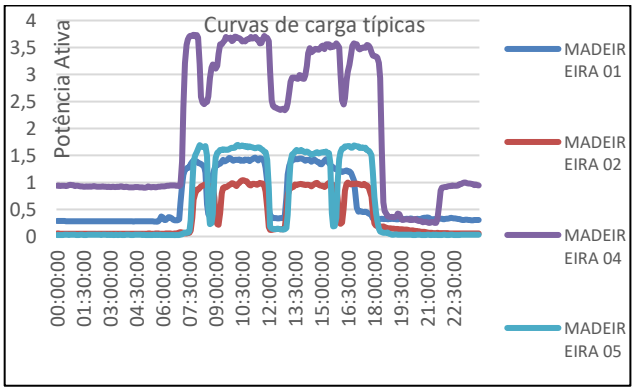


Figura 3 - Curvas de carga típicas - setor venda de madeiras

### 4.3 Seleção dos atributos de entrada

A definição dos atributos de entrada para o classificador SVM é uma importante etapa do processo de aprendizagem de máquina para identificação automática dos padrões atípicos de consumo de energia elétrica.

Inicialmente, utilizou-se apenas dados da curva de carga (96 pontos). Entretanto, observou-se que o desempenho do classificador não foi satisfatório. Neste sentido, foi necessário introduzir outros parâmetros, envolvendo informações adicionais da curva de carga (razão mínimo/máximo e fator de carga), além de informações do funcionamento da empresa (início e fim do expediente). O número do CNAE (Classificação Nacional de Atividades Econômicas) foi introduzido considerando a possibilidade de análise de outras atividades econômicas. A Figura 4 apresenta os atributos de entrada considerados.

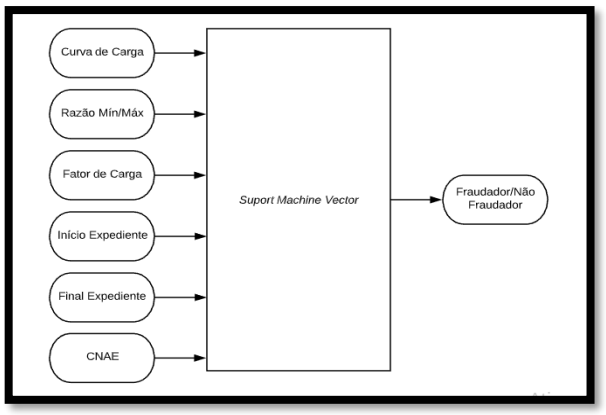


Figura 4 - Atributos de entrada do classificador SVM

Os atributos apresentados na Figura 4 são descritos abaixo:

Curvas de carga: 96 dados de potência ativa (kW).
Razão Mínimo/ Máximo: razão entre o consumo mínimo da curva de carga pelo consumo máximo.
Fator de Carga (FC): razão entre a demanda média e a demanda máxima da curva de carga.
Início e fim do expediente: dados funcionais da empresa.
Código CNAE: classificação de classes consumidoras segundo o IBGE [11]. Para o setor de madeireiras o CNAE é 4744-0/02.

A tabela 1, mostrada a seguir, apresenta para o mês 3, exemplos de valores destes atributos.

Tabela 1 - Exemplos atributos - consumidores típicos

| Consumidor | Razão Mín/Máx | Início Expediente | Final Expediente | Fator de Carga |
|---|---|---|---|---|
| | MÊS 03 | | | |
| Madeireira 01 | 0,18 | 7:00:00 | 16:00:00 | 0,44 |
| Madeireira 02 | 0,02 | 7:00:00 | 17:00:00 | 0,33 |
| Madeireira 04 | 0,08 | 8:00:00 | 18:00:00 | 0,51 |
| Madeireira 06 | 0,01 | 7:00:00 | 17:00:00 | 0,32 |

### 4.4 Definição dos consumidores atípicos

Como na base de dados disponibilizada não haviam consumidores classificados como fraudadores, foi necessário criar alguns consumidores atípicos para serem inseridos na base. Neste sentido foram considerados consumidores com curvas de carga, razão mín/máx, fator de carga e início e fim expediente diferenciados, para o treinamento e testes.

Com relação às curvas de carga foram considerados os dados dos consumidores 3 e 6, além de curvas de carga com redução de consumo na ponta e no vale. Para a razão mín/máx da curva de carga foram considerados valores atípicos os valores entre 0,01 a 0,06. O Fator de Carga típico deste segmento comercial está entre 031 e 0,52, segundo a Celesc e comprovados pelos cálculos realizados nas curvas de carga consideradas. Assim, considerou-se atípicos os FC abaixo de 0,3 e acima de 0,52. Na tabela 2 abaixo, são apresentados exemplos dos atributos para os consumidores atípicos criados.

Tabela 2 - Consumidores atípicos criados

| Fraudador | Fator de Carga | Consumo médio | Início Exp. | Final Exp. | Razão Mín/ Máx |
|---|---|---|---|---|---|
| 1 | 0,57 | 0,42 | 07:00:00 | 16:00:00 | 0,29 |
| 2 | 0,45 | 0,49 | 07:00:00 | 16:00:00 | 0,20 |
| 3 | 0,37 | 0,47 | 07:00:00 | 16:00:00 | 0,17 |
| 4 | 0,47 | 0,25 | 08:00:00 | 17:00:00 | 0,08 |
| 5 | 0,35 | 0,32 | 09:00:00 | 17:00:00 | 0,05 |
| 6 | 0,36 | 0,31 | 07:00:00 | 16:00:00 | 0,05 |
| 7 | 0,61 | 1,64 | 07:00:00 | 16:00:00 | 0,11 |
| 8 | 0,49 | 1,74 | 07:00:00 | 16:00:00 | 0,08 |
| 9 | 0,49 | 1,72 | 07:00:00 | 16:00:00 | 0,08 |
| 10 | 0,40 | 0,51 | 08:00:00 | 17:00:00 | 0,01 |
| 11 | 0,31 | 0,57 | 08:00:00 | 17:00:00 | 0,01 |
| 12 | 0,33 | 0,59 | 08:00:00 | 17:00:00 | 0,01 |

### 4.5 Método de treinamento e testes

O método de validação dos dados utilizado foi a validação cruzada (*Cross Validation*), a qual é uma técnica utilizada para avaliação de desempenho de modelos de aprendizagem de máquina, onde um conjunto é utilizado para treino e outro conjunto é utilizado para teste e avaliação do desempenho do modelo.

Um dos métodos de aplicação do *Cross Validation*, é o K-fold, o qual consiste em dividir a base de dados de forma aleatória em K subconjuntos com aproximadamente a mesma quantidade de amostra de cada um deles. Para cada iteração, treino e teste, o conjunto formado por K-1 subconjuntos são utilizados para treinamento e o restante utilizado para teste e avaliação do desempenho do modelo. Esta técnica garante que cada subconjunto será utilizado para teste em algum momento da avaliação do modelo.

Neste trabalho foi utilizado o método 7-*fold*, sendo que os dados foram divididos em 7 partes, onde em cada iteração é utilizado seis partições dos dados para treinamento e a partição restante para teste. Na segunda iteração, a segunda partição é utilizada para teste enquanto as demais para treino. Este processo foi repetido 7 vezes conforme apresentado no Quadro 1, até que toda a base de dados passe pelas etapas de treino e teste gerando uma métrica de desempenho para o modelo. O 7-fold foi escolhido em função do número de sete meses que foi selecionado para os dados dos consumidores.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Iteração 01 | teste | treino | treino | treino | treino | treino | treino |
| Iteração 02 | treino | teste | treino | treino | treino | treino | treino |
| Iteração 03 | treino | treino | teste | treino | treino | treino | treino |
| Iteração 04 | treino | treino | treino | teste | treino | treino | treino |
| Iteração 05 | treino | treino | treino | treino | teste | treino | treino |
| Iteração 06 | treino | treino | treino | treino | treino | teste | treino |
| Iteração 07 | treino | treino | treino | treino | treino | treino | teste |

Quadro 1 - Processo de validação cruzada 7-*fold*

### 4.6 Funções R utilizadas

Com relação ao software R, foram utilizados os seguintes pacotes: install.packages("tidyverse"); install.packages("e1071"); install. packages("caret"); library(tidyverse); library(e1071); library(caret); library(readr); library(readxl).

Para criação da base de dados no R foi utilizado o seguinte exemplo de comando para importação dos dados do Excel: MES03 <- read_excel("C:/Users/Desktop/Dados/2020-2/TCC3/DADOS/AnaliseGrafica.xlsx", sheet = "M03", col_names = c("CONSUMIDOR","POTENCIA_ATIVA" ,"FATOR_CARGA","CNAE","RAZAO_MAXMIN","INICIO_EXPEDIENTE","FINAL_EXPEDIENTE","POTENCIA_MEDIA_MENSAL","FRAUDOUNFRAUDE"),col_types = c("text", "numeric","numeric", "numeric","numeric","date", "date", "numeric", "text")).

Para o treinamento e testes, foram utilizados os seguintes comandos:

```
metodotreino <- trainControl(method = "cv", number = 7)
meumodelo <- train(BASEDEDADOS$FRAUDOUNFRAUDE~., data = CURVADECARGA,trControl = metodotreino, method = "svmLinear")
```

```
meumodelo <- train(FRAUDOUNFRAUDE~., data = BASEDEDADOS,trControl = metodotreino, method = "svmLinear")
```

Para a matriz de confusão:

```
pred01 <- predict(meumodelo, BASEDEDADOS)
tab01 <- table(Previsto = pred01 ,Atual = BASEDEDADOS$FRAUDOUNFRAUDE)
```

## 5. Resultados

Inicialmente foram escolhidas aleatoriamente quatro curvas de carga para cada mês, para cada um dos quatro consumidores selecionados do setor de comércio de madeiras. Como temos sete meses de dados considerados, foram 112 dados de consumidores não fraudadores. Foram acrescidos a este conjunto, os dados criados de 84 consumidores atípicos, utilizando-se o procedimento explicado na seção 4.4 deste artigo, totalizando-se assim 196 dados, com os atributos descritos na Figura 4. O processo de validação cruzado foi aplicado a este conjunto de dados, gerando a matriz de confusão apresentada no Quadro 2 apresentado abaixo.

Quadro 2 - Matriz de confusão 1

| | | Real | |
|---|---|---|---|
| | | 0 | 1 |
| Previsto | 0 | 112 | 7 |
| | 1 | 0 | 77 |

Os *label* "0" e "1" foram considerados como não fraudadores e fraudadores, respectivamente. As métricas de desempenho foram calculadas pelo software R e estão apresentadas abaixo:

Precisão = 94 %
Acurácia = 96%
Recall = 100%
Estatística Kappa = 93%

Como pode ser observado no Quadro 2, sete consumidores foram considerados fraudadores, mas o classificador considerou como não fraudador, caracterizando como falsos negativos.

Na sequência foi realizado um novo teste considerando outros 20 consumidores típicos (não fraudadores), de forma aleatória e diferentes dos anteriores, e onze consumidores atípicos, também com atipicidade diferentes dos anteriores. Os resultados obtidos deste teste estão apresentados no Quadro 3.

Quadro 3 - Matriz de confusão 2

| | | Real | |
|---|---|---|---|
| | | 0 | 1 |
| Previsto | 0 | 20 | 2 |
| | 1 | 0 | 9 |

Na matriz de confusão apresentada no Quadro 3, pode-se constatar que o classificador considerou dois consumidores como não fraudadores, mas eles eram na realidade, fraudadores, caracterizando como dois falsos negativos. Conforme pode ser observado nos resultados apresentados nos quadros 2 e 3, não se teve caso de falso positivos, o que é mais interessante para a aplicação real, pois para uma concessionária de energia o pior é considerar um consumidor como fraudador, enquanto que uma inspeção posterior comprova que o mesmo não é. As métricas de desempenho neste teste estão apresentadas abaixo:

Precisão = 91 %
Acurácia = 94%
Recall = 100%
Estatística Kappa = 92%

### 6. Comentários finais

O grande desafio da modelagem do classificador para este caso foi a definição dos atributos de entrada. Quando somente a curva de carga foi utilizada como atributo de entrada, o classificador não conseguia identificar um consumidor atípico. Entretanto, quando se colocou outros parâmetros como atributos de entrada, além da curva de carga, incorporando mais informações sobre a mesma, conseguiu-se uma melhora substancial no desempenho do classificador. Neste caso, a utilização do fator de carga como atributo de entrada foi extremamente positivo. O código CNAE, apesar de não apresentar diferença para a situação estudada, pois todas as empresas pertenciam ao mesmo ramo de atividade econômica, deverá ser importante quando se considerar consumidores de outros segmentos econômicos.

Nos casos testados, o desempenho do classificador foi bem positivo, conseguindo-se precisão e acurácia superiores a 90%. Naturalmente, quando se considerar consumidores de outros setores da economia, este desempenho tende a piorar um pouco. Entretanto, resultados preliminares apontam que o classificador SVM desenvolvido é promissor para o problema, considerando os atributos de entrada selecionados.

O problema de identificação automática de consumo atípico, sinalizando uma possível fraude, é um tema que vem a reboque dos conceitos associados as redes elétricas inteligentes e medidores inteligentes, sendo importante para a redução das perdas comerciais de energia elétrica. As perdas comerciais é uma perda econômica não somente para a concessionária de energia, mas também para a sociedade como um todo, a qual tem que se submeter a aumentos tarifários para compensar esta redução de receita por parte da distribuidora. As técnicas de aprendizagem de máquina e o classificador SVM são uma boa proposta para resolver este problema, podendo ser utilizado junto ao banco de dados da concessionária , ou de forma descentralizada nos medidores inteligentes.

## Referências

[1] AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **Perdas de Energia Elétrica na Distribuição.**2019.Disponível em: <https://www.aneel.gov.br/documents/654800/18766993/Relat%C3%B3rio+Perdas+de+Energia_+Edi%C3%A7%C3%A3o+1-2019-02-07.pdf/d7cc619e-0f85-2556-17ff-f84ad74f1c8d> . Acesso em: 27 fev. 2020.

[2] RAMOS, Caio Cesar Oba. **Caracterização de Perdas Comerciais em Sistemas de Energia Através de Técnicas Inteligentes.** 2014. 144 f. Tese (Doutorado) – Curso de Ciências da Computação, Escola Politécnicas da Universidade de São Paulo, São Paulo, 2014.

[3] STRAUCH, M.T. **Desenvolvimento de metodologia para cálculo de perdas elétricas em redes de distribuição de baixa tensão.** Dissertação de mestrado – Universidade Salvador, 2002. Disponível em:< http://biblioteca.unifacs.br/bitstream/tede/355/1/Dissertacao%20Mariana%20Torres%202002%20texto%20completo.pdf> Acesso em : 22 abr 2020

[4] CELESC. "**Gatos" e Fraudes.** 2019. Disponível em: <https://omunicipio.com.br/celesc-acusa-empresas-da-regiao-por-furto-de-energia-eletrica/>. Acesso em: 03 fev. 2020.

[5] FERREIRA, Hamilton Melo. **Uso de Ferramentas de Aprendizado de Máquina para Prospecção de Perdas Comerciais em Distribuição de Energia Elétrica.** 2008. Dissertação de mestrado – Universidade Estadual de Campinas Faculdade de Engenharia Elétrica e de Computação, Campinas (SP) – Brasil. 2008. Disponível em:<http://repositorio.unicamp.br/bitstream/REPOSIP/259083/1/Ferreira_HamiltonMelo_M.pdf> . Acesso 23 fev. 2020.

[6] TAN, Pang-Ning; et. al. ***Introdution to Data Mining:*** *Second Edition*.2018. Disponível em: < https://www-users.cs.umn.edu/~kumar001/dmbook/index.php >. Acesso em: 05 jun. 2020.

[7] HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data mining**: **concepts and techniques**. 3a edição. Whaltman: Editora Morgan Kaufmann, 2011.

[8] RITTER, Matias; THEY,Ng Haig; KONZEN, Enéas. **Introdução ao software.** 2019. Versão: 2.0. UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL,2019. Disponível em:< http://professor.ufrgs.br/sites/default/files/matiasritter/files/apostila_introducao_ao_r_-_ritter_they_and_konzen.pdf> Acesso em: 22 Jun.2020.

[9] SCHMULLER, Joseph. **Análise estatística com R para leigos**. Rio de Janeiro, Alta Books, 2019. Disponível em: <https://www.ufrb.edu.br/ccaab/images/AEPE/Divulga%C3%A7%C3%A3o/LIVROS/An%C3%A1lise_Estat%C3%ADstica_com_R_para_Leigos_-_2%C2%AA_Edi%C3%A7%C3%A3o_-_Joseph_Schmuller_-_2019.pdf>. Acesso em: 20 fev. 2020.

[10] ABRAHAM, Guilherme; SIMAO, Jorge G. S.; TEIVE, Raimundo C. G. Identificação de Fraudes de Energia Elétrica em Consumidores Comerciais - Uma Aplicação voltada aos Medidores Inteligentes. **Anais do IX Computer on the Beach - COTB**. Florianópolis. 2018.

[11] CNAE 2.0 - Classificação Nacional de Atividades Econômicas. https://CNAE.IBGE.GOV.BR/classificacoes. 2006.