Smart Grid Stability Prediction with Machine Learning

GIL-VERA VICTOR DANIEL
Faculty of Engineering, Luis Amigó Catholic University,
Trans. 51 A N° 67 B-90,
COLOMBIA

Abstract: - Smart grids refer to a grid system for electricity transmission, which allows the efficient use of electricity without affecting the environment. The stability estimation of this type of network is very important since the whole process is time-dependent. This paper aimed to identify the optimal machine learning technique to predict the stability of these networks. A free database of 60,000 observations with information from consumers and producers on 12 predictive characteristics (Reaction times, Power balances, and Price-Gamma elasticity coefficients) and an independent variable (Stable / Unstable) was used. This paper concludes that the Random Forests technique obtained the best performance, this information can help smart grid managers to make more accurate predictions so that they can implement strategies in time and avoid collapse or disruption of power supply.

Key-Words: analysis; artificial intelligence; control, machine learning; smart grid; stability.

Received: June 28, 2021. Revised: July 15, 2022. Accepted: September 16, 2022. Published: October 6, 2022.

1 Introduction

Smart grids are networks that control power delivery and provide several advantages, including the development and effective management of renewable power sources [1]. They are primarily used to solve energy supply problems by ensuring the transfer of information and electricity between power plants and appliances [2]; they also enable devices to communicate between suppliers and consumers, thus managing demand, preserving the distribution network, reducing costs, and saving energy [3].

In essence, a smart grid has advanced technology and incorporates information and communication technologies (ICT), utilizing technology for metering, communications, and control in the facilities' generating, transmission lines, substations, feeders (circuits), and meters [4]. The objectives of smart grids are; to generate faster performance for the benefit of the end consumer (services, tariffs, quality, and continuity of supply), reduce power outages, increase security and energy efficiency, reduce pollution, help control energy consumption, reduce and prevent outages by anticipating equipment damage and making changes in the electrical transmission path, reduce the vulnerability of transmission networks to attacks or failures and facilitate their rapid location in urban and rural areas [5].

According to [6], modern electric power systems' technical and commercial disturbances are often referred to as "smart grid", encompassing everything integrated into them, what uses the grid services and what interacts with them. On the other hand, [7] defines them as a complex system of technological, electricity trading, and service subsystems articulated to the business, legislative, political, and social sectors. Technically speaking, smart grids are comprised of transmission and distribution networks, production, consumption, and storage facilities, as well as related operational and investment decision-making systems. They also have close ties to other energy sources and domains due to the coupling of sectors and electrification of energy domains like building heating and cooling, transportation, and industrial processes [7]. The key to making the best use of abundant energy resources is smart grid engineering, which enables the efficient dispatching of power generated by hybrid renewable energy sources (RES) over long distances via DC transmission lines using high voltage DC (HVDC) transmission technology [8].

Smart grids enable efficient and dependable energy access using computing and digital communication technologies by integrating renewable energy generation technologies into the transmission system [9]. The reality in which utilities operate, coupled with innate values like business culture, technology, process maturity, and the current market, as well as the socioeconomic and environmental situation of their concession region, are what drive the deployment of smart grids [10]. These generate benefits for utilities, better grid management, increased customer choice, greater understanding of energy use, reduced electricity cost, increased communication with customers and their appliances, use of more renewable energy sources, and integration of electric vehicles [11].

They can offer various advantages that lend themselves to a more stable and effective system, and their primary functions include real-time monitoring and reaction, allowing the system to constantly change to an ideal condition. This is one of its key qualities [12]. Self-healing enables them to identify anomalous signals, carry out adaptive reconfigurations, and isolate disturbances, reducing or eliminating electrical disturbances during storms and disasters. They can also reduce power outages and shorten their duration when they do occur [13]. Rapid isolation enables the system to quickly isolate affected portions of the network from the rest of the system to prevent the spread of outages and enable faster restoration. Anticipation enables the system to automatically search for issues that could cause greater disturbances [14].

While grid operators manage the system's balance, provide supply stability and security, physically connect producers and consumers and facilitate energy transactions, smart grids also provide services that enable an electricity system's efficient and secure running [15]. Smart grids aim to improve the functioning of energy markets, use existing transmission infrastructures more effectively, increase the capacity of renewable energy sources, electric vehicles, heat pumps, and other energy-saving technologies, and give all stakeholders—including small-scale actors like distributed energy resource owners—more flexibility [16]. Fig.1 presents the main benefits of smart grids.

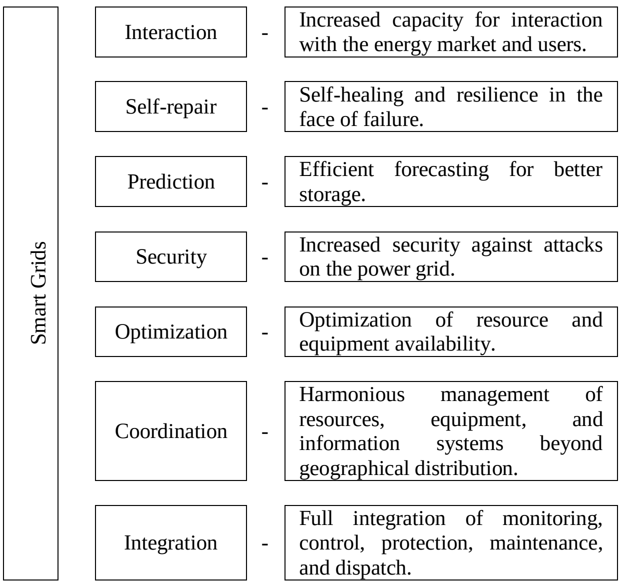

Fig. 1: Benefits of smart grids

In a smart grid, data on consumer demand is gathered, supply circumstances are compared centrally and customers are supplied with pricing information to determine their usage because the entire process is time-dependent, it is crucial to understand and plan for disturbances and fluctuations in energy consumption and production introduced by system participants dynamically, considering not only technical considerations but also how participants react to changes in energy prices [17].

In power system operation and planning, dynamic security assessment and prediction are critical to ensure uninterrupted electricity supply to consumers and improve system reliability [18]. The ability of smart grids to maintain balance over time is referred to as stability, i.e., avoid blackouts regardless of consumer demand (Hz) [19]. Globally, 50 Hz / 60 Hz frequencies are employed in electric power distribution and generation systems, the frequency of the electric signal increases in times of excess generation, therefore, measuring the frequency of the grid at each customer's location is sufficient to give the manager the necessary information on the present grid energy balance, so that it can price its energy supply and alert consumers, while it reduces in times of underproduction [19].

In the review of the state of the art, the scientific databases Scopus and WoS were used, only research articles were considered and the fields of knowledge were delimited to energy, engineering, and computer science, the search period was from 2019 to September 2022. The search equation used was:

TITLE-ABS-KEY ("smart grid" AND "stability" AND ("prediction" OR "forecasting")) AND (LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019)) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (SUBJAREA, "COMP) OR LIMIT-TO (SUBJAREA, "ENER")) AND (LIMIT-TO (SRCTYPE, "j"))

The research question considered was: Q1. How has the prediction/forecasting of smart grid stability been performed?

Most of the identified research related to smart grid stability prediction uses simulated data and deep learning techniques. In the research developed by [20], they claim that measuring the grid frequency of each customer is sufficient to provide the grid manager with all the necessary information about the energy balance so that it can price its energy supply and inform consumers. According to [21], grid stability is affected by the fluctuating nature of renewable energy sources, in this research they employed the Simulated Annealing (SA) algorithm to optimize the hyperparameters and improve the predictability of the grid stability prediction model, which obtained high performance. In the research conducted by [22], they predict the stability of smart grids using multidirectional short-term memory (LSTM). Meanwhile, [23] employed a symmetric non-negative latent factor model based on matrix factorization. In the research developed by [24], they concluded that neural networks can achieve high performance in predicting network stability; however, they claim that most existing machine learning-based approaches can only examine a specific type of stability, and feature engineering is hardly performed due to the limited size of the training data, which may present a misleading indicator of the stability status.

As mentioned above, this paper aimed to train different models to predict the stability of smart grids using machine learning techniques (Random Forests, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, ANN-MLP, Naïve Bayes), compare the performance of each technique and identify the optimal one to predict the stability of this type of grids. The utility of this study in practical applicability is the identification of the optimal technique in terms of accuracy that can help smart grid managers worldwide to make more accurate predictions about the stability of this type of network so that they can implement strategies in time to avoid collapse or breakdowns in the power supply to the nodes that make up the network. A free database of 60,000 observations with information from consumers and producers on 12 predictive characteristics (reaction times, power balances, and gamma-price elasticity coefficients) and an independent variable (stable/unstable) was used. The rest of the paper contains the following sections: in the second section generalities about smart grids are presented, in the third section generalities about machine learning, in the fourth section the method used in the models' training, and the fifth section the results and the discussion. Finally, the paper concludes.

2 Machine Learning

It is a subfield of computer science and artificial intelligence (AI) that focuses on using data and algorithms to simulate how people learn, increasing their accuracy gradually [25]. Machine learning models are used to learn patterns from data in two ways: supervised or unsupervised learning. The former starts from a labeled data set, i.e., the value of the target variable is known, while the latter uses unlabeled data, i.e., the value of the target variable is unknown. Machine learning and data analytics are interdependent and related fields of study that primarily focus on acquiring decisive knowledge [26]. Models are developed using training data and evaluated with test data. Machine learning is currently widely employed in many fields of knowledge to generate predictions and facilitate decision-making. The objective of Machine Learning is to let computers learn how to carry out tasks without being explicitly taught to do so [27]. It is viable to construct algorithms that instruct a machine to carry out the steps required to solve a problem for simpler tasks, but for activities with a greater level of complexity, it is more beneficial to assist the machine in developing its algorithm rather than outlining each step [28]. Machine learning can be used for classification (to predict the membership of a class or label) and regression (to predict a numerical value) tasks. Threesome several specialized tools or programs allow the use of machine learning; some of them are Keras, TensorFlow, KNIME, Shogun, IBM Watson, Apache Mahout, R, Apache Spark MLlib, Weka, Oryx 2, RapidMiner, H20.ai, and Pytorch.

There are several techniques (Random Forests, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, ANN-MLP, Naïve Bayes), that can be employed in the construction of classification or regression models, each of these differing from the others in terms of parameterization. Different research focused on predictive modeling has employed machine learning techniques, specifically, the logistic regression assumes that the independent variable y can take the discrete values {0,1. Equations (1) and (2) describe the relationship between the dependent and independent variables.

$$y = \sigma \sum_{i=1}^{n} (b_0 + \sum_{i=1}^{n} b_i x_i) \qquad (1)$$

$$\sigma(u) = \frac{1}{1 + e^{-u}} \qquad (2)$$

This technique is used primarily for classification tasks. The composition of a sigmoidal function φ(sig): R → [0, 1] over the class of linear functions is the logistic regression class hypothesis [29]. The K-Nearest Neighbors (K-NN) technique saves all the data in the training set and classifies the test sample data based on the Euclidean distance (3), this technique calculates the distance between the data points in the training set, chooses the K entries that are closest to the new data point, and then assigns the label with the highest frequency in the K entries as the class label for the new data point [29].

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3)$$

The Support Vector Machines (SVM) technique, optimally divides two classes by determining the distance between the nearest points in any class' training set [29]. It is possible to map features from a finite-dimensional space into a higher-dimensional space, enabling linear separation despite the dimensional space. This technique provides the best decision boundary that separates the space into classes [30]. The Bayes' Theorem (4), on the other hand, forms the foundation of the Naive Bayes technique, to find the probability when certain other probabilities are known [30].

$$P(\,Y|X\,) = \frac{P(\,Y\,)\ P(\,X \vee Y\,)}{P(\,X\,)} \quad (4)$$

P(Y|X): the probability that Y occurs when X occurs. P(X|Y): the probability that X occurs when Y occurs.
P(Y): the probability that Y occurs.
P(X): the probability that X occurs.

The X variable represents the set of characteristics and is given as *X = (X1, X2, X3, ... Xn).* See equation (5):

$$P(Y|X1,.,Xn) = \frac{P(X1|Y\,)\ldots P(\,Xn|Y)}{P(X1)\ldots P(Xn)} \quad (5)$$

The decision tree technique refers to classifiers, h: X → Y, that move from the root node to a leaf to forecast the label associated with an instance of variables; these are built as branch-like fragments. This technique includes all the predictors with the dependence assumptions between the predictors, and each tree has nodes (root and leaves) that represent the class labels, with the data attribute with the highest priority in decision making being selected as the root node [31]. For the construction of decision trees, it is necessary to calculate two types of entropy using one-attribute (6) and two-attribute (7) frequency tables.

$$E(S) = \sum_{i=1}^{n} (-p_i log_2 p_i) \quad (6)$$

$$E(T,X) = \sum_{c \in X}^{\cdots\cdot} P(c)E(c) \quad (7)$$

The gain function (8) is obtained as follows:

$$Gain(T,X) = Entropy(T) - Entropy(T,X) \quad (8)$$

In equation (8) T represents the target variable, X the feature on which it will be divided, and (T, X) the entropy calculated after dividing the data on the feature X. Random Forests is a technique based on decision trees, which are assembled by bags and trained independently [32]; this technique forecasts an output based on features using a collection of decision trees. The prediction is the outcome of consecutive binary decisions that are divided orthogonally in the multivariate space of variables; in essence, it is a meta-learning of numerous separately built trees [32].

Finally, artificial neural networks are parameterized nonlinear regression models that seek to emulate the way the human brain processes information, i.e., a large number of interconnected processing units that play the role of biological neurons, which work simultaneously to process information. The activation function (softmax, tanh, relu) is in charge of returning output from an input value, often the set of output values in a certain range such as (0,1) or (-1,1) [33]. As universal approximators, multilayer perceptrons are neural network models that can approximate any continuous function. They are made up of perception, which is neurons. A perceptron takes *n* characteristics as input (x = x1, x2, ..., xn), and each of these features is associated with a weight (9). Since a perceptron requires numeric input features, non-numeric input features must be translated before being used [33].

$$u(x) = \sum_{i=1}^{n} w_i \,.\, x_i \quad (9)$$

3 Problem Formulation

Smart grids are the future of energy supply. Their instability can cause problems in the supply of energy to consumption nodes, for this reason, is important to predict their stability. In this type of network, generation must match demand at all times, a reserve must be maintained for immediate outages, and sufficient capacity must be provided for voltage stability.

Identifying the optimal machine learning technique (higher accuracy) to predict the stability of this type of network, allows for building reliable predictive models, which can be used in the prediction of their stability (Stable / Unstable). This study aimed to compare various machine learning approaches to identify the best technique for predicting a smart grid's stability. The database used contains the results of stability simulations of a star network (three consumption nodes and one generation node) as presented in Fig.3.

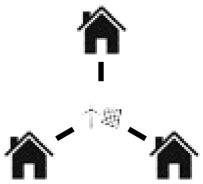

Fig. 2: 4-Node Star Smart Grid

4 Problem Solution and Discussion

A free database accessible from the following link was used to build the models: https://onx.la/46d79, the dataset contains 60,000 observations, twelve primary predictive characteristics, and one dependent variable. The database's structure is shown in Table 1.

Table 1. Database structure

Variable	Description	
V0	Target Variable (Unstable=0/Stable=1)	
V1	Reaction time	Power producer
V2		Consumer 1
V3		Consumer 2
V4		Consumer 3
V5	Power balance	Power producer
V6		Consumer 1
V7		Consumer 2
V8		Consumer 3
V9	Price elasticity coefficient (gamma)	Power Producer
V10		Consumer 1
V11		Consumer 2
V12		Consumer 3

It should be made clear that the price elasticity coefficient refers to the percentage variation in electricity demand in response to small percentage variations in price data, and the reaction time refers to the response time of network participants to adjust consumption and/or production in response to price changes, and the power balance refers to the nominal power produced or consumed at each network node. The models were trained in a ratio of 75/25 (75% for training and 25% for testing), thanks to this division it is possible to identify the accuracy of the models, which were developed in Python using Google Colab. This tool provides free virtual machines with graphics cards to perform machine learning algorithms, which have the same power as platforms such as AZURE or AMAZON Web Services. These Google virtual machines are restarted every 12 hours, allow running and programming in Python in a web browser, do not require configuration, allow free access to Graphics Processing Units (GPUs), and allow sharing content. This tool can be used by students, data scientists, or artificial intelligence researchers.

Colab files are Jupyter notebooks that enable the blending of executable code and rich text in a single document, as well as graphics, HTML, and LaTeX. These notebooks are stored in a Google Drive account and can be shared with others for comments or editing. Colab allows the use of the most popular Python libraries to analyze and visualize data, such as Pandas, Numpy, Matplotlib, Keras, and Tensorflow, among others. This tool allows importing own data from a Google Drive account and GitHub, it also allows importing image datasets, training image classifiers, and evaluating classification and regression models. It should be noted that these notebooks run code on Google's cloud servers, which allows taking advantage of the power of Google hardware regardless of the computer power on which it is used. Table 2 presents the libraries and optimal parameters for each of them.

Table 2. Optimal parameters and libraries

Model	Library & Optimal Parameters
Decision Trees	DecisionTreeClassifier: {'criterion':'gini','class_weight': 'balanced', 'max_depth': 5, 'max_features': 'log2, 'splitter': 'best'}
k-Nearest Neighbors	KNeighborsClassifier: {'n_neighbors': 4}
Logistic Regression	LogisticRegression: {'C': 17, 'max_iter': 9600, 'penalty': 'l2', 'tol': 1e-2}
SVM	SVC: {'C': 120, 'kernel': 'RBF', 'tol': 0.01}
Naive Bayes	GaussianNB: {'max_features': 'auto', 'var_smoothing':1e-8}
Random Forests	RandomForestClassifier: {'n_estimators': 60}
ANN - MLP	MLPRegressor: {'activation': 'relu', 'hidden_layer_sizes': 4, 'learning_rate': 'constant', 'solver': 'adam', 'learning_rate_init': 0.5}

A confusion matrix, which is a matrix representation of the prediction's outcomes made, was used to assess the accuracy of the constructed models (Table 3).

Table 3. Confusion matrix

Current		Predicted	
		Negative	Positive
	Negative	TN	FP
	Positive	FN	TP

TN: values that were negative in the prediction and were also negative in the real values.
TP: values that were positive in the prediction and were also positive in the real values.
FN: values that were negative in the prediction and were not negative in the real values.
FP: values that were positive in the prediction and were not positive in the real values.

From the values of the confusion matrix, the metrics presented in equations (10), (11), (12), (13), (14), and (15) were calculated.

Accuracy: percentage of correct predictions.

$$Accuracy = \frac{(TP + TN)}{Total} \quad (10)$$

Sensitivity, Exhaustiveness, or Recall: percentage of positive cases detected.

$$Recall = \frac{TP}{(TP + FN)} \quad (11)$$

Specificity: percentage of negative cases detected.

$$Specificity = \frac{TN}{(TN + FP)} \quad (12)$$

Precision: percentage of correct positive predictions

$$Precision = \frac{TP}{(TP + FP)} \quad (13)$$

F1 Score: a harmonic measure of precision and completeness, 1 denotes perfect completeness and accuracy.

$$F1 - Score = 2\left(\frac{Precision * Exhaustiveness}{Precision + Exhaustiveness}\right) \quad (14)$$

Receiver operating characteristics curve (ROC): where AUC=1 is ideal, AUC = 0.5 the model cannot differentiate between classes, and AUC = 0 means that the prediction matches the classes.

$$1 - Specificity = \frac{FP}{(TN + FP)} \quad (15)$$

Table 4 presents a summary of the metrics obtained by each of the models evaluated; these metrics are ordered from the model with the best F1 score to the model with the lowest score.

Table 4. Training results

Model	Accuracy	Precision	Recall	Specificity	F1-Score
Random Forest	0.935	0.930	0.967	0.885	0.948
Decision Tree	0.920	0.910	0.962	0.856	0.936
ANN - MLP	0.890	0.870	0.954	0.802	0.910
SVM	0.873	0.858	0.937	0.783	0.896
K- NN	0.780	0.761	0.878	0.659	0.815
Naive-Bayes	0.509	0.537	0.637	0.361	0.583
Logistic Regression	0.476	0.402	0.643	0.365	0.495

This result allows us to identify that the model with the best performance was Random Forest (Accuracy=0.935, F1-Score=0.948). Other models that performed well were Decision Trees (Accuracy=0.920, F1-Score=0.936) and ANN-MLP (Accuracy=0.890, F1-Score=0.910). The ANN-MLP obtained an F1-Score>0.90; however, its Accuracy=0.870, which shows that the ability to make correct positive predictions is lower than the previous two models. The Naive Bayes and Logistic Regression models were the models that registered the lowest capacity to identify negative cases (Specificity), for these two models this metric was lower than 0.40, which makes these models not very efficient when making predictions. Finally, the least efficient model was the Logistic Regression, with an Accuracy=0.476. The F1-Score metric is reliable when the classes are balanced. Fig.5 presents the ROC/AUC curve (Receiver Operating Characteristics Curve) of the Random Forest model, it can be seen that it has an adequate fit in the upper left corner, moving away from the main diagonal.

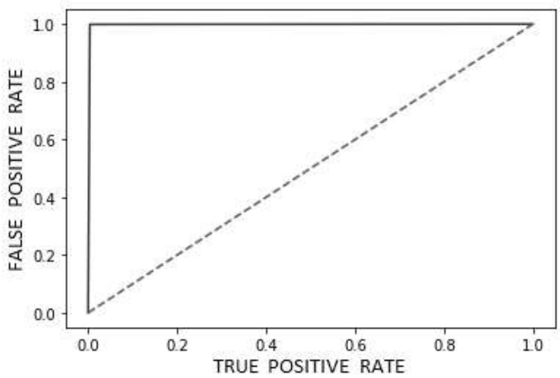

Fig. 5: Positive rates comparison

These findings coincide with the results of the research conducted by [34], where they employed the Random Forest technique to categorize smart grid zones depending on energy usage (high/low), each zone was subdivided into several subzones and assigned to Random Forest branches. In this research, the authors confirm the effectiveness of this technique compared with others (SVM, K-NN, and Naïve Bayes) and conclude that it can identify the exact location of energy availability in minimum time, which allows providing quick responses to grid users.

In the research developed by [35] on the prediction of customer abandonment using machine learning, where they point out that the least accurate techniques are Naïve Bayes and Logistic Regression. Additionally, it is consistent with the study done by [36] on the performance comparison of machine learning algorithms to detect dementia from clinical datasets, where they highlight that the Random Forests technique is one of the most accurate.

It should be noted that the objective of employing this type of technique in predictive modeling is that they discover by themselves patterns that generalize well the data that were not analyzed instead of memorizing data that they learned during training; all accuracy metrics should be evaluated to decide which is the best and not only focus on the accuracy metric. You should also analyze the models that are more separated from the random case, and not only rely on high accuracies since it is possible to have an imbalance in the classes and/or problems of under-or over-training, i.e., if in the smart grid training database most of the measurements are classified in the "Stable" category and only a few in the "Unstable" category, it is easy to guess that a new smart grid measurement will also be "Stable". There must be a balance between the number of "Stable" and "Unstable" measurements in the training database.

5 Conclusion

Smart grid stability needs to be predicted to increase supply reliability, efficiency, and consistency. There are great advantages to implementing smart grids in urban and rural areas, as they encourage the development of renewable energies, contribute to the reduction of polluting gases, reduce environmental impact and damage to the ecosystem caused by the construction of electrical infrastructure works, which is why it is vital to predicting their stability in advance to avoid failures and collapses in the system.

In this study, a comparison of various machine learning techniques for predicting the stability of the smart grid was conducted. The Random Forests technique obtained the best results in the metrics that were studied (Accuracy, Precision, Recall, Specificity, and F1 Score). When one class is less frequent than others, this technique can automatically balance data sets; it is less computationally expensive and does not require a graphics processing unit (GPU). This technique is commonly used in classification exercises since, unlike artificial neural networks, it doesn't need a lot of data to be effective. However, it is not correct to state that this technique is superior to others for making predictions/forecasts in any area of knowledge; the objective of the researcher and the quantity and quality of the available data plays a very important role. In addition, aspects such as non-normalization of the data, non-identification of optimal parameters, and inadequate processing can considerably affect its performance, is very important to normalize the data, fill in missing data with null values and eliminate inconsistencies before training the classification models.

Future research can focus on the construction of constructing predictive models using combined Machine Learning techniques (Bagging, Boosting, Random Subspaces, and others) and compares presented in this work. Finally, Google Colab facilitated the training of models and the identification of the optimal model for predicting the stability of smart grids, as it has advanced libraries for data analysis pre-installed and allows cloud saving and code compilation in blocks.

References:

[1] Lamnatou, C., Chemisana, D., & Cristofari, C., “Smart grids and smart technologies about photovoltaics, storage systems, buildings and the environment”, Renewable Energy, vol.185, p.1376–1391, 2021. DOI: 10.1016/j.renene.2021.11.019.

[2] Pandraju, T. K. S., Samal, S., Saravanakumar, R., Yaseen, S. M., Nandal, R., & Dhabliya, D., “Advanced metering infrastructure for low voltage distribution system in smart grid-based monitoring applications”, Sustainable Computing: Informatics and Systems, vol.35, p. 100691, 2022. DOI: 10.1016/j.suscom.2022.100691.

[3] Stright, J., Cheetham, P., & Konstantinou, C., “Defensive cost-benefit analysis of smart grid digital functionalities”, International Journal of Critical Infrastructure Protection, vol.36, p.100489, 2022. DOI: 10.1016/j.ijcip.2021.100489.

[4] Judge, M. A., Khan, A., Manzoor, A., & Khattak, H. A., “Overview of smart grid implementation: Frameworks, impact, performance and challenges”,

Journal of Energy Storage, vol.49, p.104056, 2022. DOI: 10.1016/j.est.2022.104056.

[5] Panda, D. K., & Das, S., "Smart grid architecture model for control, optimization and data analytics of future power networks with more renewable energy", Journal of Cleaner Production, vol.301, p.126877, 2021. DOI: 10.1016/j.jclepro.2021.126877.

[6] Dileep, G., "A survey on smart grid technologies and applications", Renewable energy, vol.146, p.2589-2625, 2020. DOI: 10.1016/j.renene.2019.08.092.

[7] Selvam, M. M., Gnanadass, R., & Padhy, N. P., "Initiatives and technical challenges in smart distribution grid", Renewable and sustainable energy reviews, vol.58, p. 911-917, 2016. DOI: 10.1016/j.rser.2015.12.257.

[8] Khoury, D., Keyrouz, F., "A predictive convolutional neural network model for source-load forecasting in smart grids", WSEAS Transactions on Power Systems, vol.14, p.181-189, 2019.

[9] Mollah, M. B., Zhao, J., Niyato, D., Lam, K. Y., Zhang, X., Ghias, A. M., Koh, L. & Yang, L., "Blockchain for future smart grid: A comprehensive survey", IEEE Internet of Things Journal, vol.8, No.1, p.18-43, 2020. DOI: 10.1109/JIOT.2020.2993601.

[10] Liu, D., Zhang, Q., Chen, H., & Zou, Y., "Dynamic energy scheduling for end-users with storage devices in smart grid", Electric Power Systems Research, vol.208, p.107870, 2022. DOI: 10.1016/j.epsr.2022.107870.

[11] Yapa, C., de Alwis, C., Liyanage, M., & Ekanayake, J., "Survey on blockchain for future smart grids: Technical aspects, applications, integration challenges and future research", Energy Reports, vol.7, p.6530-6564, 2021. DOI: 10.1016/j.egyr.2021.09.112.

[12] Fan, D., Ren, Y., Feng, Q., Liu, Y., Wang, Z., & Lin, J., "Restoration of smart grids: Current status, challenges, and opportunities", Renewable and Sustainable Energy Reviews, vol.143, p.110909, 2021. DOI: 10.1016/j.rser.2021.110909.

[13] Ashrafi, R., Amirahmadi, M., Tolou-Askari, M., & Ghods, V., "Multi-objective resilience enhancement program in smart grids during extreme weather conditions", International Journal of Electrical Power & Energy Systems, vol.129, p.106824, 2021. DOI: 10.1016/j.ijepes.2021.106824.

[14] Shobole, A. A., & Wadi, M., "Multiagent systems application for the smart grid protection", Renewable and Sustainable Energy Reviews, vol.149, p.111352, 2021. DOI: 10.1016/j.rser.2021.111352.

[15] Emmanuel, M., Rayudu, R., & Welch, I., "Modelling impacts of utility-scale photovoltaic systems variability using the wavelet variability model for smart grid operations", Sustainable Energy Technologies and Assessments, vol.31, p.292-305, 2019. DOI: 10.1016/j.seta.2018.12.011.

[16] Ullah, K., Hafeez, G., Khan, I., Jan, S., & Javaid, N., "A multi-objective energy optimization in smart grid with high penetration of renewable energy sources", Applied Energy, vol.299, p.117104, 2021. DOI: 10.1016/j.apenergy.2021.117104.

[17] Babar, M., Tariq, M. U., & Jan, M. A., "Secure and resilient demand side management engine using machine learning for IoT-enabled smart grid", Sustainable Cities and Society, vol.62, p. p.102370, 2020. DOI: 10.1016/j.scs.2020.102370.

[18] Mukherjee, R., & De, A., "Development of an ensemble decision tree-based power system dynamic security state predictor", IEEE Systems Journal, vol.14, no.3, p. 3836-3843, 2020. DOI: 10.1109/JSYST.2020.2978504.

[19] Tiwari, S., Jain, A., Ahmed, N. M. O. S., Alkwai, L. M., Dafhalla, A. K. Y., & Hamad, S. A. S., "Machine learning- based model for prediction of power consumption in the smart grid- smart way towards the smart city", Expert Systems, p. e12832, 2021. DOI: 10.1111/exsy.12832.

[20] Breviglieri, P., Erdem, T., & Eken, S., "Predicting Smart Grid Stability with Optimized Deep Models", SN Computer Science, vol.2, no.2, p.1-12, 2021. DOI:10.1007/s42979-021-00463-5

[21] Massaoudi, M., Abu-Rub, H., Refaat, S. S., Chihi, I., & Oueslati, F. S., "Accurate Smart-Grid Stability Forecasting Based on Deep Learning: Point and Interval Estimation Method", In 2021 IEEE Kansas Power and Energy Conference (KPEC), p. 1-6, 2021. DOI: 10.1109/KPEC51835.2021.9446196

[22] Alazab, M., Khan, S., Krishnan, S. S. R., Pham, Q. V., Reddy, M. P. K., & Gadekallu, T. R., A multidirectional LSTM model for predicting the stability of a smart grid. IEEE Access, 8, p.85454-85463, 2020. DOI: 10.1109/ACCESS.2020.2991067

[23] Song, Y., Li, M., Luo, X., Yang, G., & Wang, C., Improved symmetric and nonnegative matrix factorization models for undirected, sparse and large-scaled networks: A triple factorization-based approach. IEEE Transactions on Industrial Informatics, vol.16, no.5, p.3006-3017, 2019. DOI:10.1109/TII.2019.2908958

[24] Massaoudi, M., Chihi, I., Sidhom, L., Trabelsi, M., Refaat, S. S., & Oueslati, F. S. (2019, November). Performance evaluation of deep recurrent neural networks architectures: Application to PV power forecasting. In 2019 2nd International Conference on Smart Grid and Renewable Energy (SGRE) (pp. 1-6). IEEE. DOI: 10.1109/SGRE46976.2019.9020965

[25] Zhang, Y., Xin, J., Li, X., & Huang, S., Overview on routing and resource allocation-based machine learning in optical networks, *Optical Fiber Technology*, vol.60, pp.102355, 2020. DOI: 10.1016/j.yofte.2020.102355.

[26] Ibrahim, M. S., Dong, W., & Yang, Q., "Machine learning is driven smart electric power systems: Current trends and new perspectives", Applied

Energy, vol.272, p.115237, 2020. DOI: 10.1016/j.apenergy.2020.115237.
[27] Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K., "Applications of machine learning to machine fault diagnosis: A review and roadmap", Mechanical Systems and Signal Processing, vol.138, p.106587, 2020. DOI: 10.1016/j.ymssp.2019.106587.
[28] Kotsiopoulos, T., Sarigiannidis, P., Ioannidis, D., & Tzovaras, D., "Machine learning and deep learning in smart manufacturing: the smart grid paradigm", Computer Science Review, vol.40, p.100341, 2021. DOI: 10.1016/j.cosrev.2020.100341.
[29] Shalev-Shwartz, S., & Ben-David, S., "Understanding Machine Learning: From Theory to Algorithms". Cambridge University Press, New York, EEUU, 2014.
[30] Wang, M., & Chen, H., "Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis", Applied Soft Computing, vol.88, p. 105946, 2020. DOI: 10.1016/j.asoc.2019.105946.
[31] Naganandhini, S., & Shanmugavadivu, P., "Effective diagnosis of Alzheimer's disease using modified decision tree classifier", Procedia Computer Science, vol.165, p.548-555., 2019. DOI: 10.1016/j.procs.2020.01.049.
[32] Golden, C. E., Rothrock Jr, M. J., & Mishra, A., "Comparison between random forest and gradient boosting machine methods for predicting Listeria spp. prevalence in the environment of pastured poultry farms", Food Research International, vol.122, p.47-55, 2019. DOI: 10.1016/j.foodres.2019.03.062.
[33] Heidari, A. A., Faris, H., Aljarah, I., & Mirjalili, S., "An efficient hybrid multilayer perceptron neural network with grasshopper optimization", Soft Computing, vol.23, no.17, p.7941-7958, 2019. DOI: 10.1007/s00500-018-3424-2.
[34] Durairaj, D., Wróblewski, Ł., Sheela, A., Hariharasudan, A., & Urbański, M. "Random forest-based power sustainability and cost optimization in smart grid", Production Engineering Archives, vol.28, no.1, p. 82-92, 2022. DOI: 10.30657/pea.2022.28.10.
[35] Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C., "A comparison of machine learning techniques for customer churn prediction", Simulation Modelling Practice, and Theory, vol.55, p.1-9, 2015. DOI: 10.1016/j.simpat.2015.03.003.
[36] Miah, Y., Prima, C. N. E., Seema, S. J., Mahmud, M., & Shamim Kaiser, M., "Performance comparison of machine learning techniques in identifying dementia from open access clinical datasets", In Advances on Smart and Soft Computing, Springer, Singapore, p. 79-89, 2021. DOI: 10.1007/978-981-15-6048-4_8.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Victor Daniel Gil Vera has performed the normalization of the database, trained the predictive models in Python, and performed the statistical analysis.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This research was funded by the Universidad Católica Luis Amigó and was one of the results of the research project entitled "Implementation of Smart Grids in Colombia: a multidimensional analysis" - Cost Center [0502020950].