IoT-based SCADA System for Smart Grid Stability Monitoring using Machine Learning Algorithms

Adriana Babilonia Risco
Facultad de Ingeniería Mecánica
Universidad Nacional de Ingeniería
Lima, Perú
ababiloniar@uni.pe

Renzo Iván González Salinas
Facultad de Ingeniería Mecánica
Universidad Nacional de Ingeniería
Lima, Perú
renzo.gonzalez.s@uni.pe

Alhiet Orbegoso Guerrero
Facultad de Ingeniería Mecánica
Universidad Nacional de Ingeniería
Lima, Perú
alhiet.orbegoso@gmail.com

Daniel Leonardo Barrera Esparta
Facultad de Ingeniería Mecánica
Universidad Nacional de Ingeniería
Lima, Perú
daniel.barrera.e@uni.edu.pe

2021 IEEE XXVIII International Conference on Electronics, Electrical Engineering and Computing (INTERCON) | 978-1-6654-1221-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/INTERCON52678.2021.9532668

***Abstract*— This paper proposes a different approach for a Smart Grid Stability Monitoring system using an IoT-based SCADA and Machine Learning algorithms. Different models based on Machine Learning are trained and evaluated using a real dataset to select the best algorithm to predict the electrical grid stability. In addition, this work emphasizes the use of the MQTT protocol to develop an IoT-based platform that connects the company's SCADA interface to the cloud and improves the interaction between the users and the electricity company. The model with the best prediction is Extra Trees algorithm with the best R2 index performance. Compared to previous works, this study used an IoT-based SCADA system and Machine Learning algorithms, in order to improve the interaction between the users and the electricity service provider as it is desired for a Smart Grid.**

Keywords*— *Advanced Metering Infrastructure, Smart Grid Stability Monitoring, Internet of Things.

I. Introduction

A Smart Grid is a digitized electrical network that uses information as well as communication technology to perceive and react against the instant changes produced in consumer usage [1]. In essence, the Smart Grid attempts to provide a grid adapted to the modern world, where they can manage the demand and the increase in fluctuating energy from Renewable Energy Sources (RES) [1-3].

One of Smart Grid's advantages is the ability to establish a two-way communication between the producer and the consumer, and considering the large amount of data involved, reliable and real-time communication is needed [1] [4] [5]. IoT allows any everyday object to connect to the internet, using its data to enhance its performance, and offering more control over its use [6]. This technology is becoming immensely popular across industries such as manufacturing, transportation, energy, retail, healthcare, agriculture, among other areas [7].

Intelligent control and monitoring system for the Smart Grid's network is a priority, and it can be achieved using a SCADA system [8-10]. Its main functions are to collect sensor data measurements, process, and display information, monitor processes, and control local or remote equipment [11]. These functions are also among the principal requirements of the electrical power industry because of their need to monitor the status of the electrical grid and control the electric power transmission and delivery [2]. The next step for control and monitoring systems would then be the integration of IoT and SCADA systems [12].

II. State of the Art

There is a vast number of research papers that propose solutions for the different challenges that currently exist for a Smart Grid implementation [3-6] [8-10] [13]. However, one of these solutions focuses on the fluctuations of the supply caused by RES and the concerns about the protection of the consumer data, sent to a central station to be evaluated. In this study, the concept of Decentral Smart Grid Control (DSGC) is introduced as a solution. This concept refers to a decentralized system that calculates the electrical grid stability using local parameters. In consequence, it does not need to send the consumer data to the electrical company central station [3].

In [13], DSGC limitations are covered, and with new system simulations, the authors discovered useful insight to enhance the model. In [14], MQTT is presented as a flexible and secure option for a suitable communication mechanism to support Distributed Energy Resource (DER) applications. In [4], the authors make a comparative between IoT Protocols for Smart Grid communication, where MQTT stands out due to its low complexity and easy implementation. In [10] the authors developed an IoT-based SCADA integrated with Fog for Distribution Automation system. This helps the user to locate faults in the system using smart meters.

In [6], the authors state that IoT – integrated Smart Grid Systems are already deployed, but with its potential not fully discovered. This work states the importance of using Cloud Computing as a host for the SCADA system, in order to provide access to a collection of shared computing resources. Finally, although in [9] an IoT-based SCADA system was designed and implemented using MQTT protocol and an ESP32 board, it was used for a photovoltaic system's data monitoring. None of the previous works present a solution including an IoT-based SCADA system applied to Smart Grid stability monitoring using the DSGC approach.

This study presents a system prototype using an IoT-based SCADA system and Machine Learning algorithms, which will integrate the current services and improve the interaction between the users and the electricity service provider.

III. Decentral Smart Grid Control, IoT and Machine Learning

A. Definition of DSGC

Decentral Smart Grid Control is a proposal where the electricity price is encoded directly in the frequency of the grid, with this information the decentralized control of the energy balance is performed without the need to implement a large central infrastructure that performs the control. When the energy demand exceeds its generation, the price rises in a way that encourages users to reduce their consumption [3]. The model considers the producers and consumers as rotating machines, then it stablishes a relationship between electricity price and network frequency, the result of the model is shown in (1) [15].

$$\frac{d^2\theta_i}{dt^2} = P_i - \alpha_i \frac{d\theta_i}{dt} + \sum_{j=1}^{N} K_{ij} \sin(\theta_j - \theta_i) - \frac{\gamma_i}{T_i}\big(\theta_i(t-\tau_i) - \theta_i(t-\tau_i-T_i)\big) \ \forall i \in \{1, \dots, N\} \quad (1)$$

Where:

i: is an index that represents the participant number of the grid.
N: is the total number of grid participants.
θ_i y θ_j: are the rotor angles for a specific time of the grid participants i and j.
$\frac{d\theta_i}{dt}$: variation in rotor angle for participant i (s^{-1}).
$\frac{d^2\theta_i}{dt^2}$: it is an indicator of grid stability, a negative value means that the grid is unstable, and a positive value means that it is stable (s^{-2}).
P_i: is the mechanical power produced or consumed (s^{-2}).
α_j: is a damping constant (s^{-1}).
K_{ij}: coupling force between grid participants (s^{-2}).
γ_i: is a coefficient proportional to the price elasticity of each participant (s^{-1}).
t: refers to a specific moment in time (s).
τ_i: is the reaction time of the grid participants to a change in the price of electricity (s).
T_i : is the time interval to measure the average of the frequency and thus obtain the price of electricity from the grid regardless of the reaction time (s).

The work carried out in [13] was based on Eq. (1) to simulate a Four Node Star Grid system, where the central node is a producer, and the other 3 nodes are consumers. They assume certain ranges for the input values such as power (produced or consumed), the coefficients proportional to the price elasticity, and the reaction time of the participants (adaptation time to price changes). The simulation results are available online as *Electrical Grid Stability Simulated Data*, and contain 14 attributes of which 12 are inputs corresponding to the grid characteristics (P_1, P_2, P_3, P_4, γ_1, γ_2, γ_3, γ_4, τ_1, τ_2, τ_3 and τ_4) and 2 outputs, the stability as a category (stable or unstable) and its numerical value (*stab*).

Although the DSGC does not need a central control unit, this is necessary for a company or public regulatory authority to remotely monitor grid devices. This will guarantee security and trust between producers and consumers, as well as the grid stability. For this purpose, it is necessary to have a real-time data visualization system. Among the most prominent systems for Smart Grids there are AMI (Advanced Metering Infrastructure) technology and SCADA (Supervisory Control and Data Acquisition) systems [16].

SCADA systems are an automation tool for various processes that performs control, monitoring, data collection and analysis actions remotely. Its main function is to analyze the data to alert or correct possible errors. It is composed of an HMI, MTU, RTUs / PLCs, sensors, and actuators. They can be easily integrated into the Smart Grids and, properly configured, are safe against cyberattacks [17].

B. Internet of Things (IoT) and Protocols

IoT refers to physical devices that receive and transfer data over wireless networks without human intervention, which is possible thanks to the integration of simple computing devices with sensors in all kinds of objects. A traditional IoT system works by constantly sending, receiving, and analyzing data in a feedback loop.

The IoT protocols commonly used in Smart Grids are CoAP, DDS, MQTT, OPC UA, among others [4]. MQTT (Message Queuing Telemetry Transport) is a publish / subscribe protocol initially designed for SCADA systems, it works over TCP / IP or other network protocols with bidirectional support. It has 3 operation modes called Quality of Service (QoS), QoS0 is the fastest and least reliable, since the data is sent, but does not receive confirmation, with QoS1 it is ensured that the message is delivered at least once and with QoS2 the data has to be sent exactly once, avoiding duplicates.

C. Machine Learning

Machine Learning (ML) is a branch of Artificial Intelligence that creates systems that can learn by themselves. They automatically learn, from a certain amount of training data, to predict future behaviors from new data without the need to find the mechanical equations that govern them [18].There are best-known algorithms as decision trees (CART), regression algorithms (LiR, LoR) and instance-based algorithms (K-NN). Ensembled algorithms combine several of these algorithms into a single predictive model in order to decrease variance (bagging methods) or reduce bias (boosting methods) [19].

With ML, a solution for classification or regression tasks can be performed. For classification, the result is a class or category, while for regression is a continuous number or value. Our simulated data contains both types of results, the stability as a category (stable or unstable) and its numerical value. However, an analysis based on regression problem will be performed.

IV. Comunication architecture

This paper proposes that communication is done using the MQTT protocol with a QoS2. The client sends the characteristics of its grid, for this, it uses a cloud service such as IBM's Watson IoT Platform, which offers functionalities such as device registration, connectivity, quick visualization, data storage. and the creation of dashboards. The electricity company receives this data and predicts the network stability using ML algorithms, which have been implemented in

Python language. Once the prediction is done, it returns a value to the cloud service so that it can be viewed by the customer through a device such as a computer, tablet, or smartphone.

At the same time, all this information is shown in the interface of the company's SCADA system (HMI), designed in Java language, so that it can fulfill its function of monitoring the status of the grid.

V. INTERFACE AND COMMUNICATION DESIGN

A. SCADA

For the SCADA interface, and according to the communication architecture, the electrical grid stability value is obtained from the ML software in a .txt file, the SCADA interface then takes it and displays the different inputs with its calculated stability on the right side of the screen, as shown in Fig. 1. An animated display for the switching on and off of the system is also included, as well as a menu and a plot tab. The ML algorithm performance is analyzed, it graphs both the stability value of the training dataset and the value predicted with ML, showing that the model is working with very few error margins from the original values.

B. IoT and Cloud Connection.

The communication is done using the MQTT protocol implemented using Eclipse Paho MQTT library, this Python library allows the client to publish its electrical network data in JSON format in a topic of IBM's Watson IoT Platform, and the company can subscribe to it and receive the data from client grid.

Once the stability analysis is done with Machine Learning, the same method is used to return the stability prediction. In addition to this, in Watson IoT Platform the data can be visualized in real time through a dashboard using any device with Internet connection, as shown in Fig. 2. The integration with the ML system was done by reading its .txt text output file, which keeps a record of all data received, and is constantly updated with new data.

C. Machine Learning Interface

The Electrical Grid Stability dataset has 10,000 instances, 20% of the data was reserved for an evaluation of the final model developed. The remaining 80% is used for training, where 10-fold cross-validation were used for the standard test setup. The algorithms were evaluated using the MAE, MSE, and R2 metrics. Furthermore, the dataset was standardized using Pipelines and the model was built for each fold of the cross-validation to avoid data leakage.

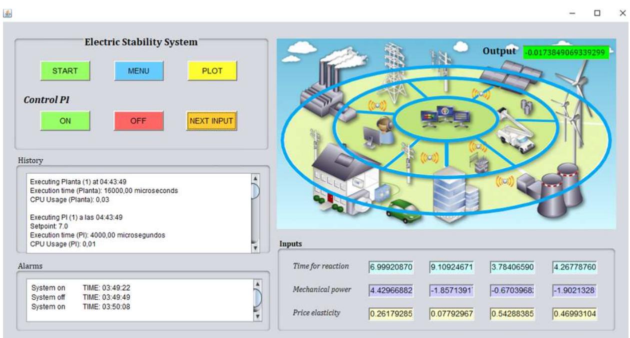

Fig.1. Main window of the SCADA system.

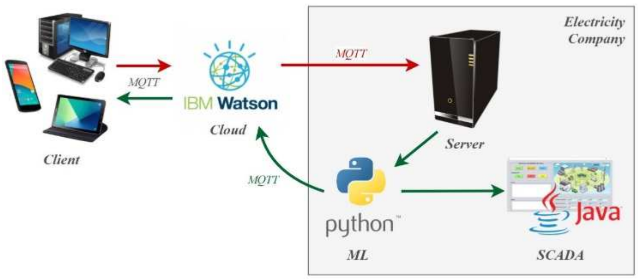

Fig. 2. Smart Grid communication architecture.

The algorithms tested were LiR, LASSO and EN as linear algorithms and CART, SVM and k -NN as non-linear ones. Ensemble algorithms were also tested, two boosting and two bagging methods were evaluated, AdaBoost (AB) and Gradient Boosting (GBM), and Random Forest (RF) and Extra Trees (ET), respectively.

- *Forecasting*

The gradient augmentation model was finally evaluated on our validation dataset. After preparing, standardizing, and training the model with the entire data set, and scaling the inputs for the validation data set, the predictions were generated.

VI. RESULTS

The results of the linear and non-linear algorithms selected using standardized data are shown in Table I.

TABLE I. EVALUATION OF LINEAR AND NON-LINEAR ALGORITHMS

Algorithm	MAE	MSE	R^2
LiR	0.017476	0.000484	0.641672
LASSO	0.031186	0.001354	-0.001411
Elastic Net	0.031186	0.001354	-0.001411
k-NN	0.013604	0.000317	0.765378
CART	0.015358	0.000404	0.701425
SVM	0.031225	0.001357	-0.003325

It can be observed that algorithms have low performances and that the best result is achieved with the k-NN algorithm, approaching an R2 value of 0.765. The number of neighbors for k-NN is fitted, being 5 the default number. Then, the test is performed for all odd values from 1 to 21, including the default value 5. The optimal configuration was achieved using the k-NN algorithm for a number of neighbors k=9, approaching an R2 value of 0.780.

Then, an evaluation of four different assembled algorithms is done with two Boosting and other two Bagging classes, respectively. The results are shown in Table II. It is observed that the Extra Trees and Gradient Boosting algorithms obtain a better approximation with an R2 of 0.912 and 0.903, respectively. In addition, an optimization of the GBM algorithm is performed. The optimization results are shown in Table III.

TABLE II. ASSEMBLED ALGORITHM EVALUATION

Algorithm	MAE	MSE	R^2
Ada Boost	0.014853	0.000318	0.765053
Gradient Boosting	0.008782	0.000131	0.903287
Random Forest	0.009146	0.000142	0.894555
Extra Trees	0.008250	0.000118	0.912898

TABLE III. GRADIENT BOOSTING OPTIMIZATION

Number of booster stages (n)	Gradient Boosting		
	MAE	MSE	R^2
50	0.011542	0.000208	0.846449
100	0.008782	0.000131	0.903287
150	0.007979	0.000112	0.916956
200	0.007690	0.000106	0.921723
250	0.007513	0.000102	0.924618
300	0.007396	0.000099	0.926688
350	0.007315	0.000097	0.928095
400	0.007250	0.000096	0.929284

Finally, an evaluation for this trained model is performed using the validation dataset. These data have not been used in any previous stage. The results are shown in Table IV. The graphical display on the dashboard of the real stability and its prediction is shown in Fig. 3.

TABLE IV. VALIDATION RESULTS

Gradient Boosting with 400 boost stages	MAE	MSE	R^2
	0.00715449	0.00009274	0.93361512

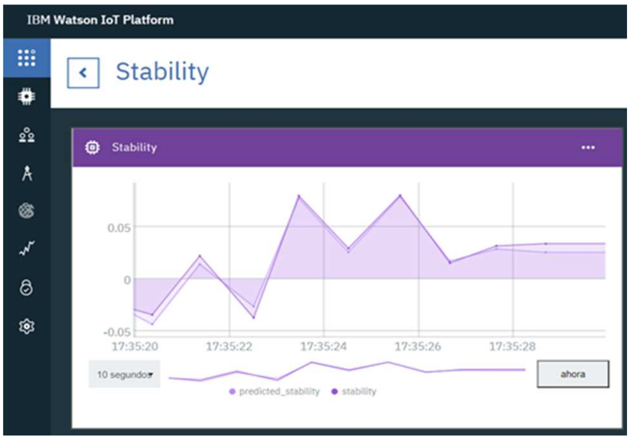

Fig. 3. IBM Watson dashboard.

VII. CONCLUSIONS

The design of this prototype for Smart Grid Stability Monitoring using an IoT-based SCADA system and Machine Learning algorithms has been completed. An architecture based on MQTT protocol with a QoS2 using IBM's Watson IoT Platform was proposed to perform the prediction of the grid stability using ML implemented in Python language. The model with the best prediction was the Extra Trees algorithm with an R2 of 0.912.

Compared to previous works, this study used an IoT-based SCADA system and Machine Learning algorithms, improving the interaction between the users and the electricity service provider as it is desired for a Smart Grid. Future research should dedicate to the development of the SCADA interface to average the individual stability values and provide an accurate prediction of the grid behavior, along with proposals for consumer data security.

REFERENCES

[1] K. Sharma and H. Monga, "Smart Grid: Future of Electrical Transmission and Distribution", International Journal of Communications, Network and System Sciences, vol. 13, no. 04, pp. 45-54, 2020. Available: 10.4236/ijcns.2020.134004 [Accessed 8 April 2021].

[2] The Smart Grid: An Introduction. United States of America: Litos Strategic Communication, 2008, pp. 6-15.

[3] B. Schäfer, M. Matthiae, M. Timme and D. Witthaut, "Decentral smart grid control (2015 New J. Phys. 17 015002)", New Journal of Physics, vol. 17, no. 5, p. 059502, 2015. Available: 10.1088/1367-2630/17/5/059502 [Accessed 8 April 2021].

[4] L. Tightiz and H. Yang, "A Comprehensive Review on IoT Protocols' Features in Smart Grid Communication", Energies, vol. 13, no. 11, p. 2762, 2020. Available: 10.3390/en13112762 [Accessed 9 April 2021].

[5] A. Veichtlbauer, A. Heinisch, F. von Tüllenburg, P. Dorfinger, O. Langthaler and U. Pache, "Smart Grid Virtualisation for Grid-Based Routing", Electronics, vol. 9, no. 11, p. 1879, 2020. Available: 10.3390/electronics9111879 [Accessed 10 April 2021].

[6] A. N. Pramudhita, R. A. Asmara, I. Siradjuddin and E. Rohadi, "Internet of Things Integration in Smart Grid," 2018 International Conference on Applied Science and Technology (iCAST), Manado, Indonesia, 2018, pp. 718-722, doi: 10.1109/iCAST1.2018.8751518.

[7] IoT Analytics Research, Top 10 IoT Application areas 2020. 2020.

[8] K. Sayed and H. Gabbar, "SCADA and smart energy grid control automation", Smart Energy Grid Engineering, pp. 481-514, 2017. Available: 10.1016/b978-0-12-805343-0.00018-8 [Accessed 10 April 2021].

[9] L. O. Aghenta and M. Tariq Iqbal, "Design and implementation of a low-cost, open source IoT-based SCADA system using ESP32 with OLED, ThingsBoard and MQTT protocol", AIMS Electronics and Electrical Engineering, vol. 4, no. 1, pp. 57-86, 2020. Available: 10.3934/electreng.2020.1.57 [Accessed 10 April 2021].

[10] R. Tom and S. Sankaranarayanan, "IoT based SCADA integrated with Fog for power distribution automation", 12th Iberian Conference on Information Systems and Technologies (CISTI), 2017. Available: 10.23919/cisti.2017.7975732 [Accessed 10 April 2021].

[11] R. Radvanovsky and J. Brodsky, Handbook of SCADA/Control Systems Security, 2nd ed. New York: CRC Press, 2016, pp. 31-33.

[12] Y. Saleem, N. Crespi, M. H. Rehmani and R. Copeland, Internet of Things-aided Smart Grid: Technologies, Architectures, Applications, Prototypes, and Future Research Directions, New York, Cornell University Library, 2017.

[13] V. Arzamasov, K. Böhm and P. Jochem, "Towards Concise Models of Grid Stability," 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Aalborg, Denmark, 2018, pp. 1-6, doi: 10.1109/SmartGridComm.2018.8587498.

[14] J. C. Hastings and D. M. Laverty, "Modernizing wide-area grid communications for distributed energy resource applications using MQTT publish-subscribe protocol," 2017 IEEE Power & Energy Society General Meeting, Chicago, IL, USA, 2017, pp. 1-5, doi: 10.1109/PESGM.2017.8274486.

[15] Schäfer, B., Grabow, C., Auer, S., Kurths, J., Witthaut, D., & Timme, M. (2016). Taming instabilities in power grid networks by decentralized control. The European Physical Journal Special Topics, 225(3), 569–582.

[16] Nga, D., See, O., Quang, D., Xuen, C. and Chee, L., 2012. Visualization Techniques in Smart Grid. Smart Grid and Renewable Energy, 03(03), pp.175-185.

[17] Cardwell, L., & Shebanow, A. (2016). The Efficacy and Challenges of SCADA and Smart Grid Integration. Journal of Cyber Security and Information Systems, 1(3), 1–7.

[18] Shaikhina, T., Lowe, D., Daga, S., Briggs, D., Higgins, R., & Khovanova, N. (2015). Machine Learning for Predictive Modelling based on Small Data in Biomedical Engineering. IFAC-PapersOnLine, 48(20), 469–474. doi:10.1016/j.ifacol.2015.10.185

[19] Fayyazifar, N., & Samadiani, N. (2017). Parkinson's disease detection using ensemble techniques and genetic algorithm. 2017 Artificial Intelligence and Signal Processing Conference (AISP).