Customer Experience-Centric Network Investment and Interventions Through AI

Maximizing Customer Experience Impact of Network Interventions and Investments with Generative AI and Machine Learning

A technical paper prepared for presentation at SCTE TechExpo24

Juan David Rodriguez Lamus
Director of Business Analytics
Liberty Latin America
juan.d.rodriguez@lla.com

Mauricio Romero
Advanced Analytics Senior Director
Liberty Latin America
mromero@lla.com

Table of Contents

List of Figures

List of Tables

1. Introduction

Liberty Latin America has launched AI-based Cx Tech, a program that aims to understand how technical experience influences customer experience and churn. Traditionally, network engineering addressed user experience issues by finding and troubleshooting potential broadband disruptions using metrics like Codeword Error Rate. The AI-based Cx Tech program enhances this by using advanced analytics, big data, and machine learning to find critical events and metrics in HFC and FTTH networks.

However, telcos often struggle to correlate direct network KPIs with churn because the KPIs are just the beginning of the problem. A customer's decision to churn is also heavily influenced by how effectively the issue is addressed afterward. This requires a shift towards a more comprehensive approach in network intervention strategies.

Different customers have different usage profiles, and the same kind of network issues can affect them differently. Therefore, customer-centric prioritization is essential to address these variances in impact effectively. The AI-based Cx Tech program incorporates this perspective, ensuring that interventions are tailored to the specific needs and experiences of diverse customer segments.

A key tool from this program, Lighthouse, is designed to optimize network interventions to maximize the impact on reducing customer technical calls. Recently, Lighthouse's HFC network segmentation was enhanced with generative AI call classifiers from customer transcripts and threshold optimizers from machine learning models. By using Bayesian Optimization with Gaussian Processes and Genetic Algorithms, this improvement has aligned segmentation more closely with customer experience, thereby improving service quality and satisfaction.

Besides improving the understanding of network performance, there is a need for a proactive network intervention redesign. Such a redesign enables prompt reactions to customer pain points, further enhancing the effectiveness of technical solutions and ultimately reducing churn.

2. The Impact of Technical Call Reiteration on Churn

In 2022, the Advanced Analytics and Technology and Information teams at Liberty Latin America (LLA) started the Cx Tech project. Launched by the Chief Customer Office and the Chief Technology and Product Office, this collaborative effort aimed to deeply understand the complexities of technical experience using innovative analytics and engineering tools. The project spans a comprehensive scope, analyzing customer experiences end-to-end from device configuration, through the access network and core systems, to large-scale outages.

While traditional network KPIs target potential issues from a purely engineering perspective, they often fall short in fully addressing all customer needs and usage profiles. To bridge this gap, the approach has shifted towards using customer technical support calls as primary indicators of problems, following the principle that "customer complaints often signal deeper issues."

As part of this strategy, the Cx Tech collaboration integrated in 2023 Generative AI call classifiers along with churn databases to delve into the effects of repeated technical calls on customer churn. GenAI-enabled analytics not only discern which calls were indeed technical support interactions but also to capture essential symptoms mentioned by customers such as "slow browsing," "connectivity intermittence," and "streaming buffering."

Our analysis for one of LLA's markets, depicted in Figure 1, showed that churn rates climb dramatically with repeated technical calls. Specifically, users who called four or more times to resolve technical issues showed a churn rate 4.4 times higher than the market average. This insight has revolutionized the way technical support is measured and analyzed at LLA, shifting from traditional metrics like Mean Time to Repair (MTTR) and Mean Holding Time (MHT) to focusing on minimizing call reiteration and achieving resolution on the first try.

This proactive approach ensures that technical support becomes more aligned with the actual experiences and frustrations of customers, thereby reducing churn and enhancing overall customer satisfaction.

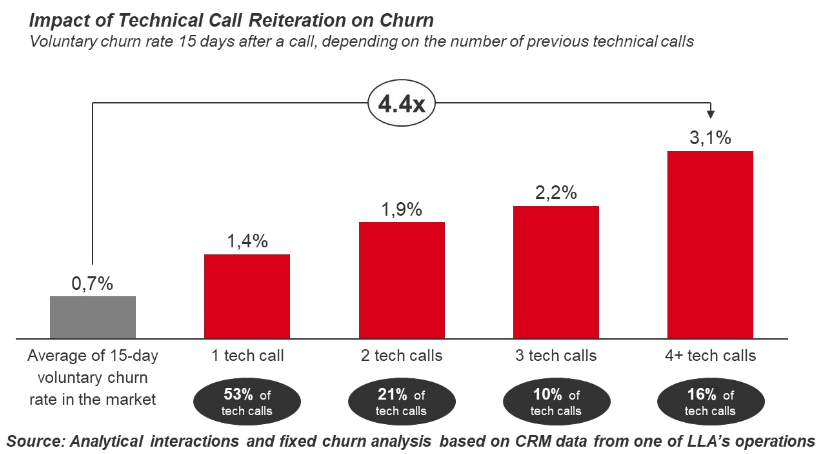


Figure 1 – Impact of technical call reiteration on churn

3. Concentration of Customer Calls in HFC Fiber Nodes

The concentration of customer technical calls was also analyzed by the Cx Tech team. It was concluded that technical calls tend to be considerably concentrated in nodes. Figure 2 shows the distribution of technical calls and nodes for different call rate ranges. As shown in the graph, the study showed that 14.6% of the nodes concentrated 31% of the technical calls in the network while 15.7% of nodes had no technical calls in one of LLA's operations. Other operations also showed concentration of calls in nodes.

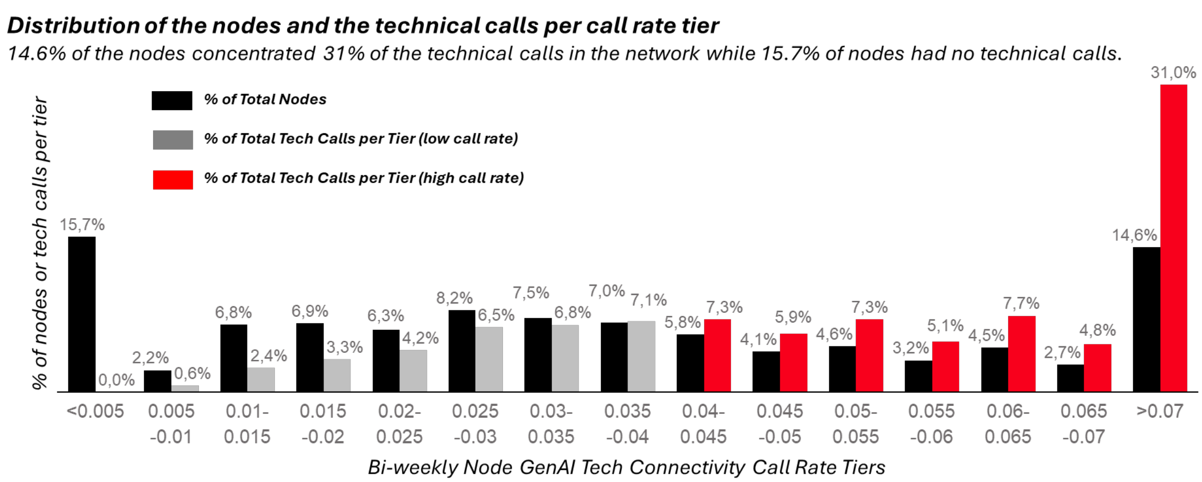


Figure 2 - Distribution of Tech Calls and Nodes per Call Rate Tier in One of LLA's Operations

4. AI based Cx Tech and the Lighthouse Tool

One of the chapters of the Cx Tech program is Lighthouse, an initiative that focuses on understanding the health of the access network and its implications for the customer-perceived experience. With the expansion of machine learning and AI applications, many vendors, operators, and academics have started to experiment with novel approaches to better diagnose issues and align prioritization with customer experience [1].

The Cx Tech teams have developed a suite of machine learning predictive models to find users with high propensity to call in the future. Additionally, a multidimensional segmentation model was developed using data from the Servassure® NXT and Viavi XperTrack® management systems, which houses over 2 billion records per country. The following models were implemented and evaluated:

1. HFC and FTTH User Propensity to Call XG-Boost Predictive Model: Trained with cable modem data extracted from Servassure® management system, the OLTs and customer technical calls, this model predicts the propensity to call within a 30-day window following the inference.

2. HFC Node XG-Boost Regression Model: This model predicts the call rate of a node within a 15-day window following the inference.

3. Sybil - Interaction Transformer sequence predictor for churn: This model calculates the propensity to churn from a user using GenAI post call analytics multidimensional data from the earlier calls generated by the engine.

4. Lighthouse HFC Network Classification Algorithm: Uses selected KPIs and transformations from the machine learning models to segment network nodes based on relevant KPIs and call rates. Calls used to train the model are extracted from the GenAI post-call analytics engine to ensure the customer complaint is related to connectivity. Metrics and thresholds are selected using a combination of Bayesian Gaussian Optimizer and Genetic Algorithms.

Customer experience analysis found that traditional engineering KPIs were effective at finding network components at very critical performance levels but struggled to differentiate performance issues that were bad enough to cause customer complaints yet not at a critical level. To address this, traditional feature engineering was employed to include time series and added data transformations. This led to the development of a new set of features to be assessed with supervised learning algorithms using tech calls as the classification label.

Additionally, SHAP values from the machine learning models have enabled the identification of new features and KPIs that were not part of the traditional engineering metrics. These findings provide deep insights into less obvious yet significantly impactful network behaviors affecting customer experience.

Besides standard metrics and thresholds recognized in industry best practices, the analytics team incorporated several interesting findings from the machine learning models' implementation and later analysis:

- **Null Values in Reported Metrics:** Databases showed null values intermittently, even when modems were online, which significantly increased the propensity for customers to call.

- **Impact of Upstream Deviations:** Upstream deviations had a much greater impact on the propensity to call than previously thought.

- **Partial Service Events:** Analysis found that many null values were caused by "partial service" events, where modems temporarily stop using part of the spectrum due to issues, leading to reduced capacity and potential service disruption.

- **Limitations of CER Reporting:** Many impactful events were not captured by Codeword Error Rate (CER) metric, highlighting the need for more metrics to troubleshoot customer experience issues.

- **Importance of SNR Variability:** High variability in Upstream and Downstream Signal-to-Noise Ratio (SNR) levels, even when overall SNR range was acceptable, correlated with increased propensity to call.

- **Number of Working Carriers in Upstream:** Upstream connection impairment can result in a user losing all but one carrier and high Tx power levels for the remaining carrier, users in this condition have very high propensity to call, even when no codeword errors are present and SNR is in acceptable range.

- **Generalized and Persistent CCER:** Persistent Correctable Codeword Errors (CCER) around partial service events signaled more significant underlying issues not captured by standard metrics.

Traditional metrics and performance systems did not prioritize tracking some of these events, which proved crucial for understanding perceived customer experience. This includes the frequency and length of partial service events, SNR variability, generalized CCER reporting across nodes, and the number of working carriers. Figure 5 illustrates an example of a frequent recaller experiencing intermittent partial service and its representation in raw data.

Freq	Time	CER	CCER	SNR	RX	TX	T3	T4	Partial Service
26000	9/05/2024 4:05	0	0	39	9	45	0	0	none(0)
26000	9/05/2024 5:05	0	0	38	9	45	0	0	none(0)
26000	9/05/2024 6:05	0	0.009 3	37	9	51 5	0	0	none(0)
26000	9/05/2024 8:05		2	37	8				t4Timeout(600), usStationMaintFailur 1
26000	9/05/2024 9:05			40 4	8				t4Timeout(600), usStationMaintFailur
26000	9/05/2024 10:05			0	9				t4Timeout(600), usStationMaintFailur
26000	9/05/2024 12:05			38	9				other(1)
26000	9/05/2024 14:05			42	9				none(0)
26000	9/05/2024 15:05	0	0	40	9	46			none(0)
26000	9/05/2024 16:05	0	0	40	9	45	0	0	none(0)
26000	9/05/2024 17:05	0	0	41	9	45	0	0	none(0)
26000	9/05/2024 18:05	0	0.083 3	37	9	45	0	0	none(0)
26000	9/05/2024 19:05	0	0.033	38	9	45	17	0	none(0)
26000	9/05/2024 20:05	0	0	39	9	45	0	0	none(0)
26000	9/05/2024 21:05	0	0	39	9	45	0	0	none(0)
26000	9/05/2024 22:05	0	0	38	9	45	0	0	none(0)

***May 9th**: The customer reported intermittent internet connectivity issues, with the service coming and going since the morning.*

Partial Service Event Impact	Other anomalies that are not supposed to be service impacting
1 Partial Service	3 Correctable Errors (CCER)
2 Null values in critical metrics	4 SNR in good theorical levels with variability
	5 Upstream Tx Power Spike

Freq	Time	CER	CCER	SNR	RX	TX	T3	T4	Partial Service
32400	12/05/2024 6:05	0	0	42	9	46	0	0	none(0)
32400	12/05/2024 7:05	0	0	38 4	2	51 5	0	0	continue(5)
32400	12/05/2024 8:05			41	9	46		0	none(0)
32400	12/05/2024 9:05		2	42	9				t4Timeout(600), usStationMaintFailur 1
32400	12/05/2024 10:05			41	9				t4Timeout(600), usStationMaintFailur
32400	12/05/2024 12:05			39	9				other(1)
32400	12/05/2024 13:05			42	9	45			none(0)
32400	12/05/2024 14:05	0	0	42	9	46	0	0	none(0)
32400	12/05/2024 15:05			41	9	46		0	none(0)
32400	12/05/2024 16:05	0	0	42	9	46	0	0	none(0)
32400	12/05/2024 18:05	0	0	41	9	46	1	0	none(0)
32400	12/05/2024 19:05	0	0	41	9	46	0	0	none(0)
32400	12/05/2024 20:05	0	0.01 3	41	9	46	0	0	none(0)
32400	12/05/2024 21:05	0	0.015	42	9	46	0	0	none(0)
32400	12/05/2024 22:05	0	0.001	42	9	46	0	0	none(0)
32400	12/05/2024 23:05	0	0.009	42	9	46	0	0	none(0)

***May 12th**: The customer contacted the call center to report intermittent internet connectivity issues they had experienced for four days.*

Partial Service Event Impact	Other anomalies that are not supposed to be service impacting
1 Partial Service	3 Correctable Errors (CCER)
2 Null values in critical metrics	4 SNR in good theorical levels with variability
	5 Upstream Tx Power Spike

Figure 3 - Raw data captured from the cable modem from a frequent caller

5. Lighthouse HFC Detailed Network Classification Algorithm

Industry vendors offer high-quality and sophisticated troubleshooting tools, such as Servassure® NXT and Viavi XperTrack®, which are extensively used in LLA to support and operate the network. However, the default graphical interfaces and reports provided by these tools may not include some indicators found by machine learning models, and the aggregation of their time series data may require further transformations to better correlate with customer technical claims. Additionally, these systems have hundreds of configurable thresholds that must be customized by telecom operators [2] [3] [4] [5].

The Lighthouse segmentation tool was developed to complement the traditional technical troubleshooting toolset. It adds more information with different time series transformations and machine learning-based adjusted thresholds to maximize its correlation with customer calls. Lighthouse is based on three principles agreed upon by the tech, analytics, and care teams:

- **Consider Customer Technical Calls**: Nodes flagged with selected indicators and thresholds should have higher call rates, ensuring interventions maximize impact on customer experience and churn.
- **Affect Multiple Users:** The segmentation should focus on issues affecting multiple users, rather than individual in-home issues, to ensure effective and efficient network interventions.
- **Frequent Issues Over Time:** The segmentation targets chronic or frequent issues, updated every 15 days, ensuring that flagged nodes have recurring problems even if they are intermittent.

The construction of the segmentation and resulting interventions involves 6 steps, explained below:

a) **Fiber Node technical call rate:** Originally, Cx Tech models were trained using technical tickets recorded by call center agents, which could include non-connectivity issues (e.g., TV interface, remote control, Wi-Fi password changes). Liberty Latin America recently implemented "GenAI post-call analytics” to improve the understanding of customer calls using Large Language Models. The Generative AI post call analytics architecture, is comprised of an inhouse developed pipeline that automatically captures all call recordings and whatsapp transcripts from the call center, produces transcripts from the audio recordings, enrich data to include the calling customer, calling number, agent, country, timestamp of the call can be stored, and sends the resulting structures transcripts with metadata to a Bedrock environment in LLA’s AWS environment.

   The call is then processed in Bedrock using Anthropic Claude Large Language Model to find the reason for calling (intent), call resolution, generate a set of alarm flags such as threat to cancel or negative sentiment, and create a summary of the entire call.

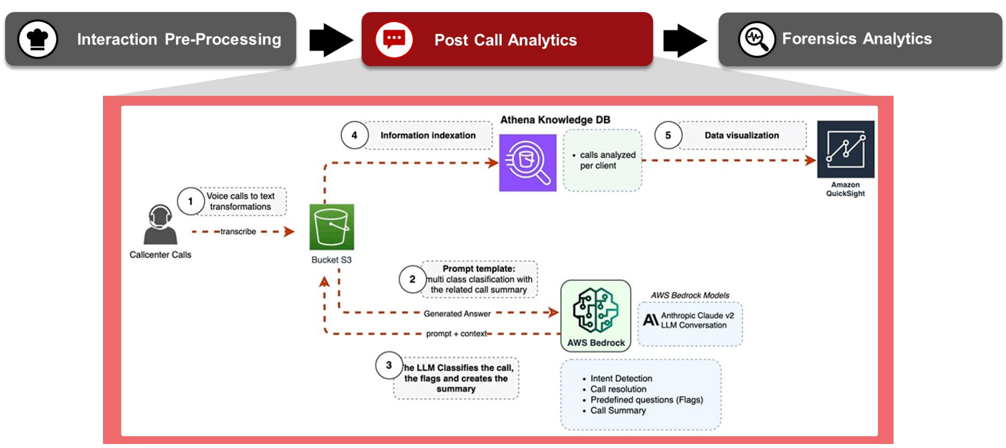
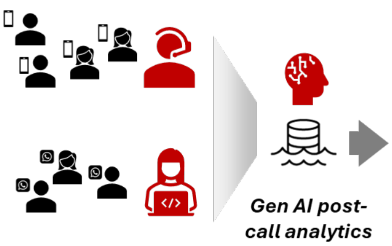


Figure 4 - Liberty Latin America Generative AI Post Call Analytics high level architecture

   The output from GenAI post call analytics is then filtered to include only calls that are related to Internet connectivity and grouped by fiber node to calculate the Node Technical Call Rate as described by Figure 5.



Interact - ID	Fiber Node	Gen AI intent Classifier	Valid call
A0001	FN01	Internet Loss	1
A0032	FN01	Fiber cut	0
A0053	FN01	Internet Intermittent	1
A0045	FN01	TV	0
A0045	FN01	Billing	0

$$\text{Node Call Rate} = \frac{sum(GenAI\ connectivity\ calls)}{Unique\ users\ in\ the\ node}$$

Fiber Node	Unique users	Tech Connectivity calls	Tech call rate
FN01	200	20	0.10
FN02	150	3	0.03
FN03	400	20	0.05
FN04	170		0.15

Figure 5 - Construction of the GenAI technical interaction call rate per Fiber Node

b) **Metrics selection and calculation of Affected Hourly Measurements:** Candidate tech indicators are selected based on the relevant features found by machine learning user models including:

- Partial service events
- Number of functional upstream carriers
- Variance and valued from Upstream SNR
- Number of measurements above Tx Upstream value
- Tx Upstream variance
- Number of measurements above Rx Downstream value
- Rx Downstream variance
- Number of measurements above CER % value (Upstream/Downstream)
- Number of measurements above CCER % value (Upstream/Downstream)
- Number of events above T3 counts.
- Number of events above T4 counts.

A first threshold (threshold 1) detects "Affected Hourly Measurements" (e.g., 1% CER). The percentage of affected measurements per hour is calculated for each indicator, carrier and modem, then, the percentage of affected hour measurements are calculated over the total number of measurements. Using binary flags over thresholds prevents outlier values from skewing the results, unlike average aggregations.

$$\begin{gathered}\textit{\% Affected hour measurements} \\ \textit{(per indicator, day per modem and carrier)}\end{gathered} = \frac{(Number\ of\ affected\ hour\ measurements)}{Total\ measurements\ per\ day}$$

Indicator	Threshold 1	hour	Modem	Frequency (MHz)	Value	Affected hour-measurement
Partial service	>1	Apr-01 15:00	1835D14	26.0	1	1
Partial service	>1	Apr-01 16:00	1835D14	26.0	0	0
CER	>1%	Apr-01 15:00	1835D14	26.0	2%	1
CER	>1%	Apr-01 16:00	1835D14	26.0	0.1%	0

Figure 6 - Calculation of Affected Hourly Measurements.

c) **Daily Affected Modem calculation**: A second threshold flags modem-frequency pairs with a high percentage of deviations during the day (e.g., 40%). Setting a daily flag ensures the calculation is not skewed by users with an outlier day, as it counts as 1 even if all measurements were deviated on that day. This approach helps find persistent issues in the next aggregation phase. The worst-performing carrier is then selected for each node and indicator.

$$\begin{gathered}\textit{\% Daily affected modems} \\ \textit{per indicator, node and carrier}\end{gathered} = \frac{(Number\ of\ Daily\ Affected\ Modems)}{Total\ day\ modems} \;\Rightarrow\; \begin{gathered}\textit{\% Daily Affected Modems} \\ \textit{per indicator, and node}\end{gathered} = \max_{\text{over carriers}} (\%\ Daily\ Affected\ Modems)$$

Indicator	Day	Modem	Fiber Node	Frequency (MHz)	Deviated hours	Non-deviated hours	Threshold 2	Deviated %	Daily Affected Modem Flag
Partial service	Apr-01	1835D14	FN01	26.0	3	21	>4%	12.5%	1
Partial service	Apr-01	1835D14	FN01	32.4	0	24	>4%	0%	0
CER	Apr-01	1835D14	FN01	26.0	5	19	>20%	20.8%	1
CER	Apr-01	1835D14	FN01	32.4	1	23	>20%	4.2%	0

Figure 7 - Calculation of Daily Affected Modem flag.

d) **Impairment flag for the node per indicator:** A third threshold flags nodes with a high percentage of daily affected modems for each indicator. The percentage of deviated day-modems over all day-modem measurements is calculated for the preceding 7 days for each indicator for the worst carrier. This calculation ensures that nodes with impairment flags have been affected for a significant share of the users and during several days.

Indicator	Fiber Node	Frequency (MHz)	Deviated day-modems	Non-deviated day-modems	Threshold 3	Deviated %	Deviated node	
Partial service	FN01	26.0	280	1120	>10%	20%	1	Worst carrier selected
Partial service	FN01	32.4	70	1330	>10%	5%	0	
CER	FN01	26.0	140	1260	>30%	10%	0	Worst carrier selected
CER	FN01	32.4	0	1400	>30%	0%	0	

Figure 8 - Calculation of Node Impairement flag.

e) **Categorize and execute interventions:** Nodes are classified into four categories based on deviation flags and ticket rates:

- **Critical Nodes:** High call rates and at least one node impairment flag. These nodes are prioritized and addressed by specialized teams. Found deviations are certified. Maintaining a small number of critical nodes has proven effective in reducing call rates and technical-driven churn.
- **High Call Rate and No Generalized Node Impairments:** Likely nodes with isolated user issues. These cases are reviewed by technical support and treated individually, as they do not affect the entire node.
- **Deviation Flags and Low Call Rate:** These nodes often had high call rates in the past but were not resolved in time. They are assigned to an intervention team with independent capacity, ensuring they don't compete with critical nodes.
- **Low Ticket Rates and No Deviations:** Nodes with minimal issues and no significant deviations.

Nodes recently intervened are quarantined to avoid actioning on nodes that are being stabilized.

f) **Coordination with call center operations:** The resulting network segmentation and Daily Affected Modem counts are shared with call center operations. Users with a high percentage of Daily Affected Modem counts are automatically flagged daily in the call center agent display. This allows them to be escalated to advanced support with priority when they call. Agents receive a brief on the customer's symptoms, found deviations, and the node's status. Additionally, advanced support technicians are informed about nodes marked as "critical" to prevent ineffective troubleshooting and avoid unnecessary truck-roll dispatches.

6. Lighthouse Machine Learning Optimization Implementation

Optimizing segmentation thresholds is complex due to the highly non-linear, nested, and time-consuming nature of the objective function. The function requires optimizing 82 parameters across a wide range of possible values, making exhaustive testing computationally expensive. This situation resembles the optimization of hyperparameters in Neural Networks, a problem efficiently addressed by machine learning algorithms like Bayesian Optimization and Genetic Algorithms [6].

6.1. Bayesian Optimization

Bayesian Optimization evaluates the target function using a Gaussian Process combined with Bayesian principles. It builds a probabilistic model with a first set of points and then optimizes a simple acquisition/utility function using the posterior distribution. This approach allows smart testing of thresholds, aiming for a positive marginal impact with each try, thus avoiding the need for a greedy grid search. It can efficiently optimize 82 parameters with more than 10 possible values without running all combinations [7].

6.2. Genetic Algorithms

Genetic Algorithms are effective for optimizing highly non-linear and complex functions. These algorithms resemble natural selection processes, including mating, mutating, and selection. The process begins with a first "population" of solutions, scoring the function, selecting the best options, and producing mutations to improve performance. Key parameters include [8] [9]:

- **Mutation Probability:** Defines the likelihood of a parameter changing randomly, enabling the exploration of different areas and avoiding local minima paths.
- **Crossover Probability:** The likelihood that characteristics of a parent cohort pass to the next generation.
- **Parents Proportion:** The share of possible solutions passed to the next generation in each iteration.

Combining these parameters ensures a balance between efficient searching and minimizing the risk of ending with a suboptimal local minimum.

6.3. Optimization Architecture

The implementation of the machine learning optimization module is divided into two stages:

Stage 1 – Optimization of Daily Deviated Modem Flags: In this stage, Daily Affected Modem flags are optimized for each metric, focusing only on flags with good classification capabilities and statistically significant differences in calls between flagged and non-flagged groups. This ensures that only the most relevant flags are used.

Stage 2 – Optimization of Threshold 3: Threshold 3 is the minimum percentage of Affected Daily Modems required for a node to be flagged as impaired.

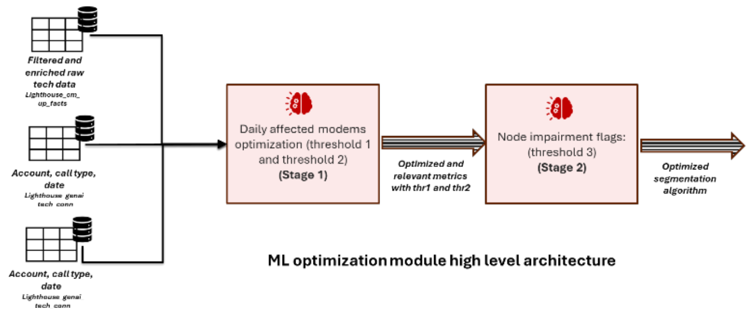


Figure 9 - Machine Learning optimization stages high level architecture

The target function for optimization is constructed by multiplying the "lift" by the "compensated share of positives" as detailed in Figure 10:

- **Lift:** Calculated by dividing the call rate of positives by the average call rate.
- **Compensated Share of Positives:** Calculated by dividing the number of positives by all measurements, then subtracting a compensation function. This function disincentivizes solutions with extremely low or high positive share values.

$$\textit{Target Funcion} = \textit{Call rate lift} \times \textit{Compensated Share of positives}$$

$$\textit{Call rate lift} = \frac{\textit{Call rate positive}}{\textit{Average Call rate}} \qquad \textit{Positive Share} = \frac{\textit{Number positives}}{\textit{All measurements}}$$

$$\textit{Compensated Share of positives} = \textit{Positive Share} - \textit{Comp function(Pos. Share)}$$

Figure 10 - Target function for machine learning optimization

Statistical Testing: In each optimization stage, results are tested for statistical significance:

- **Daily Affected Modem Thresholds:** Tested using a Chi-square test over the number of calling users for modems flagged as positive and negative.
- **Node Impairment Thresholds:** Evaluated using a T-test to compare the mean call rate of positive and negative nodes.

7. Results and Optimized Thresholds

7.1. Resulting thresholds from machine learning optimization

Error! Reference source not found. displays the resulting values for Threshold 1 and Threshold 2, which are used to flag Daily Affected Modems for the selected metric. Consistent with findings from earlier machine learning models, customers generally show greater sensitivity to deviations in upstream metrics. Partial service events, users connected to only one upstream carrier, and high variability in SNR and TX metrics are good predictors of customer-affecting issues.

Uncorrectable codeword error rate (CER) is the metric traditionally tracked to find customer-affecting issues. The Bayesian optimization results showed that downstream CER is perceived by customers at very low values (0.1% for downstream). Users with such low CER values in 6% or more of the daily measurements had 2.77 times higher call rate than the mean and represented 1.32% of the daily samples.

Results for OFDM carrier (DOCSIS® 3.1 downstream) were similar, users with low levels of CER (0.1%) had 8.01 times higher call rate but represented only 0.12% of the daily samples. On the contrary, Upstream CER was not found to be statistically significant by either of the two algorithms. As shown in the frequent caller example, this might be explained because users with severe upstream issues tend to experience intermittent partial service events triggered by T4, where the CER metric is not available.

Table 1 - Metrics selected by the machine learning optimization for Daily Affected Modem

Thr_1	Thr_2	score	metric	lift	positive_share	positive_call_rate	negative_call_rate	p_chi2	algorithm	metric_group	Interpretation
	4.00	0.231	partial_service_flag	4.33	6.94%	0.81%	0.14%	0.00E+00	Bayesian optimization	upstream	Users with partial service in 4% of the measurements or more have 4.33x higher call rate
	5.00	0.046	snr_stdev	3.82	1.61%	0.71%	0.18%	2.68E-146	Bayesian optimization	upstream	Users with more than 3.82 dB of standard deviation in snr in a day have 3.82x higher call rate
28.85	14.60	0.028	snr	2.93	1.45%	0.54%	0.18%	4.40E-62	Genetic optimization	upstream	Users with less than 28.85 dBm of snr in 14.6% of the measurements or more have 2.93x higher call rate
	4.00	0.045	one_carrier_flag	1.88	5.09%	0.35%	0.18%	2.55E-47	Bayesian optimization	upstream	Users with only one carrier in 4% of the measurements or more have 1.88x higher call rate
	0.50	0.029	tx_stdev	1.42	6.96%	0.26%	0.18%	5.74E-16	Bayesian optimization	upstream	Users with more than 0.5 dB of standard deviation in tx in a day have 1.42x higher call rate
9.68	4.00	0.016	rx_positive	1.26	6.09%	0.23%	0.18%	2.62E-06	Bayesian optimization	upstream	Users with more than 9.68dBm Rx in 4% of the measurements or more have 1.26x higher call rate
51.28	4.01	0.007	tx	1.14	5.05%	0.21%	0.18%	2.06E-02	Bayesian optimization	upstream	Users with more than 51.28dBm tx in 4.01% of the measurements or more have 1.14x higher call rate
1.01	4.14	0.008	t3	1.11	7.32%	0.21%	0.18%	3.74E-02	Bayesian optimization	upstream	Users with 1 T3 in 4.14% of the measurements or more have 1.11x higher call rate
0.10	6.32	0.023	cer	2.77	1.32%	0.48%	0.17%	4.47E-44	Bayesian optimization	downstream	Users with more than 0.1% cer in 6.32% of the measurements or more have 2.77x higher call rate
- 8.87	4.10	0.016	rx_negative	2.62	1.01%	0.45%	0.17%	6.82E-29	Bayesian optimization	downstream	Users with less than -8.87 dB of rx in 4.1% of measurements in a day have 2.62x higher call rate
0.10	4.14	0.025	ccer	2.50	1.67%	0.43%	0.17%	1.89E-40	Bayesian optimization	downstream	Users with more than 0.1% ccer in 4.14% of the measurements or more have 2.50x higher call rate
36.27	12.74	0.035	snr	1.88	4.03%	0.32%	0.17%	1.50E-34	Bayesian optimization	downstream	Users with less than 36.27 dBm of snr in 12.74% of the measurements or more have 1.88x higher call rate
	1.23	0.047	snr_stdev	1.65	7.24%	0.28%	0.16%	1.73E-35	Bayesian optimization	downstream	Users with more than 1.23 dB of standard deviation in snr in a day have 1.65x higher call rate
0.10	15.58	0.008	cer	8.01	0.12%	1.90%	0.23%	1.65E-09	Bayesian optimization	ofdm	Users with more than 0.1% cer in 15.58% of the measurements or more have 8.01x higher call rate
	4.34	0.008	partial_service_flag	4.12	0.27%	0.98%	0.23%	6.37E-05	Genetic optimization	ofdm	Users with partial service in 4.34% of the measurements or more have 4.12x higher call rate
0.01	4.00	0.005	plc_cer	3.26	0.21%	0.77%	0.24%	1.63E-02	Bayesian optimization	ofdm	Users with more than 0.01% plc cer in 4.0% of the measurements or more have 3.26x higher call rate
36.68	4.00	0.013	mean_mer	2.37	0.93%	0.56%	0.23%	7.03E-04	Bayesian optimization	ofdm	Users with less than 36.68 dBm of mean mer in 4% of the measurements or more have 2.37x higher call rate
	1.04	0.018	mean_mer_stdev	2.27	1.40%	0.54%	0.23%	9.05E-05	Bayesian optimization	ofdm	Users with more than 1.04 dB of standard deviation in mean mer in a day have 2.27x higher call rate
1.65	6.27	0.020	high_profile	1.83	2.34%	0.43%	0.23%	8.02E-04	Genetic optimization	ofdm	Users with high profile lower than 2 in 6.27% of measurements in a day have 1.83x higher call rate

Table 2 shows the values for threshold 3 and the selected metrics. As expected, most critical issues like partial service (18.0%) or one carrier in upstream (13.6%) have lower values for threshold 3, meaning that with lower frequency of affected modems detected, the call rate of the node increases significantly. Other issues like CCER (correctable codeword error rate), need to be generalized to have a significant impact in the node call rate, thus have high threshold 3 values (43.2%).

Table 2 - Threshold 3 results

Metric flag	Threshold 3	Lift	Metric group	positive_share	negative_share	Interpretation
tx_upstream_node_imp_flg	38.3%	1.85	upstream	0.12%	99.9%	Nodes with 38.5% or more Daily Affected Modems for high Tx Upstream have 1.85x higher call rate
one_carrier_flag_upstream_node_imp_flg	13.6%	1.38	upstream	4.94%	95.1%	Nodes with 13.6% or more Daily Affected Modems for one Upstream carrier have 1.38x higher call rate
partial_service_flag_upstream_node_imp_flg	18.0%	1.30	upstream	7.66%	92.3%	Nodes with 18% or more Daily Affected Modems for partial service in Upstream have 1.3x higher call rate
t3_upstream_node_imp_flg	96.6%	1.28	upstream	0.04%	100.0%	Nodes with 96.6% or more Daily Affected Modems for t3 in Upstream have 1.28x higher call rate
snr_stdev_upstream_node_imp_flg	50.9%	1.28	upstream	0.32%	99.7%	Nodes with 50.9% or more Daily Affected Modems for standard deviation of SNR in Upstream have 1.28x higher call rate
snr_upstream_node_imp_flg	10.9%	1.26	upstream	3.52%	96.5%	Nodes with 10.9% or more Daily Affected Modems for SNR in Upstream have 1.26x higher call rate
rx_positive_upstream_node_imp_flg	65.7%	1.12	upstream	4.86%	95.1%	Nodes with 65.7% or more Daily Affected Modems for high Rx in Upstream have 1.12x higher call rate
tx_stdev_upstream_node_imp_flg	61.5%	1.09	upstream	0.99%	99.0%	Nodes with 61.5% or more Daily Affected Modems for standard deviation of SNR in Upstream have 1.12x higher call rate
cer_downstream_node_imp_flg	12.2%	2.74	downstream	0.40%	99.6%	Nodes with 12.2% or more Daily Affected Modems for cer in Downstream have 2.74x higher call rate
ccer_downstream_node_imp_flg	43.2%	2.17	downstream	0.08%	99.9%	Nodes with 43.2% or more Daily Affected Modems for ccer in Downstream have 2.17x higher call rate
rx_negative_downstream_node_imp_flg	16.7%	1.83	downstream	0.32%	99.7%	Nodes with 16.7% or more Daily Affected Modems for ccer in Downstream have 2.17x higher call rate
snr_downstream_node_imp_flg	26.7%	1.49	downstream	1.19%	98.8%	Nodes with 26.7% or more Daily Affected Modems for SNR in Upstream have 1.49x higher call rate
snr_stdev_downstream_node_imp_flg	11.8%	1.15	downstream	21.18%	78.8%	Nodes with 11.8% or more Daily Affected Modems for standard deviation of SNR in Upstream have 1.15x higher call rate
mean_mer_stdev_ofdm_node_imp_flg	10.2%	1.29	ofdm	4.15%	95.9%	Nodes with 10.2% or more Daily Affected Modems for low mean mer in OFDM have 1.29x higher call rate
high_profile_ofdm_node_imp_flg	10.3%	1.20	ofdm	5.81%	94.2%	Nodes with 10.3% or more Daily Affected Modems for low high profile OFDM have 1.20x higher call rate
mean_mer_ofdm_node_imp_flg	49.6%	1.04	ofdm	0.32%	99.7%	Nodes with 49.6% or more Daily Affected Modems for low mean mer in OFDM have 1.04x higher call rate
Any_node_imp_flg		**1.27**		**39.51%**	**60.5%**	**p-value: 4.24E-29**

7.2. Additional Testing in New Markets and Network Architectures

The thresholds previously described made a significant difference compared to traditional network assessment KPIs. They were used to optimize and complement other technical KPIs, such as QoE (Quality of Experience), which detects poorly performing nodes.

Although this represents a step change in understanding customer network experience from multiple dimensions, segmentation still relies on broad classification algorithms (explained in Section 6). However, sensitivity and metric symptoms may vary due to specific customer usage profiles or network architecture. For instance, areas with multiple layers of amplifiers or demographic profiles such as commercial hubs or remote residential areas exhibit different tech support engagement patterns.

Consequently, the next phase of the project will involve testing the resulting thresholds in different areas and new markets to determine how much additional condition-based adjustments are needed before reaching a point of diminishing returns.

7.3. Resulting HFC Node Categorization Over the Test Window:

The resulting thresholds and optimized segmentation from the training phase were tested in a different time window, this assures that there is no data leakage from the training to the test, and that the resulting segmentations do not overfit for events exclusive from the training window. Figure 11 shows how train and test windows are separated in time.

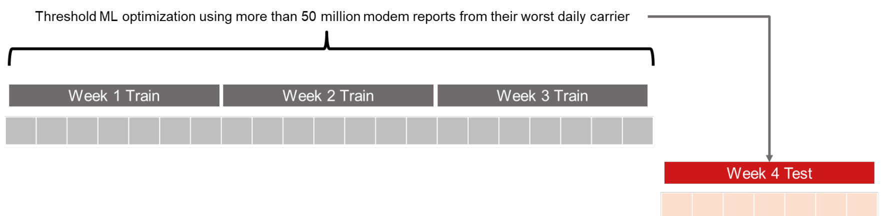


Figure 11 - Train and test windows

Figure 12Figure 12 – ML Optimized network segmentation output shows the call rates from nodes depending on the number of impairment flags in the test window. As expected, nodes with ML optimized impairment flags do show significantly higher call rates, and the call rate also increases if the node has more issues found by the algorithm, and the difference is statistically significant. In the final output from the segmentation. 8% of the nodes were flagged as "critical for intervention". Critical nodes concentrate 26.4% percent of the calls, thus, the return on investment from intervening those nodes is very significant.

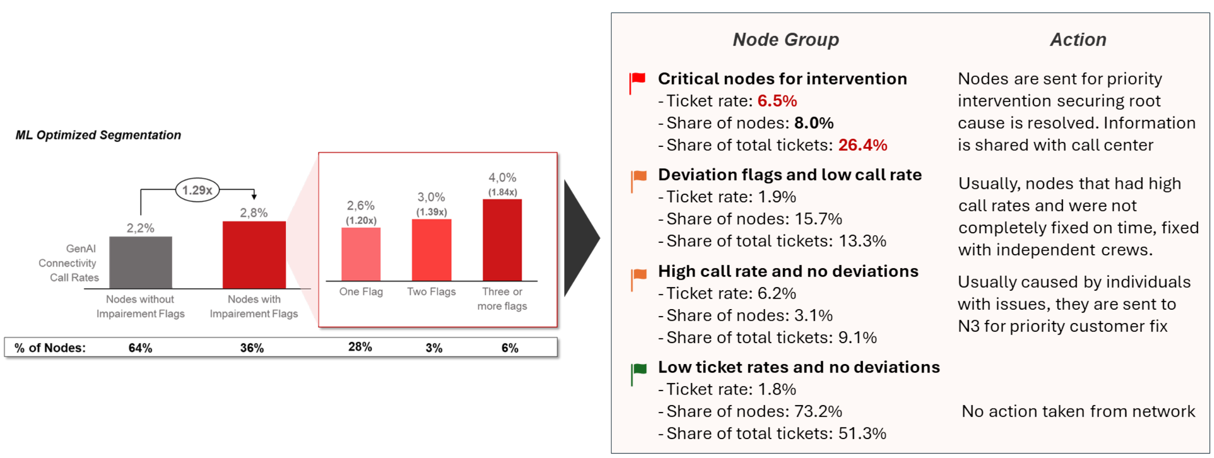


Figure 12 – ML Optimized network segmentation output

8. Impact of the Cx Tech and Lighthouse Framework in Technical Call Reduction

Two of LLA's operations have proactively implemented the described approach to segment, prioritize, and certify node interventions. This network segmentation was integrated with enhanced technical diagnosis capabilities in the call center, which enabled care agents to access new technical performance information and the flags developed, providing them with a clearer understanding of the issues at hand.

Further improvements included an upgraded technical intervention toolkit that guided field technicians in what specifically to look for, thanks to a broader and more specific set of network KPIs affecting network components. This comprehensive toolkit helps in pinpointing the exact areas needing attention, thereby enhancing the efficiency and effectiveness of field interventions.

Additionally, a better understanding of which calls were indeed technical support calls helped to find repeated callers at a network component level and modem resets that did not resolve the issue. This insight was critical in refining our approach to addressing persistent problems more strategically.

There was also a full network intervention process redesign aimed at reducing lead times and repeat interventions that yield no improvement in network KPIs. This redesign has been pivotal in minimizing unnecessary follow-ups and enhancing customer satisfaction by resolving issues more swiftly and effectively.

The combination of these enhancements has led to significant reductions in HFC call rates and truck roll rates, as shown in Figure 11 and Figure 12. In Country 1, the total tech ticket rate decreased by 21%, and the truck roll rate by 31%. In Country 2, which implemented the new intervention process later in 2024, call rates were reduced by 37%. Although the truck roll rate has not yet been reduced, it is expected to decrease as the program continues and fully integrates these enhancements.

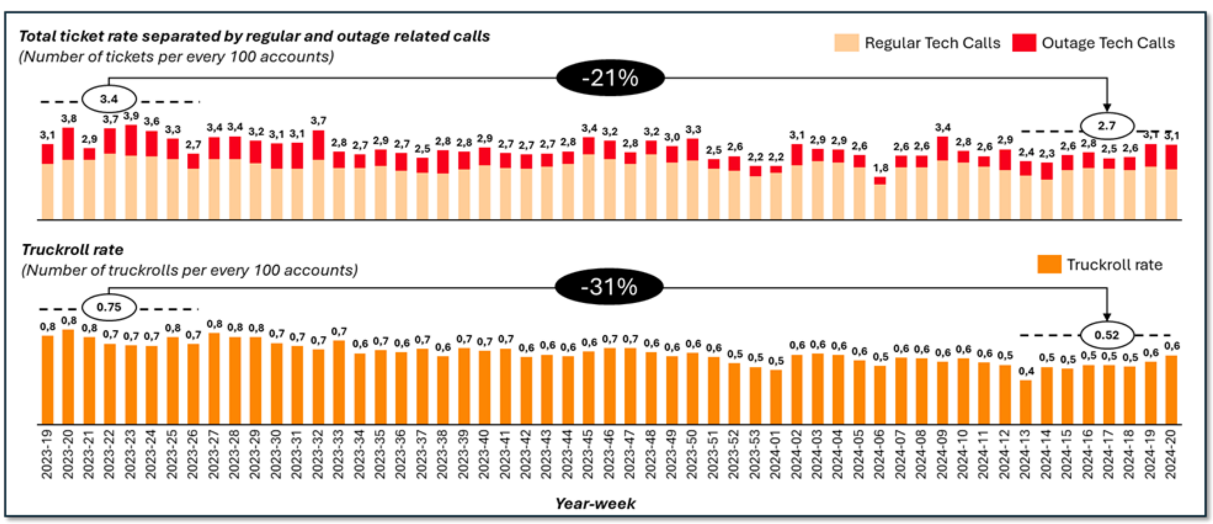


Figure 13 - Total impact of the enhanced network maintenance and technical care programs feed with project Lighthouse tools in Country 1.

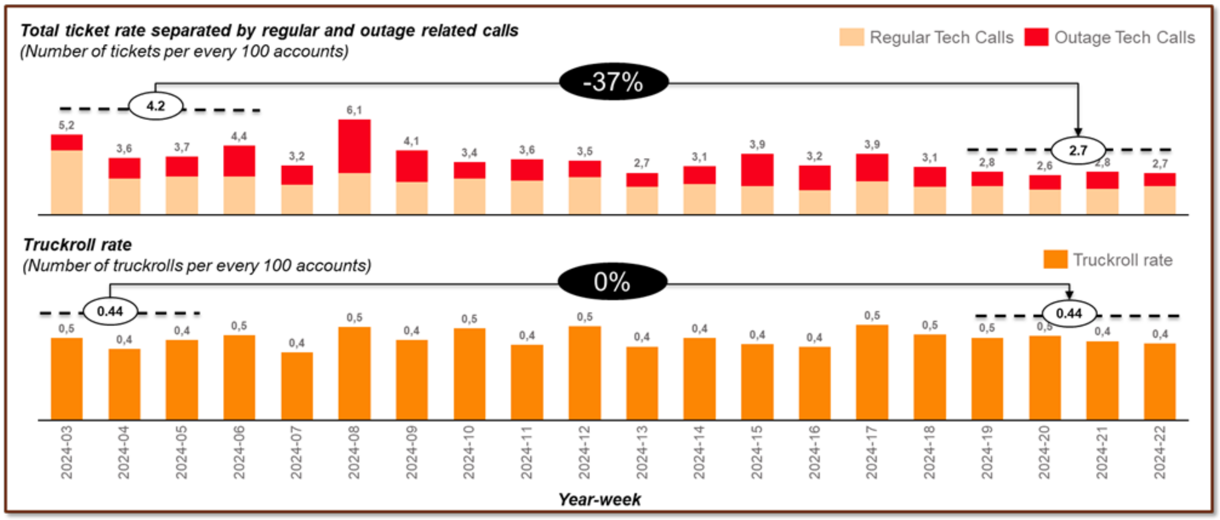


Figure 14 - Total impact of the enhanced network maintenance and technical care programs feed with project Lighthouse tools in Country 2.

9. Conclusion

Understanding the complex time series of events that led to a bad customer experience and customer dissatisfaction requires advanced data techniques and understanding of the technical data. Liberty Latin America Cx Tech program has proven that increased alignment between technical indicators, network monitoring with customer and business KPIs such as call rates or churn can be very beneficial for the business. The program has resulted in lower call rates in the markets where it was deployed, and better understanding of the events that are more impactful for customers.

New technologies like big data cloud processing platforms, machine learning and generative AI are enabling new analysis and use cases that were previously impossible or cost prohibitive. By setting up a centralized Advanced Analytics multidisciplinary team, LLA has been able to experiment with these new technologies and launch transformational use cases in a very agile way.

Abbreviations

LLA	Liberty Latin America
CER	codeword error rate
FEC	forward error correction
SNR	signal to noise ratio
Hz	hertz
Cx	customer experience
SCTE	Society of Cable Telecommunications Engineers

Bibliography & References

[1] G. G. T. Heiler, T. Haider and A. Hanbury, "Identifying the root cause of cable network problems with machine learning," *Arvix,* 2022.

[2] CommScope, Inc. , "ServAssure® NXT Performance Management," [Online]. Available: https://www.commscope.com/globalassets/digizuite/946800-servassure-nxt-performance-manager-pa-116621-en.pdf. [Accessed 5 July 2024].

[3] VIAVI Solutions Inc, "VIAVI XPERTrak," 2019. [Online]. Available: https://www.viavisolutions.com/en-us/literature/xpertrak-brochures-en.pdf. [Accessed 19 05 2024].

[4] CommScope, Inc, "Preparing for DOCSIS® 3.1," [Online]. Available: https://www.commscope.com/globalassets/digizuite/1878-wp-113924-en-preparing-for-docsis-3-1.pdf. [Accessed 18 May 2024].

[5] Cable Television Laboratories, Inc. 2010-2016, "DOCSIS® Best Practices and Guidelines: PNM Best Practices: HFC Networks (DOCSIS 3.0) CM-GL-PNMP-V03-160725," 2016. [Online]. Available: https://volpefirm.com/wp-content/uploads/2017/01/CM-GL-PNMP-V03-160725.pdf. [Accessed 19 05 2024].

[6] J. Bergstra, R. Bardenet, Y. Bengio and B. and Kégl, "Algorithms for Hyper-Parameter Optimization," *Advances in Neural Information Processing Systems,* p. 2546–2554, 2011.

[7] G. Louppe and M. Kumar, "Bayesian optimization with skopt," 2016. [Online]. Available: https://scikit-optimize.github.io/stable/auto_examples/bayesian-optimization.html. [Accessed 5 June 2024].

[8] Python Software Foundation, "geneticalgorithm 1.0.2," [Online]. Available: https://pypi.org/project/geneticalgorithm/. [Accessed 5 June 2024].

[9] The MathWorks, Inc., "What Is the Genetic Algorithm?," [Online]. Available: https://www.mathworks.com/help/gads/what-is-the-genetic-algorithm.html. [Accessed 5 June 2024].

