engineering proceedings

MDPI

Proceeding Paper

Detection of Non-Technical Losses in Special Customers with Telemetering, Based on Artificial Intelligence †

José Luis Llagua Arévalo * and Patricio Antonio Pesántez Sarmiento *

Department of Electrical and Electronic Engineering, National Polytechnic School, Quito 170525, Ecuador
* Correspondence: jose.llagua@epn.edu.ec (J.L.L.A.); patricio.pesantez@epn.edu.ec (P.A.P.S.)
† Presented at the XXXII Conference on Electrical and Electronic Engineering, Quito, Ecuador, 12–15 November 2024.

Abstract: The Ecuadorian electricity sector, until April 2024, presented losses of 15.64% (6.6% technical and 9.04% non-technical), so it is important to detect the areas that potentially sub-register energy in order to reduce Non-Technical Losses (NTLs). The "Empresa Eléctrica de Ambato Sociedad Anónima" (EEASA), as a distribution company, has, to reduce NTLs, incorporated many smart meters in special clients, generating a large amount of data that are stored. This historical information is analyzed to detect anomalous consumption that is not easily recognized and is a significant part of the NTLs. The use of machine learning with appropriate clustering techniques and deep learning neural networks work together to detect abnormal curves that record lower readings than the real energy consumption. The developed methodology uses three k-means validation indices to classify daily energy curves based on the days of the week and holidays that present similar behaviors in terms of energy consumption. The developed algorithm groups similar consumption patterns as input data sets for learning, testing, and validating the densely connected classification neural network, allowing for the identification of daily curves described by customers. The results obtained from the system detected customers who sub-register energy. It is worth mentioning that this methodology is replicable for distribution companies that store historical consumption data with Advanced Measurement Infrastructure (AMI) systems.

Keywords: energy under registration; fraudulent curves; validation indices; neural networks; non-technical losses

check for updates

Citation: Llagua Arévalo, J.L.; Pesántez Sarmiento, P.A. Detection of Non-Technical Losses in Special Customers with Telemetering, Based on Artificial Intelligence. *Eng. Proc.* **2024**, *77*, 29. https://doi.org/10.3390/engproc2024077029

Academic Editor: Walter Vargas

Published: 18 November 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In a society that is highly dependent on the availability, efficiency, and reliability of electricity, it is essential for the electric companies in the sector to have good management of the production and distribution of energy. One of the major problems that concerns this industry is electrical losses, which occur mostly in the distribution segment [1]. In the transportation of energy, losses are the difference between the electricity that enters the network and that which is delivered for final consumption, and they reflect the level of efficiency of the transmission and distribution infrastructure. Being able to reduce losses is essential to increasing the efficiency of energy distribution, and, in many cases, it can even help improve the financial sustainability of distribution companies. In Ecuador, at a general level, in recent years, in the electric sector, the level of losses has substantially improved, dropping from 22% in 2006 to 13.06% in 2021 [2]; however, recently, it has increased to 15.64% (6.6% technical and 9.04% non-technical). The advances in the identification of NTLs have been undeniable recently, and identification models have been established with the information to detect patterns or find areas where there are high percentages of NTLs, as well as methodologies based on energy balances at the level of distribution transformers and medium voltage branches. Smart Grids have evolved to the level of incorporating techniques based on artificial intelligence to solve real problems, such as NTLs in distributors, demand prediction, and adaptive protections, among others. These systems generate

a large amount of data and cybersecurity plays an important role. The purpose of this article is to develop a methodology to detect non-technical losses using readings of energy consumption every 10 min, use various classification indexes that guarantee the formation of groups, and finally implement artificial intelligence techniques for the development of pattern recognition of fraudulent and non-fraudulent curves of industrial clients owned by EEASA. Compared to the latest advances in NTL methodologies, in which the Multiversal Recurrent Algorithm with Multiple Repetitions (MV-REMR), published in November 2023, stands out [3], the lack of information on energy consumption replaced by zeros has, in this research, been solved through the use of statistical interpolation methods. The records obtained by the AMI, with sampling every 10 min, generated a matrix of 105,264 per special client for two years; to facilitate information management, the reduction to an hourly demand was carried out, and this allowed us to improve the classification of the clients using clustering techniques supported by three k-means validation indices, complying with the class balance. Densely Connected Neural Networks (DenseNet) have a better response in non-discretized systems compared to the Convolutional Neural Networks (CNN) used in MV-REMR, allowing for a better view of the consumption pattern of special clients. Likewise, the use of AMIs, due to their high reliability in data acquisition, communication, and storage, makes the proposed methodology more effective.

2. Methodological Proposal

The methodology consists of the evaluation of historical daily curves in a two-year time interval for the identification of non-technical losses through deep learning of the classification neural network. This method will help us to classify curves with anomalies and typical curves based on learning from the historical data downloaded from the distribution company's customers, identifying patterns considered with the underreporting of customers that could be suspected of energy fraud. Figure 1 summarizes this methodology, showing three clear stages for the development of this work.

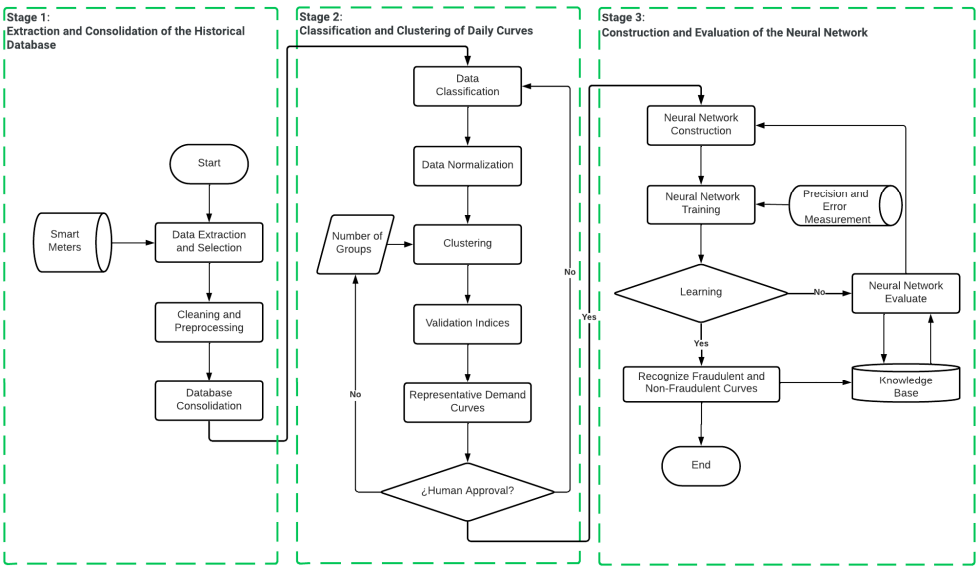


Figure 1. Methodology flowchart.

2.1. Database Extraction and Consolidation

From the historical data of the special customers (industrial and commercial) that have telemetering for the period from 31 May 2020 to 31 May 2022, a data matrix will be created for which a pre-selection of the most representative customers will be made according to their consumption characteristics and metering history.

2.1.1. Extraction and Selection

In the selection, customers have determining details, such as the characteristic of "possessing Telemetry" and "Total kW Consumption", which will be important at the time of selecting 100 of the 297 customers that meet the conditions of having historical data for greater than two years and for whom there has not been a reason for a change in the telemetering process during the designated time interval. With the stratified sample method, the number of individuals from each zone to be extracted to form the customer matrix is determined, as shown in Table 1, which specifies the number of customers chosen from each zone and the number from which the historical data will be extracted.

Table 1. Stratified sample of customers.

Zone	Description	Customers	Sample
1	Ambato	142	60
2	Pelileo	43	10
3	Pillaro	13	3
4	Baños	11	5
5	Patate	8	2
7	Pastaza	43	8
8	Palora	2	2
10	Quero	14	3
11	Tena	19	5
12	Archidona	2	2
TOTAL		297	100

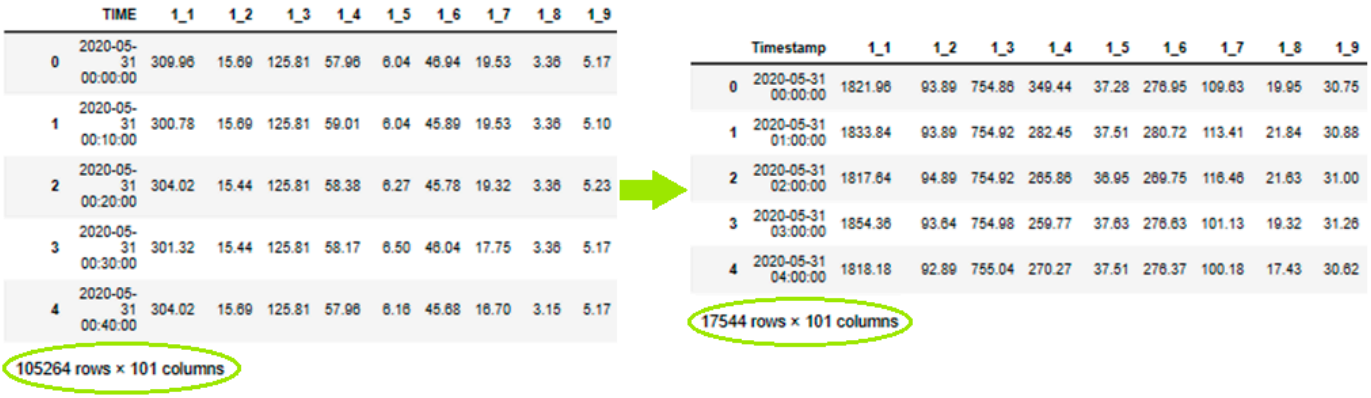
The platform "Telemetering Systems for special customers EASYmetering AMI Solutions", used by the distributor, has several clusters where they store readings of voltage [V], current [A], power [kW], active energy [kWh], and reactive energy [kVArh] in intervals of 10 min; this results in 144 data entries for each day.

2.1.2. Cleaning and Preprocessing

It is important to mention that not all the records are complete due to failures in the communication between the AMI and the server; also, not all the columns contain the same format. The work is meticulous client-by-client since each file had its own particularities which did not allow for treating all the extracted files in a general way. Instead, we followed a small procedure to eliminate inconsistencies, which is detailed as follows: 1. In the debugging phase, missing values are detected and replaced by means of interpolation techniques; 2. The interpolation considers all the points and correlates the complete behavior of the set used; and 3. Cases where there was no data had the particularity that they were followed by Not a Number (NaN) values, showing that there were connection failures with the server and the AMI meter, so the data were interpolated for the interval with these values.

2.1.3. Consolidation of the Complete Data

With the customer data processed and purified, which contained the 105,264 pieces of data on energy delivered every 10 min during the 731 days of the two years, we proceeded to consolidate the database of the 100 customers. Keeping their names assigned by the zone of their location and the place they occupied in the ranking of energy consumption, the consolidated data are reduced to 17,544 rows to obtain consumption values for each hour, as shown in the green circles of the Figure 2, where the first columns have the date and time of the reading of the energy delivered and then, in chronological order, the demand values expressed in kW that each client had, thus forming the history of the two years of AMI readings.



	TIME	1_1	1_2	1_3	1_4	1_5	1_6	1_7	1_8	1_9
0	2020-05-31 00:00:00	309.96	15.69	125.81	57.96	6.04	46.94	19.53	3.36	5.17
1	2020-05-31 00:10:00	300.78	15.69	125.81	59.01	6.04	45.89	19.53	3.36	5.10
2	2020-05-31 00:20:00	304.02	15.44	125.81	58.38	6.27	45.78	19.32	3.36	5.23
3	2020-05-31 00:30:00	301.32	15.44	125.81	58.17	6.50	46.04	17.75	3.36	5.17
4	2020-05-31 00:40:00	304.02	15.69	125.81	57.96	6.16	45.68	16.70	3.15	5.17

105264 rows × 101 columns

	Timestamp	1_1	1_2	1_3	1_4	1_5	1_6	1_7	1_8	1_9
0	2020-05-31 00:00:00	1821.96	93.89	754.86	349.44	37.28	276.95	109.63	19.95	30.75
1	2020-05-31 01:00:00	1833.84	93.89	754.92	282.45	37.51	280.72	113.41	21.84	30.88
2	2020-05-31 02:00:00	1817.64	94.89	754.92	265.86	36.95	269.75	116.46	21.63	31.00
3	2020-05-31 03:00:00	1854.36	93.64	754.98	259.77	37.63	276.63	101.13	19.32	31.26
4	2020-05-31 04:00:00	1818.18	92.89	755.04	270.27	37.51	276.37	100.18	17.43	30.62

17544 rows × 101 columns

Figure 2. Variabilit.

2.2. Classification and Clustering of the Daily Curves

2.2.1. Data Reduction and Sorting

To obtain a more effective data reduction without losing important information in the energy records, all energy values within each hour were summed, thereby obtaining a reduction from "105,264 rows × 100 columns" to "17,544 rows × 100 columns".

The classification of the data was performed with prior knowledge of the way in which EEASA's special customers record their energy consumption, knowing beforehand that most industrial sector workers work typical working days between Monday and Friday, that, on weekends, few industrial customers work fully, and that holidays are a very unfavorable option to work. The behavior of the daily curves generated on the basis of the energy consumption supplied is similar in most of the special customers, clearly affecting their shape on weekdays; for more information see [4].

Table 2 shows a summary of the classification performed, detailing the days of each group, the data within each group, and the matrix formed for the grouping with the k-means indexes.

Table 2. Classification data frames.

Ranking	Days	Data Frame
Group from Monday to Friday	500	12,000 × 100
Weekend group	173	4152 × 100
Holiday group	58	1392 × 100

2.2.2. Data Normalization

The minimum–maximum normalization methodology is used—see [5]. Figure 3 shows the curves described with the normalized data for the Group from Monday to Friday corresponding to the normalized consumptions represented in different colors for each of the 100 customers, where all values are within the interval 0 and 1.

2.2.3. Clustering

This sub-process conglomerates the daily curves into compact groups with distinct and significant properties from the rest of the groups. In this work, the clustering capability of the k-means algorithm and its variations are investigated; see [6] to understand the heuristic iterative procedure to cluster the Representative Demand Patterns into k groups. Where the procedure is fulfilled, choose centroids of k groups (randomly chosen from the set of the Daily Curves), then each Daily Curve is integrated into the nearest group (according to the optimization function), and, finally, the centroids are recalculated by averaging the Daily Curves of its members. The process is repeated until the centroids of the cluster are stable—see [7].

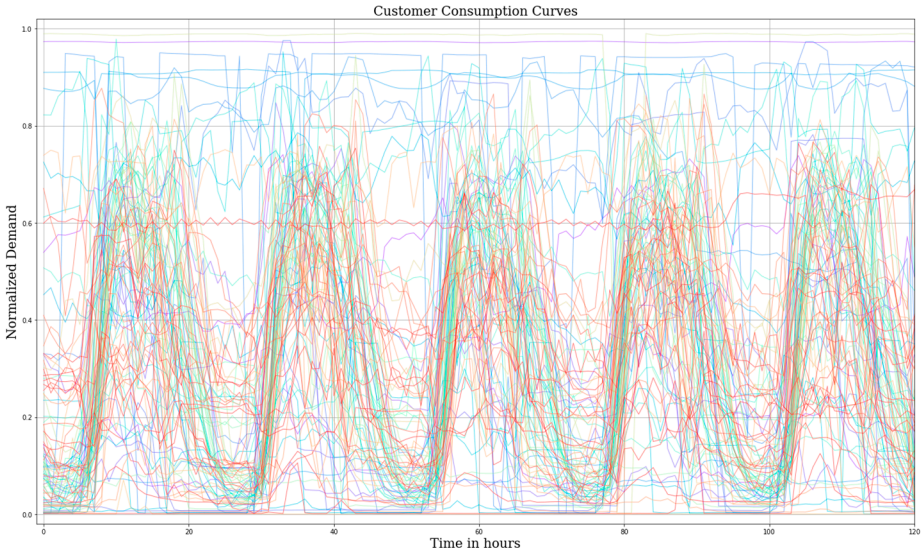


Figure 3. Demand.

2.2.4. Validation Indexes

The indexes used are Euclidean k-means, Dynamic Time Warping (DTW), Barycenter Averaging k-means (DBA), and Soft-DTW k-means, which employ evaluation coefficients such as DTW, Within Cluster Sum of Saqueis (WCSS), and Silhouette Coefficient (SC)—see [8].

In Figure 4, the red curve represents the centroids of each cluster, and the black ones represent the clients that make up the group; the value of "*n*" gives the number of clients that are in each group. The visual analysis of each group with the location of the centroids with respect to the curves that make up the cluster and the experience in the behavior of the curves of the industrial clients are important factors for the definition of the value of k and the most suitable classification method for each group.

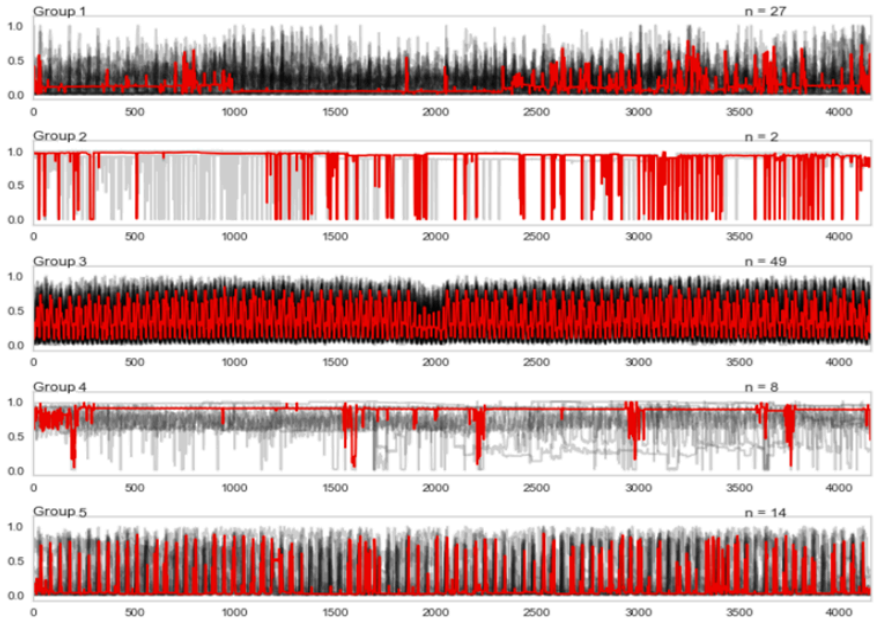


Figure 4. Grouping using the Soft-DTW k-means index for a k = 5, represented the centroid curves in red color.

2.2.5. Definition of Groups

In the visual evaluation process, it is important to carefully observe the graphs described in each evaluation index to determine the method that best describes the group, taking into account the position of the centroids calculated and fully defined in red and the behavior of the curves of the customers with respect to the centroid within each group. Figure 5 shows how they were classified for a k = 5 and the value of 'n' with the number of customers found in each group. It is important to mention that the ideal k value for this project is 5, since 'k' values equal to 4, 6, and 7 do not show homogeneity in the clusters. For more details about evaluation indexes, see [9].

It is important to emphasize that, for the previously classified groups (Monday through Friday, holidays, and weekends), the same methodology is applied, but the behavior of the evaluation indexes with respect to their daily curves is different. Any index of the three that were evaluated can be chosen, observing the best grouping, the number of individuals in the group, and the way the centroid describes the complete group.

2.2.6. Conformation of Groups

With the method and the value of k defined, the assigned values for the grouping and the calculated centroids are determined; with an array of the assignments to each group presented, the centroids calculated for each hour of energy consumption are also normalized. With a small array of rows and columns, each customer can be assigned to the group to which they belong, as shown in Figure 5, where the first column shows the customer, followed by the normalized data of the daily curves, and the last column shows the number of the group to which the customer was assigned. With the values assigned in the Data Frame, the five groups are created with the given assignments—see [10].

	0	1	...	1390	1391	Groups
1_1	0.675224	0.677938	...	0.568357	0.581090	0
1_2	0.748290	0.758575	...	0.103997	0.102846	3
1_3	0.967538	0.967640	...	0.918640	0.918844	3
1_4	0.055345	0.018177	...	0.507596	0.418882	4
1_5	0.116467	0.054153	...	0.018108	0.019235	3
...	...	...	...	...	...	...
11_4	0.019700	0.025805	...	0.195339	0.208935	0
11_5	0.013710	0.007192	...	0.103012	0.074843	4
12_1	0.088134	0.101428	...	0.168554	0.155096	4
12_2	0.170968	0.178710	...	0.272258	0.208710	4
12_3	0.603521	0.601006	...	0.610855	0.611693	4

100 rows × 1393 columns

Figure 5. Grouping assigned values.

2.2.7. Creation of Fraudulent Customers

With the lack of data on daily curves that present abnormal consumptions by the distributor, patterns that under-register energy are created to simulate typical frauds that may occur in special customers of the distributor. Therefore, for the creation of these patterns, the aim is to simulate this type of behavior with a set of random variables, simulating the different types of under-registration that can be created by using malicious activities and directly affecting the distributor and increasing the non-technical losses of the same—see [11]. Therefore, three types of frauds are created, and are described as follows: Type 1 (T1): Constant Proportional Decrease in Time; Type 2 (T2): Proportional Decrease in

Time Windows with Time Band at Maximum Consumption (Peak Hours); and Type 3 (T3): Proportional Decrease in Time Windows with Indistinct and Random Time Band.

To create the fraudulent curves, the parameters detailed in Table 3 were used, in which the values are specified to create the synthetic base that contains the curves with under-registration. With the use of statistical metrics, such as standard deviation, the minimum and maximum values of variability of the data of each group in the daily consumption curves are found.

Table 3. Creation of fraudulent curves.

	Standard Deviation		**Fraudulent Curves**		
	MIN	**MAX**	**Min. % of Admissible Variation**	**% Min. of Variation Chosen**	**Max. % Variation Chosen**
Group Monday to Friday	0.172888139	0.303566006	30.35%	35.00%	85.00%
Group Weekend	0.228457341	0.32703843	32.70%	35.00%	85.00%
Group Holidays	0.210337241	0.344457664	34.45%	35.00%	85.00%

The percentages of decrease, which are between 35% and 85%, create fraudulent curves; these are randomly chosen values and form curves with abnormal consumptions for each type of fraud (in blue color), with respect to the normal energy consumption curves (in orange color), as shown in Figure 6. Each chart has a label that shows the customer code plotted, the type of fraud performed, and the percentage that has been decreased to the normal consumption curve.

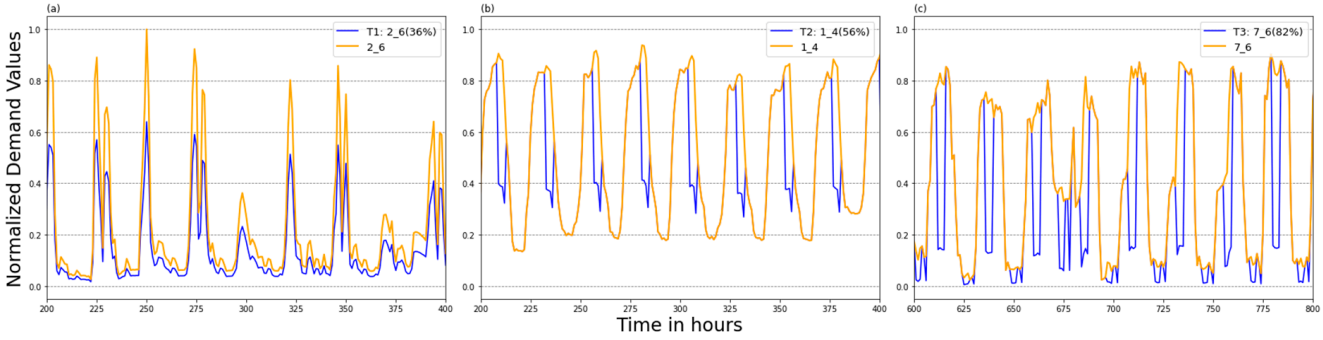


Figure 6. Normal and fraudulent consumption curves with percentage decrease. (**a**) Type 1 with 36% of customer 6 in zone 2. (**b**) Type 2 with 56% of customer 4 in zone 1 and (**c**) Type 3 with 82% of customer 6 in zone 7.

2.3. Construction of the Neural Network

The neural network model to be built is shown in Figure 7, where can see the input, hidden and output layer (rectangles blue color), the main hidden layer is constituted for three hidden layers (rectangles red color) with eighty nine, ninety, eighty nine neurons respectively and the black arrows show the density connection into NN; It is one of the three models to be built for each group is shown. The hyperparameters are parameters external to the model itself, set by the neural network programmer, for example, the selection of the activation function to use or the batch size used in training. The model parameters are internal to the neural network, for example, the weights of the neurons. Stating that the hyperparameters are the adjustable parameters used to control the training process of the model, it must be specified that the values set for each model are different since each data set has different row dimensions.

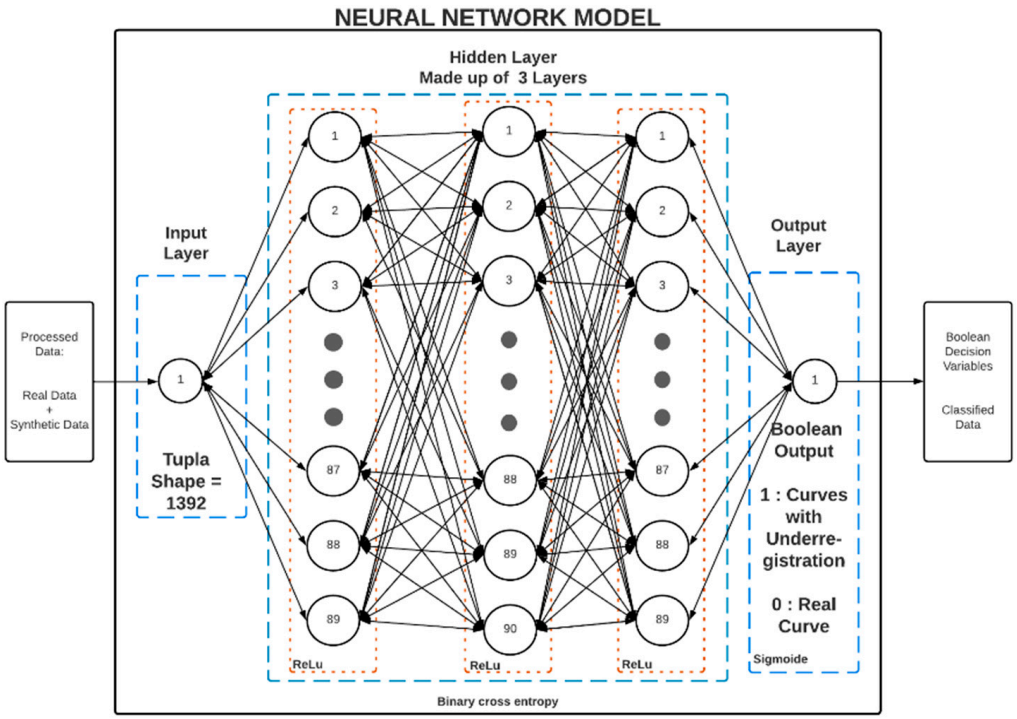


Figure 7. Model network design for the holiday group.

2.3.1. Construction of the Neural Network

For the construction of the neural network, the free software KNIME Analytics Platform 4.7.1 is used which, despite its high performance in the handling of large data sets, must be linked to the installation of the "Python Deep Learning" package with the Keras and Tensor Flow 2 libraries, which are compatible with Python 3.6.13, which comes by default in the installation of Conda in its latest version (Conda 4.12.0). Figure 8 shows the linked software so that the libraries can be used for the construction of the neural network.

2.3.2. Configuration of the Neural Network

For the configuration of the neural network model, the following nodes must be configured: Keras Input Layers, Keras Dense Layers, Keras Network Learner, Keras Network Executor, and a node for the visualization of results which will be Line Plot (local) after forming the complete data (real data and fraudulent data of the three types). The hyperparameters shown in Figure 8 exist in order to form the densely connected neural network that will be trained and tested to check if it is learning and generating knowledge in the identification and classification of curves with sub-register with a value of 1 and normal curves with values of zero. For more information on this, see the thesis work [10], where the process followed in the reading, classification (training, validation, and test data), learning, validation, and testing of the formed groups are shown in detail.

The neural network model and the data execution nodes will be shown configured in the work area, as shown in Figure 9, where all the nodes are executed and in green color, showing the correct procedure in data loading and filtering, configuration, training, validation, and testing of the neural network model. The yellow nodes represented the data conditioning, the magenta nodes show the tag assignment, the brown nodes present the neural network architecture and finally the green nodes evince the trained network model.

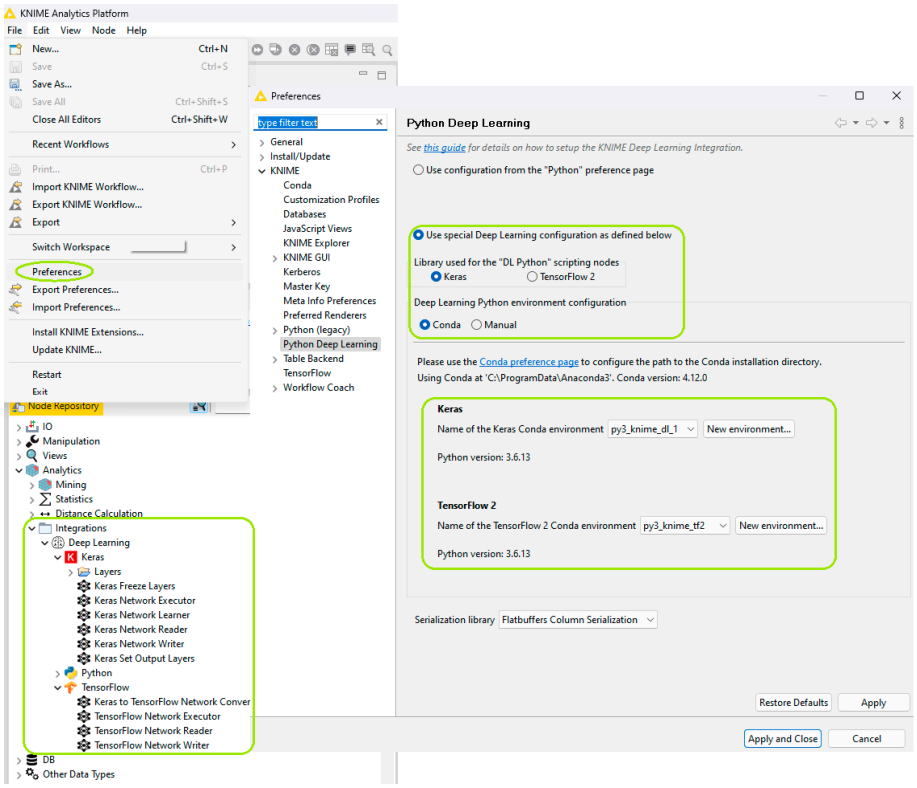


Figure 8. KNIME—Python link and deep learning libraries.

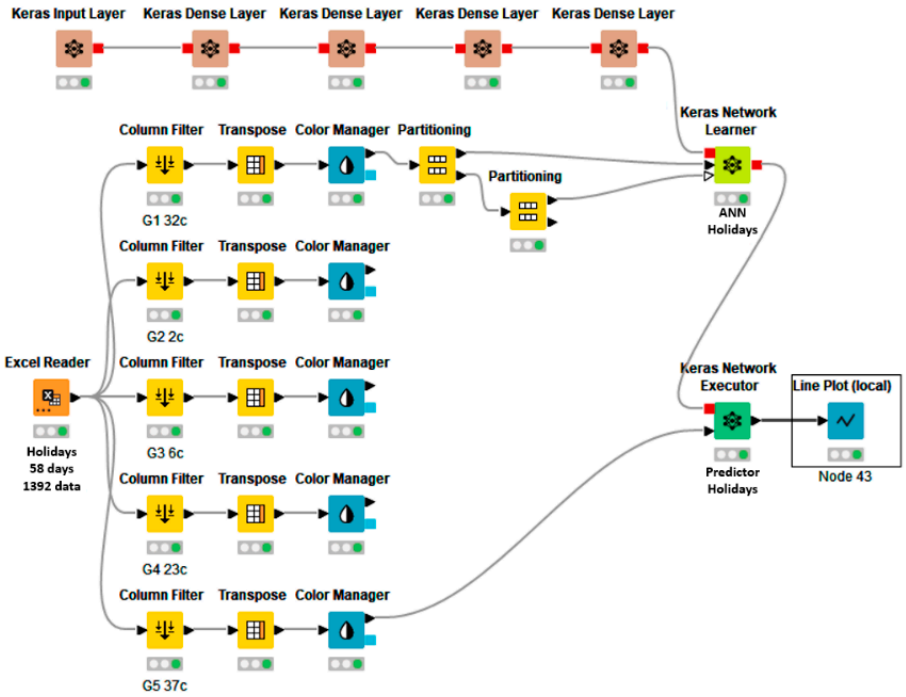


Figure 9. Completed neural network in the working environment.

This process of building the neural network model must be performed for each classification group (Monday to Friday, weekends, and holidays), so, in the end, three neural network models will have to be trained, see [10].

3. Results

The results are focused on the prediction performed by the neural network to classify real and fraudulent curves, so they will focus directly on the accuracy and loss margins of the trained neural network.

3.1. Accuracy

In the option *View: Learning Monitor* of the *Keras Network Layer*, the *Accuracy* of the training can be observed, which indicates the model accuracy and how the model is learning with the *Batches* configured during the construction of the network, as shown in Figure 10. In Figure 10 it can also be seen that the smoothed curve of the training data (in red color) is reaching the values of the smoothed curve of the validation data (in blue color), with small and acceptable error margins within the training. This confirms that Adam's optimization function and the activation functions configured in the hidden layers work together for the learning of the neural network. Also, these results can be verified between the curves of the training data (in pink color) and the validation data (in lilac color), as the values change abruptly at the beginning and the others describe the step function of the activation function *Binary cross entropy* as they pass the *Batch* configured during the learning process of the model, adjusting until the necessary knowledge base for the neural network is achieved.

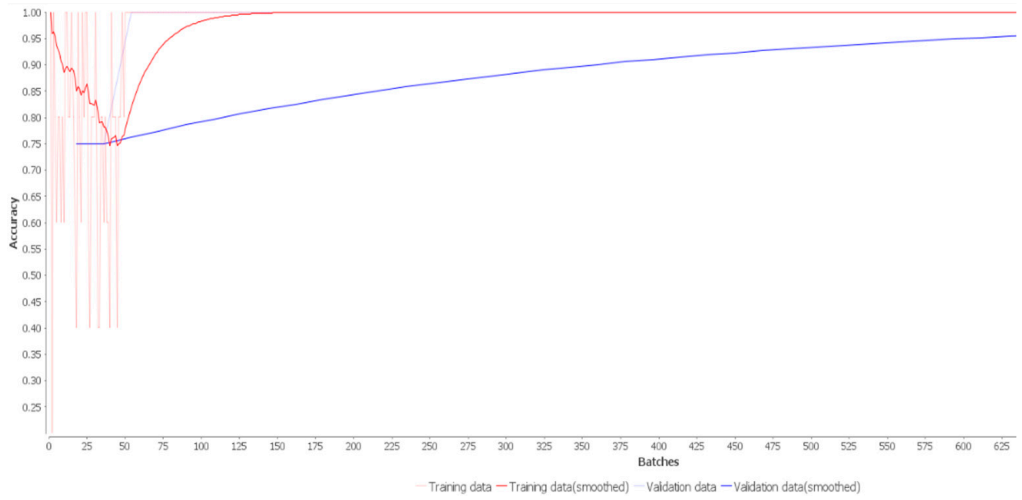


Figure 10. Accuracy curves of the neural network.

3.2. Losses

In the Loss tab, within the same View: Learning Monitor, it can be observed how the values of the error calculated by the Binary cross entropy function, used for the classification of the curve patterns, make the values desirable since the error tends toward zero and the curves converge as the Batch and Epoch are advancing, as shown in Figure 11. Here, it can also be seen that the red curve (Smoothed Training Data) and the blue curve (Smoothed Validation Data) generate increasingly similar values, reaching a minimum distance between the predicted values and the desired values. Between the purple curve (Validation Data) and the pink curve (Training Data), we can see the abrupt changes that the network model undergoes, showing how the configured hyperparameters make the model predict and learn the expected responses. When the accuracy and error curves follow the same trend, it is an indicator that it is a good model.

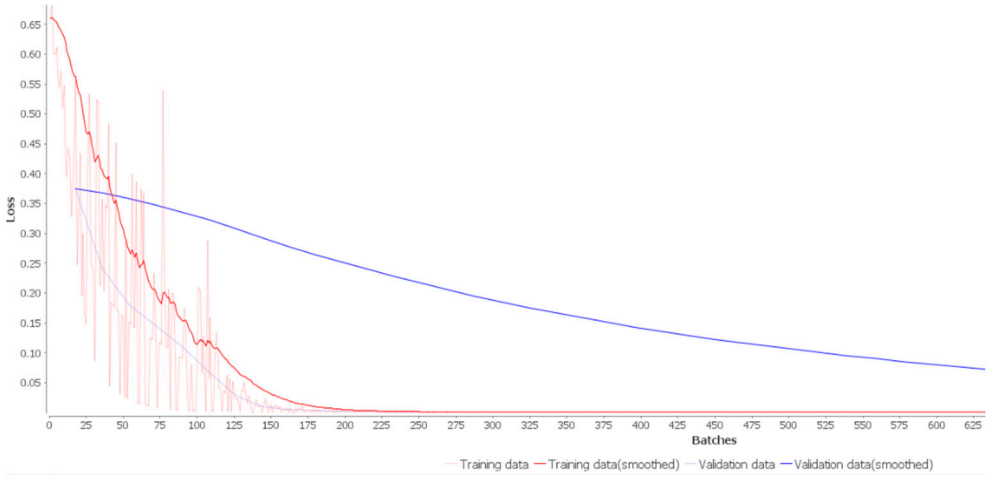


Figure 11. Losses curves of the neural network.

3.3. Tests

With the trained neural network, tests are performed with data from 20 new customers that are completely unknown to the trained network, where the network analyzes 2 months of historical data from 1 June to 31 July 2023. This period has the particularity of not containing holidays, making 61 days to be analyzed, and the results will be shown with the *Line Plot Node*.

The results of the prediction accuracy in red line of the weekend data evaluated on the neural network (18 days) are shown in Figure 12.It can be said that customers 1_2, 1_4, 1_5, 1_8, 3_1, and 7_2, present anomalous consumptions because the red dots have values close to Boolean value one, being possible customers that are performing to register abnormal consumptions or under-registering energy.

Finally, we analyze the data for the group from Monday to Friday (43 days), and we can say that customers 1_4, 1_8, and 7_2, repeat this pattern of anomalous consumption or under-registration of energy, as shown in Figure 13, where we can clearly observe that the high Boole values of these customers are approximated to one, corresponding to abnormal energy consumption.

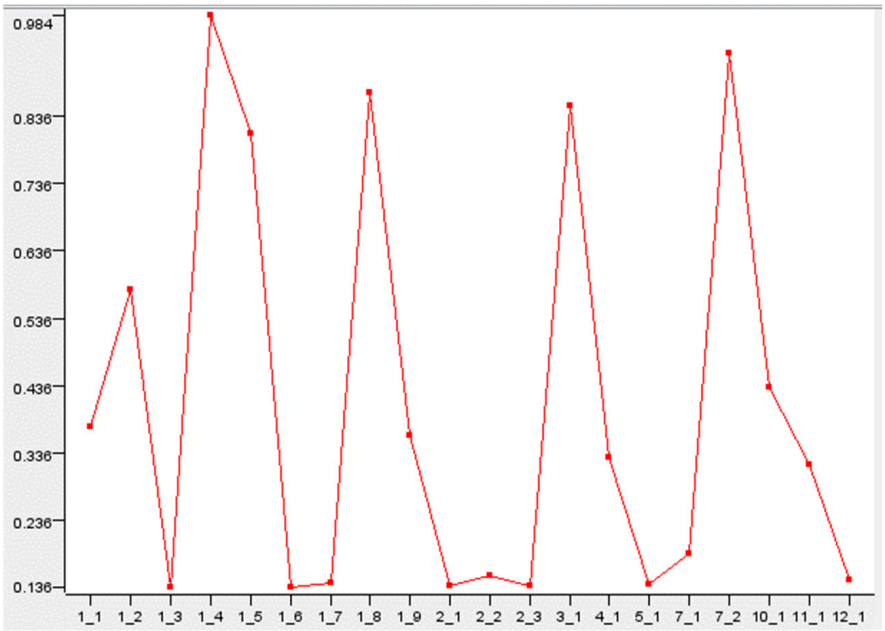


Figure 12. Weekend neural network results.

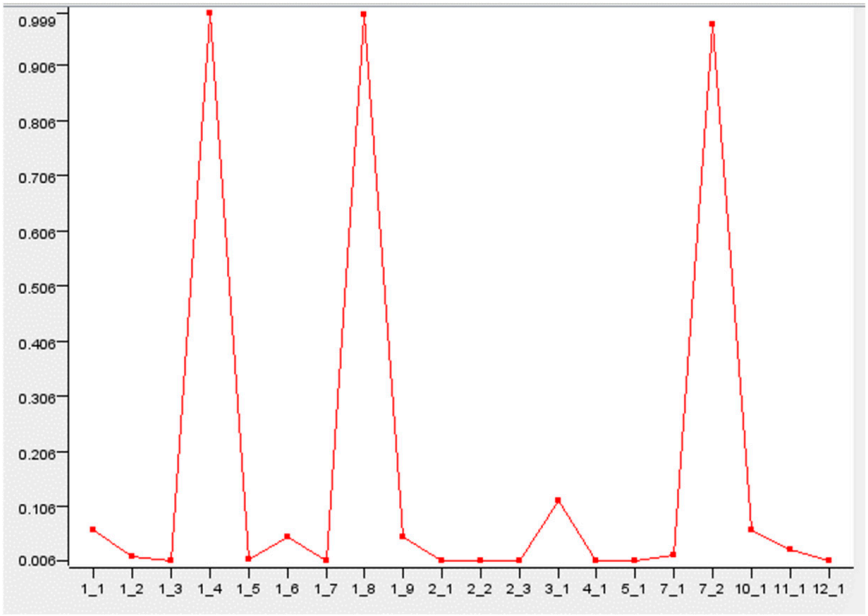


Figure 13. Results of the neural network from Monday to Friday.

With the results given, it can be affirmed that clients 1_4, 1_8, and 7_2 are candidates to be inspected for abnormal consumption. Previously to the culmination of this article, EEASA indicated that, after an inspection was carried out, one of the clients is a frequent recipient of fines for this particularity and the other two presented technical reports of corrective maintenance for problems in their machinery, causing the low energy consumption that they regularly register during their regular operation.

4. Conclusions

In the development of the neural network model, the training process is important, since the hyperparameters must be correctly chosen and configured (structure and topology of the neural network: number of layers, number of neurons in each layer, activation functions, etc.; at the level of the learning algorithm: the Epoch, the Batch Size, the Learning Rate, the Momentum, etc.). The hyperparameters (structure and topology of the neural network: number of layers, number of neurons in each layer, activation functions, etc.; at the level of the learning algorithm: Epoch, Batch Size, Learning Rate, Momentum, etc.) are important because they fulfill their function at each stage, being an essential step to achieve an acceptable neural network model, so that it fulfills the objective for which it was designed.

In the training process, if the expected results are not achieved, the hyperparameters must be changed, such as varying the number of neurons in the hidden layers in accordance with the number of variables entered, as well as choosing appropriate activation functions. This is important since one of their main functions is to model the nonlinearity of the data and these are transcendental decisions in the training of the model and since they are in charge of activating or not activating these neurons in each training cycle.

Loss and Accuracy functions should be analyzed and monitored during the model training process because the shape and trend allow us to know if the network is learning.

Overfitting (overfitting or overtraining) and underfitting (undertraining or underfitting) are references that make the neural network model unable to capture the underlying trends of the data, i.e., when generalizing the model, it does not fit with the knowledge it is expected to acquire, causing the Accuracy and Loss curves to diverge and the network model to fail to fulfill the purpose for which it was created.

Author Contributions: Conceptualization, J.L.L.A. and P.A.P.S.; methodology, J.L.L.A. and P.A.P.S.; software, J.L.L.A.; validation, J.L.L.A.; formal analysis, J.L.L.A. and P.A.P.S.; investigation, J.L.L.A.; resources, J.L.L.A.; data curation, J.L.L.A.; writing—original draft preparation, J.L.L.A. writing—review and editing, J.L.L.A.; visualization, J.L.L.A.; supervision, P.A.P.S.; project administration, J.L.L.A.; funding acquisition, J.L.L.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data can be obtained from the corresponding author on request.

Acknowledgments: To the director of this degree work, Patricio Antonio Pesántez Sarmiento, who collaborated tirelessly in the development of this degree project; to the National Polytechnic School, for the opportunity provided for professional development through its master's program; and finally, to my wife, children, and family, who are always present.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Estupiñan, J.J.; Giral, D.A.; Martínez, F. *Implementation of Algorithms Based on Support Vector Machines (SVM) for Electrical Systems: Subject Review*; Francisco José de Caldas District University: Bogotá, Colombia, 2016.
2. Bermeo, J.C.; Arguello, G.A.; Cepeda, J.C. *Annual and Multiannual Statistics of the Ecuadorian Electrical Sector*. Agency for the Regulation and Control of Energy and Non-Renewable Natural Resources. 2022, p. 30. Available online: https://www.controlrecursosyenergia.gob.ec/wp-content/uploads/downloads/2022/04/Estadistica2021.pdf (accessed on 30 July 2022).
3. Berghout, T.; Benbouzid, M.; Ferrag, M.A. Multiverse Recurrent Expansion With Multiple Repeats: A Representation Learning Algorithm for Electricity Theft Detection in Smart Grids. *IEEE Trans. Smart Grid* **2023**, *14*, 4693–4703. [CrossRef]
4. Monedero, Í.; Biscarri, F.; León, C.; Biscarri, J.; Millán, R. MIDAS: Detection of non-technical losses in electrical consumption using neural networks and statistical techniques. In *Computational Science and Its Applications–ICCSA 2006*; Springer: Berlin/Heidelberg, Germany, 2006.
5. Chicco, G.; Napoli, R.; Piglione, F. Comparisons Among Clustering Techniques for Electricity Customer Classification. *IEEE Trans. Power Syst.* **2016**, *21*, 933–940. [CrossRef]
6. Angelos, E.W.S.D.; Saavedra, O.R.; Cortés, O.A.C.; de Souza, A.N. Detection and Identification of Abnormalities in Customer Consumptions in Power Distribution Systems. *IEEE Trans. Power Deliv.* **2011**, *26*, 2436–2442. [CrossRef]
7. Rossoni, A. Estimation of Technical and Commercial Losses: Methods Based on Load Flow and State Estimator. Master's Thesis, Federal University of Rio Grande do Sul, Porto Alegre, Brazil, 2014.
8. Monedero, I.; Biscarri, F.; León, C.; Guerrero, J.I.; Biscarri, J.; Millán, R. Detection of Frauds and other Non-Technical Losses in a Power Utility using Pearson Coefficient, Bayesian Networks and Decision Trees. *Int. J. Electr. Power Energy Syst.* **2012**, *34*, 90–98. [CrossRef]
9. Jonathan, M.M.B.; Antonio, D.C. Customer Segmentation Using Electricity Smart Meter Data. Master's Thesis, Higher School of Engineering and Technology, International University La Rioja, La Rioja, Spain, 2022.
10. Llagua, J.L. *Detection of Non-Technical Losses in Special Customers with Telemetering, Based on Artificial Intelligence with Application in EEASA*; National Polytechnic School: Quito, Ecuador, 2023.
11. Pau, G.C. *Design and Implementation of a Classifier Using Neural Networks for a 3D Industrial Inspection System*; Polytechnic University of Valencia: València, Spain, 2019.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.