Fraud Detection in Electric Power Distribution: An Approach that Maximizes the Economic Return.

Pablo Massaferro*, J. Matías Di Martino*†, and Alicia Fernández*
*Universidad de la República, Montevideo, Uruguay.
† Duke University, North Carolina, USA.
{pmassaferro, matiasdm, alicia}@fing.edu.uy

***Abstract*—The detection of Non-Technical Losses is a very important economic issue for Power Utilities. Diverse machine learning strategies have been proposed to support electric power companies tackling this problem. Methods performance is often measured using standard cost-insensitive metrics such as the accuracy, true positive ratio, AUC, or F1. In contrast, we propose to design a NTL detection solution that maximizes the effective economic return. To that end, both the income recovered and the inspection cost are considered. Furthermore, the proposed framework can be used to design the infrastructure of the division in charge of performing customers inspections. Then assisting not only short term decisions, e.g., which customer should be inspected first, but also the elaboration of long term strategies, e.g., planning of NTL company budget. The problem is formulated in a Bayesian risk framework. Experimental validation is presented using a large dataset of real users from the Uruguayan Utility (UTE). The results obtained show that the proposed method can boost companies profit and provide a highly efficient and realistic countermeasure to NTL. Moreover, the proposed pipeline is general and can be easily adapted to other practical problems.**

***Index Terms*—economic return, non-technical losses, electricity theft, automatic fraud detection, example-cost-sensitive.**

I. Introduction

Dishonest customers perform different and ingenious fraudulent mechanisms to steal electric power and reduce their bills. Power distribution companies lose a substantial amount of revenue due to this problem which directly impacts on countries economy [1], [2]. This harms specially developing countries where due to a complex combination of social, economical, and cultural factors, the economic losses associated to electric fraud are even more significant. In India, for example, non-technical losses (NTL) are estimated at $4.5 billion. In Brazil, Malaysia and Lebanon NTLs represent up to the 40% of the total electricity distributed [3], [4]; while in the UK and USA, non-technical losses are estimated between 1 and 6 billion US dollars [3], [5].

Diverse maneuvers are developed by thieves to steal electricity. For example, magnets are sometimes attached to the electromechanical meters to slow down their reading; another common technique is to create an electric bypass between the input and output of the electric meter. A different modality consists of connecting users to a distribution line or transformer, bypassing all the existing meters. This second example is different in nature to the first one, in particular, because these users do not have an active contract with the distribution company, nor a history of power consumption. In the present work we focus on the first type of frauds. We assume that some kind of consumption reading is being performed (with a certain periodicity) and we analyze the history of these readings to automatically detect suspicious and potentially illegal activities.

The present work is developed in Uruguay as part of an existing collaboration between the University "Universidad de la Republica" and UTE (the national company in charge of the power generation and distribution). In Montevideo (capital of Uruguay) NTL represent approximately 13% of the total energy distributed. Electric companies perform regular customer inspections to detect and prevent NTL and fraudulent activities. Obviously, performing such controls have an associated economic cost. In addition, it is economically and timely intractable to visit all the users and therefore, the optimal number of inspection must be defined.

The goal of this work is to developed a machine learning solution, that based on customer consumption profiles and utility cost, produces an optimal reduced list of clients to be inspected. The optimization consists of the maximization of the company economic return. In other words, we aim not only at detecting those customers that commit fraud, but also, at providing an inspection priority list containing those customers for which the economic return is potentially larger.

The main contributions of this work are: (i) to present an approach for NTL detection conceived to maximize the economic return. To the best of our knowledge, this is the first work that provides such a solution in the context of NTL. (ii) To propose a flexible method that can be optimized for non trivial and realistic cost models, therefore, our solution can be used in practice as an input for managers to make long and short-term decisions. (iii) To contextualize the proposed solution in a Bayesian-Risk framework, which would inspire researchers of other fields to easily adapt our ideas to their specific discipline.

And (iv), to study and compare our solution with other cost-based classification approaches.

Section 2 describes NTL problem and the proposed solution. Then Sec. 3 provides additional implementation details and experimental validation. Finally Sec. 4 concludes this work and summarizes the main results.

Related Work

Different machine learning based approaches have been proposed in the past twenty years for NTL detection. Glauner et al. [3] present a review of the most relevant work published until 2017. Recently, Messinis and Hatziargyriou [6] presented an exhaustive survey including network oriented, hybrid and novel data oriented methods.

Supervised approaches build and learn mathematical models that describe the problem based on labeled datasets provided by power distribution companies. Several works explore the use of support vector machines (SVM) algorithm [4], [7], [8], [9] or combinations of SVM with other methods such as the Genetic Algorithm [10]. Other classification strategies have been explored as well, e.g., Neural Networks classifiers [11], k-Nearest neighbor [8], or Optimum path forest (a graph-based classifier) [12].

In recent years, the increasing shift from electromechanical and electronics meters to smart meters is originating novel countermeasures to the problem of NTL. For example Zeng et al. [13] used smart meters data to train deep neural networks considering daily consumption inputs at the top layer. They propose a Wide & Deep Convolutional Neural Network (CNN) model with two components. The Deep CNN component identify the non-periodicity of electricity-theft and the periodicity of normal electricity usage based on two dimensional (2-D) electricity consumption data. Meanwhile, the wide component captures global features of 1-D electricity consumption data. Hu et al. [14] propose a semi supervised deep-learning-based fraud detection model to handle high dimensional and unlabeled data as input.

The transition to Advanced Measurement Infrastructure (AMI) also generates new types of fraud, such as cybertampering. Several works have been carried out to face this problem. Guo et al. [15] propose an online data validation framework to verify household energy meters in a secondary network with real-time measurements from the remote terminal units of the feeder in the primary network. Other emerging threat is related to the malware on IP-based smart meters. In 2015, Guo et al. [16] use Markovian Decision Process to define a preventive maintenance strategy in order to control a malware propagation over the AMI.

Complementary work focus on engineering proper features to represent customer consumption profiles in a convenient vector space. For example, Fourier coefficients, local averages and categorical information such as: the meter type, history of theft, or credit worthiness proved to provide useful information for NTL detection [1], [7], [17], [18]. Recently, Glanuer et al. [19] included the use of neighborhood local features. In a similar direction, Massaferro et al. [20] propose a method to find an optimal grid for the development of adaptive geographical features.

Related works highlights the imbalance nature of fraud detection problem [17] and the importance of the metric used to asses the performance of NTL detection. To address classes unbalance Avila et al. [21] propose a random under sampling boosting strategy, using the area under the receiver operating curve (AUC) and the Matthews correlation coefficients (MCC) as performance metrics.

In contrast with the approaches described above, the focus of the present work is not on the circumstantial features or classifier being used. A recent review on financial fraud detection asseverates: *"The expense of a false positive in miss-classifying a legitimate transaction as fraud, is typically far less than that of a false negative [22]. Insufficient study has been performed on the disproportionate nature of these costs, with attention typically focusing on the traditional classification performance methods"*[23]. The present work is a contribution aiming specifically at reducing this gap by considering the associated economical cost of false negative and false positive cases.

We aim at designing a machine learning solution that optimizes companies economic return. Similar approaches were followed on different areas. For example, Bahnsen et al. [24] present a comparison of a set of classification approaches taking into account the monetary loss for credit card fraud detection.

II. Proposed approach

Let x_i represent a column vector with the feature values associated to the i-th sample, and y_i its label. For example, in the context of NTL, x_i can represent the history of monthly consumption, or the concatenation of the history of consumption and additional numerical features (e.g., the geographic coordinates of the customer associated to the sample) [20]. We focus on a binary classification problem where $y_i \in \{-1, 1\}$. The label $y_i = 1$ (named the positive class) is associated to a fraudulent behavior while the label $y_i = -1$ (negative class) is associated to a normal customer.

When the posterior probabilities $P(y_i = 1|\mathbf{x_i})$, and $P(y_i = -1|\mathbf{x_i})$ are available for a given sample $\mathbf{x_i}$, the classification criteria that maximizes the accuracy over a given set $X = \{\mathbf{x_1}...\mathbf{x_n}\}$ is:

$$\hat{y}_i = \underset{\tilde{y}}{\operatorname{argmax}}\{P(\tilde{y}|\mathbf{x_i})\}. \quad (1)$$

$\hat{y}_i$ denotes the predicted label while y_i represents the ground truth label (available or not). The previous classification rule is known as the Bayes minimum error approach [25]. It is easy to prove that this strategy leads to an optimal classification solution in terms of minimizing the mean classification error.

Although the previous strategy may seem appealing, it suits problems with balanced classes. Unpractical results are obtained

when it is applied for NTL. In particular, because NTL problem is very unbalanced (only a minor percentage of the total customers pursuit fraudulent activities [3], [5], [17]). Imagine for instance that only 1% of the customers are committing fraud. Then, a trivial classifier that predicts always the negative class would achieve 99% accuracy, despite that it does not provide any benefit for detecting fraud.

As an alternative, we are interested in minimizing the financial loss taking into account: (a) the cost of performing individual inspections and (b) the harm of not detecting a fraudulent case. This information could be used to define whether the number of inspections carried out is sufficient, or if re-allocation of resources is necessary in a division of the company.

Maximizing the economic return. Let m denote the number of inspections to be performed and $X_m \subset X$ an arbitrary subset of m samples of X. As before, $P(y_i = 1|\mathbf{x_i})$ denotes the probability that a given sample $\mathbf{x_i}$ is committing fraud, a_i represents the amount of money the i^{th} customer could potentially be stealing (if it does), and c_i the cost of inspecting the i^{th} customer. Given the previous definitions our approach consists of obtaining the optimal subset $\hat{X}_m = \{\mathbf{x_{i_1}}, \ ..., \ \mathbf{x_{i_m}}\}$ such that

$$\hat{X}_m = \underset{X_m}{\operatorname{argmax}} \left\{ \sum_{k=1}^{m} a_{i_k} P(y_{i_k} = 1|x_{i_k}) - \sum_{k=1}^{m} c_{i_k} \right\}. \quad (2)$$

The previous optimization scheme can be related to a Bayesian-Risk formulation of the problem [26] [27]. A cost sensitive classification loss can be expressed as

$$L(\mathbf{x}, q) = \sum_{k} P(y = k|\mathbf{x})\mu_{qk}. \quad (3)$$

μ_{qk} represents the cost associated to predict as q a member of the class k [26]. In the particular case of binary classification, μ_{11}, μ_{00}, μ_{01} and μ_{10} are associated to the true positive, true negative, false negative, and false positive costs respectively.

An optimal decision in terms of the loss $L(\mathbf{x}, \mathbf{q})$ leads to the classification rule:

The optimal prediction is the positive class if and only if the expected cost of this prediction is less than or equal to the expected cost of predicting the negative class, i.e., if and only if $P(y = -1|\mathbf{x})\mu_{10} + P(y = 1|\mathbf{x})\mu_{11} \leq P(y = -1|\mathbf{x})\mu_{00} + P(y = 1|\mathbf{x})\mu_{01}$.

The proposed solution can be seen as the application of the previous rule when $\mu_{10}(\mathbf{x_i}) = c_i$, $\mu_{11}(\mathbf{x_i}) = c_i - a_i$ and $\mu_{00}(\mathbf{x_i}) = \mu_{01}(\mathbf{x_i}) = 0$. Substituting the weight defined above and using the property $P(y = 1|x) + P(y = -1|x) = 1$ the classification rule can be simplified as: x should be classified as positive if the expected return surpass the inspection cost, i.e., $c_i < a_i P(y_i = 1|x_i)$.

Implementing a solution for NTL. Equation (2) is intuitive and mathematically express the goal of maximizing the economic return. Three crucial aspects needs to be addressed before it is practically applicable. For instance, we need to estimate: (i) the a-posterior class probability $P(y_i = 1|\mathbf{x_i})$, (ii) the amount of potential fraud a_i, and (iii) the optimum number of inspections m_{op}. In the following we address these definitions, and we propose a practical solution to estimate these quantities with the information accessible in the context of NTL detection.

*1) **Empirical estimation of the a-posterior class probability**:* A naive approach would be to exploit training data $(\mathbf{x_i}, y_i)$ to directly estimate the density function $P(y = 1| \ \cdot \ )$, for example, applying non-parametric kernel-based method [28]. This family of methods are intractable in several practical applications, specifically when the dimension of the feature space is moderate or large [29]. Certainly, it is not the right choice in the context of NTL where the dimension of the feature space is moderately larger [17], [7], [3], [30].

Inspired by the work of Zadrozny and Elkan [31] we propose to estimate the a-posterior probability of fraud in two steps. First, a classifier method is trained to estimate a score function $s(\mathbf{x})$. Then a calibration mapping function $g : [0, \ 1] \rightarrow [0, \ 1]$ is defined such that $P(y = 1|\mathbf{x}) = g(s(\mathbf{x}))$.

Classification algorithms like SVM, random forest, or neural networks are extremely efficient at learning the distribution of different classes in the feature space. Most of these techniques provide as a result a score function $s(\mathbf{x})$. When $s(\mathbf{x}) \approx 0$ the sample $\mathbf{x}$ very likely belong to the negative class, while on the other hand, $s(\mathbf{x}) \approx 1$ indicates that the sample likely belong to the positive class. Then, classification of new input samples can be performed according to: *if* $s(\mathbf{x}_i) < \lambda$ *then* $\hat{y}_i = -1$, *otherwise,* $\hat{y}_i = 1$. Most methods set by default $\lambda = 0.5$, but is also common to tune this threshold to maximize a particular performance measure (e.g., the Accuracy, F1 or AUC).

Although the score $s(x)$ provides valuable information for classification, it can not be directly interpreted as the chance of membership to a given class. For example, $s(\mathbf{x}) = 0.2$ does not necessary imply that $P(y = 1|\mathbf{x}) = 0.2$.

Score calibration can be defined as the task of obtaining a calibration function $g : [0, 1] \rightarrow [0, 1]$ such that $P(y = 1|g(s(\mathbf{x})) = s)$ converges to s as the number of samples goes to infinity [31]. One of the most used techniques on probability calibration is Platt Scaling originally proposed by John Platt to calibrate SVM [32]. This is a parametric method based on adjusting the likelihood through logistic regression as describes Eq. (4)). Parameters A and B are learned in a supervised fashion using available training data.

$$P(y = 1|x) = \frac{1}{1 + e^{As(x)+B}} \quad (4)$$

An extension of the previous technique was proposed for other algorithms. It was recently proved that a simplified extension called Temperature Scaling, is the simplest and most efficient method to calibrate the output of neural networks [33].

In the present work, Platt Scaling calibration method is compared with a non-parametric approach based on isotonic regression. Then calibration is performed using the Platt Scalling implementation included in the *scikit-learn* library for the score output of Random Forest and SVM algorithms. For NN we implement the Temperature Scaling by adding an extra layer before the finally SoftMax activation function (to minimizing the binary cross-entropy).

*2) **Estimation of the potential fraud loss**:* In order to estimate the volume of customers potential fraud a_i, we propose two alternatives. The first idea uses ubiquitous information in the context of NTL, requiring only the knowledge of customers maximum (peak) contracted power. The second alternative, uses the records of economic return obtained for fraudulent customers inspected in the past.

a) Fraud estimation using exclusively billing information: A typical domestic installation includes in addition to the power meter, a switchgear that limits the maximum power that can be consumed from the power grid (and protects the electrical infrastructure in case of failures). The maximum power contracted is ubiquitously accessible in the context of NTL.

Based on empirical observation of the data, we make the following assumption: fraudulent customers reduce their electrical bills an amount that is approximately proportional to their actual consumption, i.e, $a_i \propto e_i.$, where e_i represents the total amount of energy (in kWh) the customer actually consumes. We empirically observed that e_i has a strong correlation with the value of the maximum (peak) contracted power Mp_i.

Specifically, we have observed that the average energy consumed is approximately proportional to the maximum contracted power Mp_i. This is an interesting observation from a practical perspective, as the latter is available information (while the former is strictly unknown in the context of fraud detection). Therefore, the potential magnitude of fraud can be estimated as $a_i \propto Mp_i$.

b) Fraud estimation using actual records of economic return.: An alternative approach for the estimation of customers potential theft is to consider this problem also as a supervised learning problem. In other words, if the economic harm can be retrieved in practice (after actual inspections are performed) we can use this information to predict $a_i(x)$.

We formulate this as a regression problem where using customers information x, we predict the amount of associated potential fraud a_i. To this end, we collected real measures of the economic loss associated to cases of fraud. About 50 thousand inspections were performed in 2017 and the information associated to three thousand cases of fraud analyzed.

Random forest, SVM and a neural network where considered for numerical solution of the regression problem. These algorithms are trained on fraudulent cases using the real economic harm as the target output.

*3) **Income and Cost: Defining the optimal operating capacity**:* The optimal number of inspection that should be performed is determined by the expected amount of economic gain versus the cost associated to perform the inspections. These balance between gain and cost can be formulated in the framework described before,

$$\hat{X}_m = \underset{N, X_m}{\operatorname{argmax}} \left\{ \sum_{k=1}^{m} a_{i_k} P(y_{i_k} = 1 | x_{i_k}) - \sum_{k=1}^{m} c[N, k] \right\}, \quad (5)$$

$c[N, k]$ denotes the cost of performing the kth inspection when the infrastructure is designed to perform a nominal number of N inspections. In the following we address how to define realistic cost models $c[N, k]$.

Cost model: Defining the number of inspections that needs to be performed helps to establish and design the operational infrastructure. A realistic cost model must include at least one fixed component and one variable. Let us recall standard definitions in microeconomics. The Marginal Cost (MC) represents the derivative of the cost function with respect to the quantity, and the Average Cost (AC) the total cost divided by the number of units. In the present work we use a cost curve designed to an operating capacity of N inspections, we assume the fixed cost to be proportional to N and an extra cost when operating over the designed capacity:

$$c[m] = \begin{cases} \alpha m + \alpha N \frac{\gamma}{1-\gamma} & m \leq N \\ \beta m + \alpha N \frac{1}{1-\gamma} & m > N. \end{cases} \quad (6)$$

$\gamma = c[0]/c[N]$ sets the fixed cost, α/β the marginal cost below/above the designed capacity respectively. A smooth approximation of the previous model is also considered as illustrated in Fig. 1 (third-order polynomial approximation).

In addition to the optimal list of users to be inspected X_m, we consider now the nominal number of inspections N also as an optimization parameter. As described before, a realistic model for the inspection costs is variable and depends of the nominal capacity N set by the company infrastructure.

The components of Eq. (5) associated to the potential gain $a_{i_k} P(y_{i_k} = 1 | \mathbf{x_{i_k}})$ are independent of the cost curve. Therefore, we can primarily rank the estimated potential return of each customer. Then, the optimal number of inspections m can be computed given the nominal value N.

An important practical advantage of the proposed approach is that the more complex learning task (associated to estimating $a_{i_k} P(y_{i_k} = 1 | \mathbf{x_{i_k}})$) is performed only once, and thus several cost models can be tested efficiently. In contrast, algorithms that take into account the cost of each sample internally, e.g., as proposed in [34], need to be trained from scratch for every cost model leading to time consuming training routines.

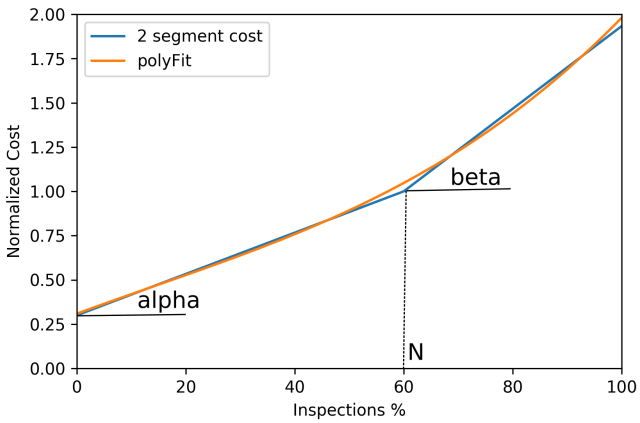


Fig. 1: Inspection costs assuming the infrastructure is planned for a nominal number of N inspections. The cost model includes: fixed, variable and extra costs. A continuous smooth approximation of the two segment model is also considered.

III. Experiments

A. Data

Two data sets are considered for experimental validation and analysis: NTL_10K_S and NTL_50K_R.

NTL_10K_S dataset is composed of the historical monthly consumption of ten thousand customers distributed across Montevideo (capital of Uruguay). To provide a first round of experiments in fully controlled conditions, synthetic frauds are simulated over real consumption of this set of clients. Therefore, the actual ground truth economic recovery can be accurately measured over this set. Random fraudulent cases are simulated in 10% of the samples uniformly distributed over time. The stolen percentage of energy is also randomly assigned.

NLT_50K_R dataset consists of fifty thousand customers in Montevideo inspected over 2017. The portion of fraudulent customers is approximately 6.0%. This data set includes the historical energy consumption and additional features such as the contracted peak power and the geographical coordinates. After fraudulent customers are detected, fines and economic penalties are implemented (whose value depend on the company estimation of the magnitude of the fraud). We use this financial information to estimate ground truth values of the economic return associated to each fraudulent customer.

B. Implementation Details.

A span of three years of monthly electrical consumption is used as feature. Experiments over *NLT_50K_R* data set also include additional features such as the record of previous irregularities and geographical coordinates. A detailed analysis of these features is provided in [20] where we focus on the impact of considering different subset of features for NTL. Algorithms are trained using 70% of the data while the other 30% is isolated for testing. This partition is randomly performed ten times, the average test performance over these experiment is reported. Classification and regression algorithms are selected from Python libraries: *sikit-learn*, *Keras*, and *Tensorflow*.

a) Training: Parameters (C, γ) for SVM and $(n\_estimators, max\_features)$ for Random-Forest are obtained by performing a logarithmic grid search of the maximum value of AUC (Area Under Receiver Operating Curve). The fully connected network considered has 3 hidden layers with 100 neurons per layer, ReLU (Rectified Linear Unit) is used as activation function. The network is optimized using Stochastic Gradient Descent (SGD) and the crossentropy is defined as loss function. The final activation is a sigmoid unit which value is considered to estimate the probability of fraud for each sample. Temperature-Scaling is adjusted during the training process. For the case of SVM and Random-Forest, score outputs are calibrated using Platt-Scalling procedure.

In addition an experiment is presented where the incidence of the calibration method in the results is evaluated. PlattScaling and Isotonic Regression algorithms are compared using the implementation contained in the *scikit-learn* library.

To estimate the volume of potential theft a_i by means of regression, SVR, Random-Forest-Regressor and a neural network are considered.

The NN used for regression has the same architecture of the classification network described before. The mean square error is considered as the optimization loss and the economic return as the output. To train the Random-Forest-Regressor four hundred trees are used with the minimum square error criterion. Each tree uses all the features but the samples are randomly chosen with replacement. For SVR, an RBF kernel is considered ($\gamma = 0.01$) and the error penalty set to $C = 1$.

b) Optimization of the number of inspections: Once the probability of fraud p_i and the estimation of the amount of economic harm a_i is obtained for each customer i, the optimal number of inspections m can be computed given a cost curve $c_N(m)$. $c_N(m)$ denotes the total cost of performing m inspections given a nominal capacity N. Algorithm 1 summarizes the main steps involved in the determination of the optimal number of inspection and the list of customers to be inspected.

In addition, the previous method can be used to estimate the optimal capacity of the division in charge of fraud detection. For example, different values N can be tested (defining a family of cost functions c_N). Moreover, N can also be chosen in order to

Algorithm 1 Estimate the optimal number of inspections m and the list of customers to be inspected X_m.

Input: Input: trained methods f and g for the estimation of $P(y=1|x)$ and $a(x)$ respectively, and cost profile c_N.

1: **for** each sample x_i **do**
2: $p_i = f(x_i)$ prob. of fraud.
3: $a_i = g(x_i)$ potential harm.
4: $G = [p_1 a_1 \ ... \ p_M a_M]$ Vector of potential gain per sample
5: $[G, ind] = \text{sort}(G, \text{descending})$
6: $m = \text{argmax}_k \left(\sum_{i=1}^{k} G[i] - c_N[k] \right)$
7: $X_m = ind[1:m]$ Set of customers to be inspected
8: **return** X_m, m

maximize the overall gain,

$$N_{op} = \text{argmax}_N \left(\sum_{i=1}^{m(N)} G[i] - c_N[m(N)] \right). \quad (7)$$

As in Alg. 1 G_i represent the potential gain associated to each individual customer (in descending order). As we will show in the following experiments, the expression given in Eq. (7) can be empirically evaluated to find the optimal capacity of operation and the operation point associated to it.

C. Results on NTL_10K_S data set

Three fraud detection strategies are compared. As baseline we consider a solution that maximizes the F1 measure. This criterion aims to find an optimal balance between classification Recall and Precision, for unbalanced problems, this is one of the many adequate measures that can be considered [6]. A second solution is provided estimating the potential theft associated to each customer a_i from its maximum contracted power M_p. The third solution is obtained by estimating customers potential gain a_i as a supervised regression problem as described in Pag. 4.

Let us denote the solutions described above as:

- SFP: Sort Fraud Probabilities. The classification threshold is defined to maximize F1 metric.
- SWFP-P: Sort Weighted Fraud Probabilities considering the contracted peak power.
- SWFP-R: Sort Weighted Fraud Probabilities using regression algorithms.

Figure 2 provides the F1 score for the solutions described above. The horizontal axis represents the number m of customers to be inspected (i.e. labeled as fraudulent). As expected, SFP solution obtains the highest performance in terms of the F1 measure. On the other hand, Fig. 3 shows the accumulated income (i.e. the economic gain without accounting for inspection costs). And finally, Fig. 4 provides the net gain (i.e., gain minus cost) varying the number of inspections m for a fixed cost profile.

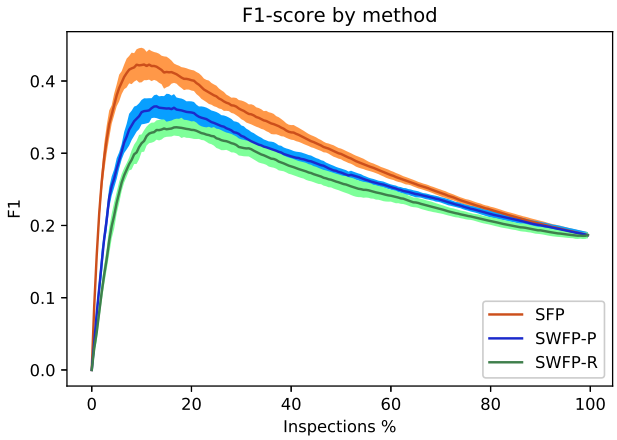


Fig. 2: F1 score for SFP, SWFP-P and SWFP-R solutions. Random Forest is used as classification and regression algorithm.

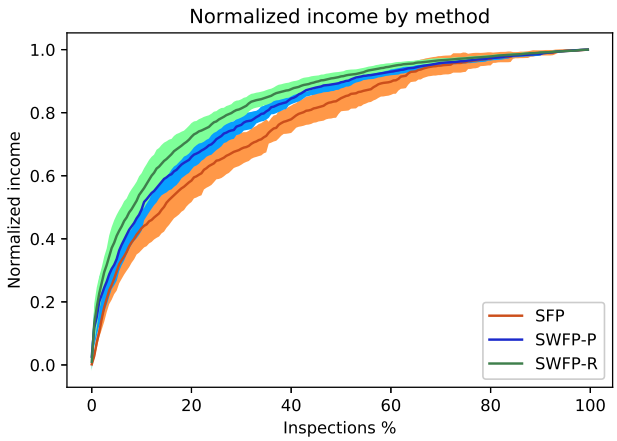


Fig. 3: Income for SFP, SWFP-P and SWFP-R solutions.

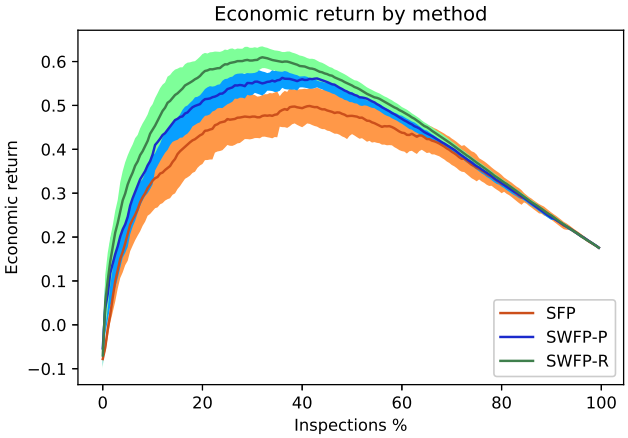


Fig. 4: Net economic return (normalized) for SFP, SWFP-P and SWFP-R solutions.

TABLE I: Comparison of calibration methods, Platt Scaling and Isotonic Regression for the three proposed solutions using RF algorithm over *NTL_10K_S* dataset.

Clibration	method	E. Return %	F1 %	precision %	recall %
Isotonic	SFP	31.5	42.9	44.5	41.3
	SWFP-P	55.0	33.5	22.3	67.6
	SWFP-R	58.2	31.5	21.0	62.7
Platt Scalling	SFP	32.3	43.6	41.1	46.5
	SWFP-P	55.7	32.8	21.9	65.4
	SWFP-R	61.0	32.7	22.7	58.4

TABLE II: Economic net return for the solutions SFP, SWFP-P and SWFP-R implemented with different classification/regression algorithms. The solutions we propose are also compared with the Cost-Based solutions proposed by Bahnsen et al. [34], [35] CostCla.CSRP and CostCla.CSDT

Alorithm	**Method**	**E. Return %**	**Inspections %**	**F1**
SVM	SFP	54.6%	44.5%	27.2%
RF	SFP	33.4%	10.5%	42.3%
NN	SFP	41.4%	27.5%	30.5%
SVM	SWFP-P	50.6%	28.5%	21.2%
RF	SWFP-P	56.3%	36.0%	30.6%
NN	SWFP-P	51.1%	34.0%	19.9%
SVM	SWFP-R	55.6%	40.5%	24.3%
RF	SWFP-R	**61.0%**	32.0%	30.6%
NN	SWFP-R	55.7%	29.5%	21.7%
CostCla.CSRP		49.2%	58.5%	25.2%
CostCla.CSDT		43.2%	59.5%	21.6%

Figures 2, 3 and 4 show results obtained using Random-Forest as classification and regression tool. Table I shows results for the three proposed solutions (SFP, SWSP-P and SWSP-R) comparing two methods of probability calibration. The experiment indicate that Platt Scalling leads to a better performance. Complementary results for SVM and NN algorithms are reported in Table II.

Algorithms aiming at maximizing the economic return were proposed in the context of credit card fraud. Some of these methods are implemented and publicly available in the Cost Sensitive Classification library [35]. We compare our solutions with two state-of-the-art cost sensitive existing methods: Cost-Sensitive-Decision-Tree (CSDT) and Cost-Sensitive-Random-Patches (CSRP), see Table II. In addition, we measure and compare how efficient these solutions are in terms of execution time. Since Costcla's algorithms require the prior knowledge of each sample individual cost, these algorithms need to be re-trained each time the cost model is updated (which becomes very inefficient when a large number of cost models wants to be tested). For example, to obtain the results reported in Table II, SWFP-R demanded approximately 9 seconds while CSDT required 4.2 hours.

The experiments presented above are obtained for a fixed cost model (fixed N). As explained in the previous section and illustrated in Alg. 1, cost curves $c_N(m)$ with nominal capacity

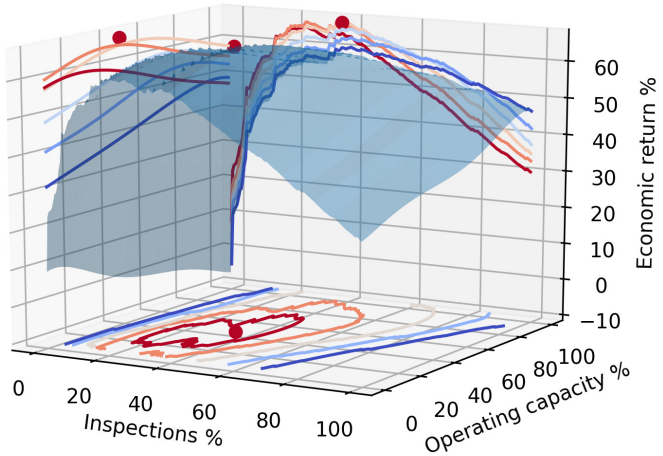


Fig. 5: Economic return depending on the number of inspections and the size of the operational capacity. Results over *NTL_10K_S* data set using RF/SWFP-R method. Red dot represent the maximum economic return and its projection is shown on the three planes.

N can be simultaneously compared. Figure 5 illustrates the net economic gain for different values of N (operating capacity) and m (actual number of inspections performed). The global maximum recovers a 68.6% of the total monetary value being stolen, this solution is obtained when the 33.5% of the customers are inspected.

Performing experiments with a fixed versus a variable value of N addresses two different practical situations companies face. In the first case (N is fixed) a company may have a fixed infrastructure (e.g. a given number of inspectors and vehicles) and wants to know which customers should be inspected to maximize the economic return. A second scenario, is when a company is determining which would be their optimal infrastructure, in this case N is also a free parameter that can be optimized.

D. Results on NLT_50K_R data set

Similar experiments are performed on NTL_50K_R dataset. This set extremely heterogeneous and in particular, the number of customers across the range of contracted power M_p is highly unbalanced. To prevent classification bias for certain values of the contracted power, oversampling techniques are considered (exclusively applied to the porting of the data selected for training). Once training data is balanced, the experiments are performed exactly as described for the previous dataset.

Table III reports the highest economic return obtained for each solution, and provide additional classification measures (the cost model is fixed in this experiment). Again, it can be seen that SFP

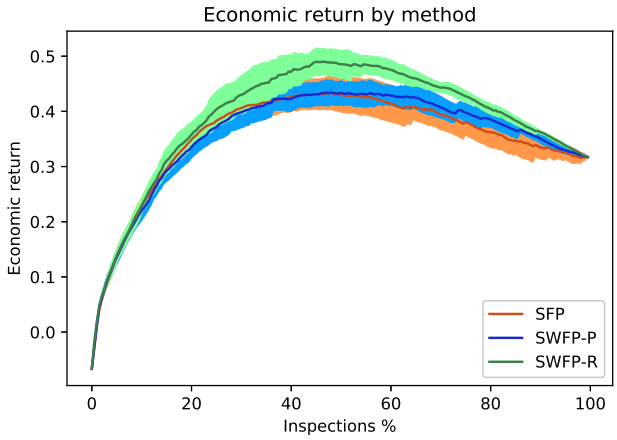


Fig. 6: Normalized net economic return versus the number of inspections (fixed cost model). Random Forest is used for classification/regression.

	SFP	SWFP-P	SWFP-R
Economic Return %	25.8	43.5	**49.0**
Inspections %	12.0	48.0	46.5
F1 %	37.1	24.9	26.1
precision %	32.8	14.8	15.6
recall %	42.7	77.3	78.8

TABLE III: Highest economic return achieved for SFP, SWFP-P and SWFP-R. Random Forest is used for classification/regression.

method achieves the highest F1 score. Figure 6 illustrates the net economic return for SFP, SWFP-P and SWFP-R solutions.

IV. Discussion and Conclusions

The experimental results show how the economic return can be drastically improved when machine learning solutions are developed with economic aspects in mind.

Random Forest, Support Vector Machines and Neural Networks proved to be adequate algorithms to implement the proposed schemes. Although all of them are extremely efficient and suitable for this task, RF was the algorithm that consistently provided the highest performance for the data at hand. As expected, when classification measures (such as the F1) are selected as maximization criteria, algorithms become optimal only with respect to that measure. This can lead to substantial economic losses when fraud detection solution are being developed. It is important to highlight that due to the large number of users, even a modest percentage increase of the relative economic return represent very large amounts of profit.

We observed that solutions that maximize the economic impact developed in the context of credit card fraud detection outperform classical approaches (compare for instance, the results reported in Table II for SFP with CSRP and CSDR). However, in the context of electric NTL and for the data collected across Uruguayan customers, we observed that the proposed strategies (both SWFP-R and SWFP-P) outperform other state-of-the-art methods. In addition, the proposed solutions are computationally more convenient than strategies such as CSRP and CSDR when different cost models need to be evaluated. Several realistic cost curves and the size of the infrastructure required can additionally be optimized as illustrated in Fig. 5. This is extremely useful as in addition of detecting fraud, the proposed approach can be used to simulate the economic impact associated to different management decisions.

Even-though we focus on this work in NTL for electric companies, the proposed pipeline is general and most of the ideas here presented can be easily adapted to other application. Furthermore the proposed approach focuses in designing an optimization criteria that is economically meaningful, and it is agnostic to the set of features or classification method at hand.

Acknowledgments

This work was partially supported by ANII and UTE: “Proyecto ANII Fondo Sectorial de Energía 14038”. The authors thank UTE for providing the datasets and for sharing their expertise in particular, authors greatly acknowledge engineers Juan Pablo Kosut and Fernando Santomauro. Additionally, J.M. Di Martino thanks Guillermo Sapiro for very fruitful discussions and the NSF, DoD, Cisco, Google, and Microsoft for research support.

References

[1] J. P. Kosut, F. Santomauro, A. Jorysz, A. Fernández, F. Lecumberry, and F. Rodríguez, “Abnormal consumption analysis for fraud detection: Ute-udelar joint efforts,” in *Innovative Smart Grid Technologies Latin America (ISGT LATAM), 2015 IEEE PES*. IEEE, 2015, pp. 887–892.

[2] P. Glauner, A. Boechat, L. Dolberg, R. State, F. Bettinger, Y. Rangoni, and D. Duarte, “Large-scale detection of non-technical losses in imbalanced data sets,” in *Innovative Smart Grid Technologies Conference (ISGT), 2016 IEEE Power & Energy Society*. IEEE, 2016, pp. 1–5.

[3] P. Glauner, J. Meira, P. Valtchev, R. State, and F. Bettinger, “The challenge of non-technical loss detection using artificial intelligence: A survey,” *International Journal of Computational Intelligence Systems 10.1 (2017): 760-775.*, 2017.

[4] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, and R. C. Green, “High performance computing for detection of electricity theft,” *International Journal of Electrical Power & Energy Systems*, vol. 47, pp. 21–30, 2013.

[5] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and F. Nagi, “Improving svm-based nontechnical loss detection in power utility using the fuzzy inference system,” *IEEE Transactions on power delivery*, vol. 26, no. 2, pp. 1284–1285, 2011.

[6] G. M. Messinis and N. D. Hatziargyriou, “Review of non-technical loss detection methods,” *Electric Power Systems Research*, vol. 158, pp. 250–266, 2018.

[7] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, “Nontechnical loss detection for metered customers in power utility using support vector machines,” *IEEE transactions on Power Delivery*, vol. 25, no. 2, pp. 1162–1171, 2010.

[8] C. C. O. Ramos, A. N. De Souza, D. S. Gastaldello, and J. P. Papa, “Identification and feature selection of non-technical losses for industrial consumers using the software weka,” *Industry Applications (INDUSCON), 2012 10th IEEE/IAS International Conference on*, pp. 1–6, 2012.

[9] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Support vector machine based data classification for detection of electricity theft," *2011 IEEE/PES Power Systems Conference and Exposition*, pp. 1–8, 2011.

[10] K. S. Yap, Z. Hussien, and A. Mohamad, "Abnormalities and fraud electric meter detection using hybrid support vector machine and genetic algorithm," *3rd IASTED Int. Conf. Advances in Computer Science and Technology, Phuket, Thailand*, vol. 4, 2007.

[11] C. Muniz, K. Figueiredo, M. Vellasco, G. Chavez, and M. Pacheco, "Irregularity Detection on Low Tension Electric Installations by Neural Network Ensembles," *International Joint Conference on Neural Networks*, no. March 2016, 2009.

[12] O. Ramos, P. Papa, and A. X. Falc, "A New Approach for Nontechnical Losses Detection Based on Optimum-Path Forest," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 181–189, 2011.

[13] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1606–1615, 2018.

[14] T. Hu, Q. Guo, X. Shen, H. Sun, R. Wu, and H. Xi, "Utilizing unlabeled data to detect electricity fraud in ami: A semisupervised deep learning approach," *IEEE transactions on neural networks and learning systems*, 2019.

[15] Y. Guo, C.-W. Ten, and P. Jirutitijaroen, "Online data validation for distribution operations against cybertampering," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 550–560, 2013.

[16] Y. Guo, C.-W. Ten, S. Hu, and W. W. Weaver, "Preventive maintenance for advanced metering infrastructure against malware propagation," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1314–1328, 2015.

[17] M. Di Martino, F. Decia, J. Molinelli, and A. Fernández, "Improving electric fraud detection using class imbalance strategies." in *International Conference on Pattern Recognition and Methods, 1st. ICPRAM.*, 2012, pp. 135–141.

[18] L. T. Faria, J. D. Melo, and A. Padilha-Feltrin, "Spatial-temporal estimation for nontechnical losses," *IEEE Transactions on Power Delivery*, vol. 31, no. 1, pp. 362–369, 2016.

[19] P. Glauner, J. Meira, L. Dolberg, R. State, F. Bettinger, Y. Rangoni, and D. Duarte, "Neighborhood features help detecting electricity theft in big data sets," in *Proceedings of the 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies*. IEEE, 2016.

[20] P. Massaferro, H. Marichal, M. Di Martino, F. Santomauro, J. P. Kosut, and A. Fernández, "Improving electricity non technical losses detection including neighborhood information," in *2018 IEEE PES General Meeting (GM) - IEEE Power and Energy Society, Portland, Oregon, USA, 5-9 aug*. IEEE, 2018, pp. 1–5.

[21] N. F. Avila, G. Figueroa, and C.-C. Chu, "Ntl detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and random undersampling boosting," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7171–7180, 2018.

[22] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, 2011.

[23] J. West and M. Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," *Computers & security*, vol. 57, pp. 47–66, 2016.

[24] A. C. Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten, "Cost sensitive credit card fraud detection using bayes minimum risk," in *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, vol. 1. IEEE, 2013, pp. 333–338.

[25] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[26] C. Elkan, "The foundations of cost-sensitive learning," in *International joint conference on artificial intelligence*, vol. 17, no. 1. Lawrence Erlbaum Associates Ltd, 2001, pp. 973–978.

[27] J. Hernández-Orallo, P. Flach, and C. Ferri, "A unified view of performance metrics: translating threshold choice into expected classification loss," *Journal of Machine Learning Research*, vol. 13, no. Oct, pp. 2813–2869, 2012.

[28] B. W. Silverman, *Density estimation for statistics and data analysis*. Routledge, 2018.

[29] N. M. Nasrabadi, "Pattern recognition and machine learning," *Journal of electronic imaging*, vol. 16, no. 4, p. 049901, 2007.

[30] K. S. Yap, S. K. Tiong, J. Nagi, J. S. P. Koh, and F. Nagi, "Comparison of supervised learning techniques for non-technical loss detection in power utility," *International Review on Computers and Software (I.RE.CO.S.)*, vol. 7, no. 2, pp. 1828–6003, 2012.

[31] B. Zadrozny and C. Elkan, "Transforming classifier scores into accurate multiclass probability estimates," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2002, pp. 694–699.

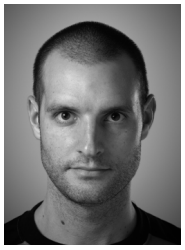
[32] J. Platt *et al.*, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, vol. 10, no. 3, pp. 61–74, 1999.

[33] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," *arXiv preprint arXiv:1706.04599*, 2017.

[34] A. C. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive decision trees," *Expert Systems with Applications*, vol. 42, no. 19, pp. 6609–6619, 2015.

[35] ——, "Ensemble of example-dependent cost-sensitive decision trees," *arXiv preprint arXiv:1505.04637*, 2015.

Pablo Massaferro received the B.Sc. and M.Sc. from Universidad de la República, Uruguay, in 2008 and 2018 respectively. Currently is a Ph.D. student and holds a position as Assistant Professor in the Signal Process Department of Universidad de la República, Uruguay. His main research areas are: pattern recognition, voice identification and anomaly detection.



J. Matias Di Martino received the B.Sc. and Ph.D. degrees in Electrical Engineering from the Universidad de la Republica, Uruguay, in 2011 and 2015 respectively. During 2016 he worked as Research Associate at Ecole Normale Superieure de Cachan, Paris. Currently, he is working as Assistant Professor at the Physics Department of the School of Engineer, Universidad de la Republica (UY). In addition, he holds a Posdoctoral Associate Researcher position at Duke University (US). His main areas of research are Applied Optics, Image Processing and Data Science.

Alicia Fernandez Full Professor of Signal Processing at the Electrical Engineering Institute(IIE), Universidad de la República. Since 1989, she works at the IIE, in telecommunication and signal processing areas. Her main research interests are signal processing and pattern recognition with focus in biomedical image analysis, biometric identification, anomaly detection and big data analysis.