SCTE·ISBE CABLE-TEC EXPO 2019 | NEW ORLEANS, LA SEPT. 30-OCT. 3

2019 Fall Technical Forum
SCTE·ISBE • NCTA • CABLELABS

Customer First: CX-Driven Augmented Operations

A Technical Paper prepared for SCTE•ISBE by

Roger Brooks, Ph.D.
Chief Scientist
Guavus, a Thales company
2860 Junction Avenue
San Jose, CA 95134 USA
roger.brooks@guavus.com

Pankaj Kumar, Sr. Mgr. Analytics, Guavus

Mudit Jain, Mgr. Analytics, Guavus

Megha Vij, Data Scientist, Guavus

Nandit Jain, Data Scientist, Guavus

Andrew Colby, Field CTO, Guavus

Table of Contents

List of Figures

Introduction

While service plan price will attract customers, it is their quality of experience (QoE) which will determine whether they churn (Ovum, 2017). Subscribers hold operators responsible for everything, from the content provider to their home, that affects their experience. When asked "What characteristics are important for a high-quality broadband service?" the top two responses were

- "100% reliable broadband service"
- "Good customer service"

When we look at the first of these, survey results (Incognitio, 2016) indicate that subscribers base their decision to recommend their broadband service provider on three factors:

- Service speed: 45%
- WiFi reliability: 31%
- Pricing and service bundling options: 21%

We can imagine numerous use cases to affect these key quality indicators (KQIs) and to which machine learning might be needed. However, appropriate data is not always available to support the use cases as significant portions of those data must come from the CPE and network devices themselves.

In this paper, we report on two examples of how machine learning can leverage data generally available from CPE and network equipment to address technical customer experience use cases.

CPE versus Network Repair

1. Use Case

A typical cable operator schedules hundreds of thousands of truck rolls annually only to find out that the service impairments cannot be resolved by the technician (Field Tech) at the home. This discovery is made after the visit to the home and requires scheduling a second truck roll for a network maintenance technician (Line Tech). Apart from the wasted Opex of the truck roll to the home, this also negatively impacts NPS due to the frustration and delayed satisfaction of the customer. Conversely, network impairments have a potential to impact multiple customers, and delays in identifying these impacts further increase the likelihood of unnecessary truck rolls to customers impacted by the same network impairment and negative customer experiences. At $60+ per truck roll, solving this one problem equates to savings in the order of $10sM per year in addition to the impact on NPS and churn. Our goal is to make the prediction at least 24 hours before the Line Tech's truck roll would have been scheduled (e.g., when the Field Tech is at the home).

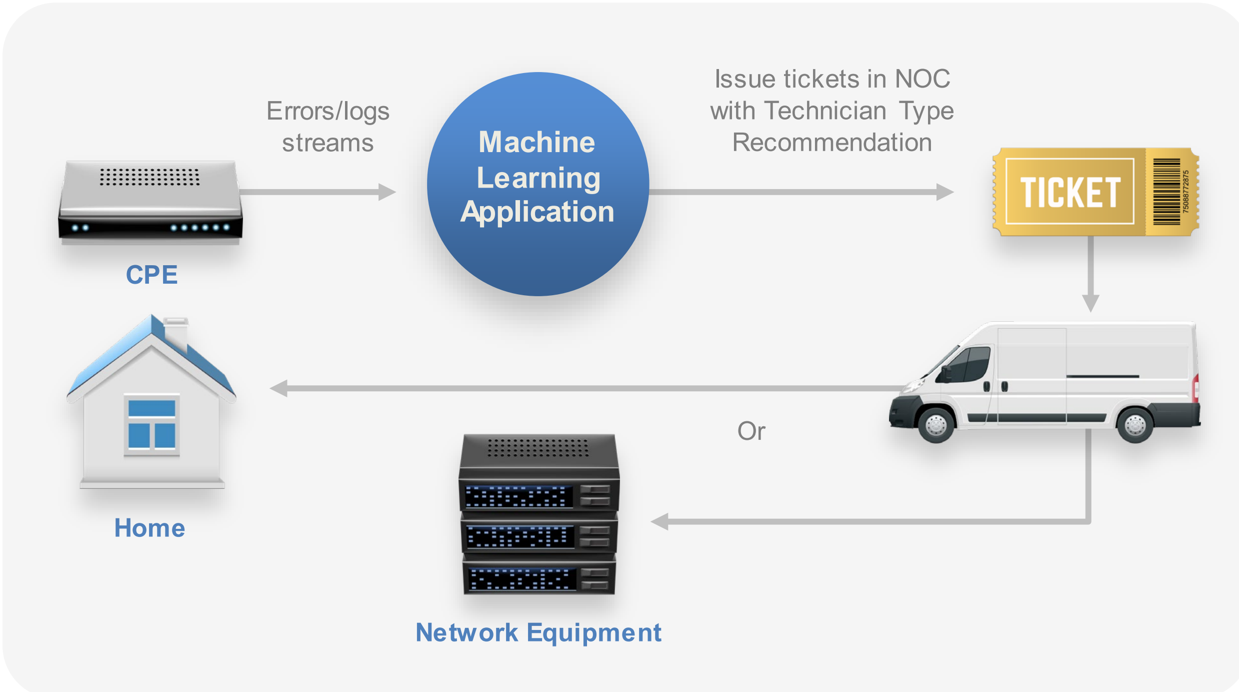

Figure 1 - Recommending that a subscriber's issue is best addressed at the customer's premises or in the network.

Present Mode of Operations (PMO)

1. Customer calls in seeking resolution of an Issue.
2. Troubleshooting results in a scheduled Field Tech Truck Roll to the home.
3. Field Tech arrives and determines that a Line Tech is needed to completely resolve the issue. (It is possible that BOTH a Field and Line Tech may be needed.)
4. Field Tech submits a request for a Line Tech to be scheduled.
5. A Line Tech is dispatched to diagnose and address the Issue in the HFC/Access network.
6. Attempts are made to verify that all services have been restored.
7. Other customers who are experiencing or will succumb to the same Issue either seek support resulting in additional Field and/or Line Tech Truck Rolls or remain silent with downgraded QoE.

Future Mode of Operations

1. --- Same as 1 of PMO ---
2. The troubleshooting process is informed by an AI prediction that the issue will require a Line Tech Truck Roll to the HFC/Access network.
3. --- Same as 5 of PMO ---
4. --- Same as 6 of PMO ---
5. Other customers who are experiencing or will succumb to the same Issue also find their services are back to normal without the need for additional truck rolls.

Operations Benefit: Unnecessary Field Tech Truck Rolls are preempted. Needed Line Tech Truck Rolls can be automatically scheduled. A decrease in support calls, from customers sharing the same network issues, is also expected.

Business Value: $10sM/year are saved in unnecessary Field Tech Truck rolls. Customers no longer experience unnecessary home visits and some issues are discovered and addressed without customers experiencing service impairments. This is expected to reduce churn and help stabilize NPS.

2. Data

In addition to data drawn from customer databases and network telemetry common to most MSOs, we leveraged telemetry from the CPE. Thematically, the data encompassed:

- Customer care event data
    - Calls – timestamp, account ID, service type(s), call type, etc.
    - Tickets – timestamp, account ID, problem type, prior resolution types, etc.
- CPE Telemetry
    - timestamp, CPE identifier, issue indicators, etc.
    - timestamp, CPE identifier, network connectivity events, etc.
- CPE Descriptor data
    - Service type, Device Type, Manufacturer, Model, MAC address, etc.

For the results reported, we utilized data from a MSO.

3. Machine Learning Approach

We created an analytics solution based on Supervised Machine Learning whose key modelling elements were an advanced clustering and classification on those clusters. To prove the concept, we constructed a training pipeline in which each subscriber's CPE data records were transformed into features reflecting the nature of the issues, the history (seven days prior to the prediction) and context of the device and customer. These features were broken into rolling windows of various time periods. The care event data was reflective of whether a Field Tech alone resolved the issue, or whether a Line Tech was needed, was used as the labels for the model.

4. Results

Separate from the data used to train the model, predictions were checked against recorded care events. It was found that roughly 90% of the subscriber accounts predicted to need a Line Tech, did in fact have the predicted truck roll after the interval (24 hours) on which the prediction was made. Despite unavailable (beyond those reflected in the data) factors leading to the scheduling of truck rolls, the 90% confidence predictions from the model captured 41% of Line Tech Truck Rolls for all subscribers who had at least one issue, within 24 - 168 hours, prior to the prediction; see Figure 2.

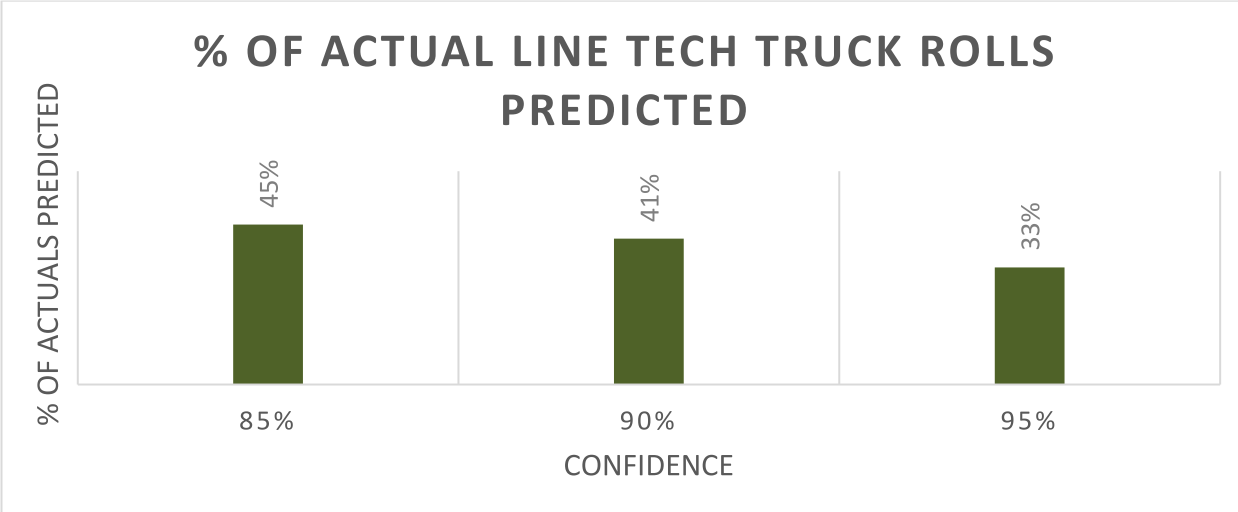

Figure 2 – The percentage of actual Line Tech Truck Rolls predicted by the ML model run at various levels of confidence.

5. Challenges Addressed

Data quality and completeness is always a challenge with these real-world projects and particularly so in operations analytics where one is essentially trying to interpret issues which would otherwise appear as statistical noise. The processes by which tickets are generated and updated further complicated the data preparation and analytics design as it led to ambiguous associations between the issues and those tickets. The machine learning had to be robust to these challenges; a task difficult even for human SMEs.

6. Related Proof of Concept

In a related proof-of-concept, we developed a similar ML model to predict which subscriber-cable modems were experiencing issues that would, if left unaddressed, result in a care event (call or truck roll). For the particular MSO customer, we used KPI telemetry data from the cable modems and found such predictions if acted upon would result in a seven-digit annual OPEX savings. The top categories of KPI attributes which drove those predictions are depicted in Figure 3.

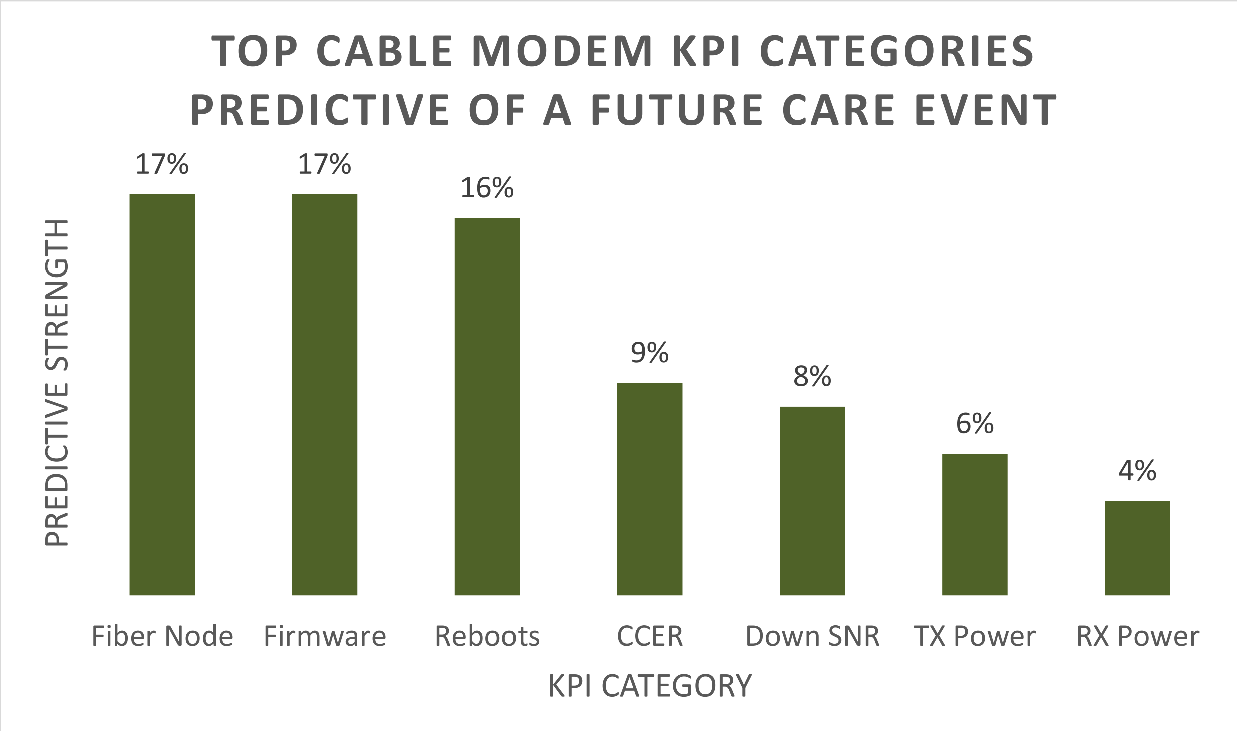

Figure 3 – The top KPI categories which are predictive of future cable modems care events.

Network Equipment Failure Prediciton

1. Use Case

We seek to predict if equipment critical to the operations of services may fail and why. We know how to apply machine learning to predict incidents from the alarms that are being issued by faulty devices; but what about non-alarm related failures? Logs are ubiquitous so in this use case, we predict service impacting equipment failures from syslog data. In particular, for tens of thousands of network node equipment, we sought failures associated with incidents which do not require on-site actions under the premise that those can be remotely and quickly resolved to take advantage of the advanced notice afforded by the predictions.

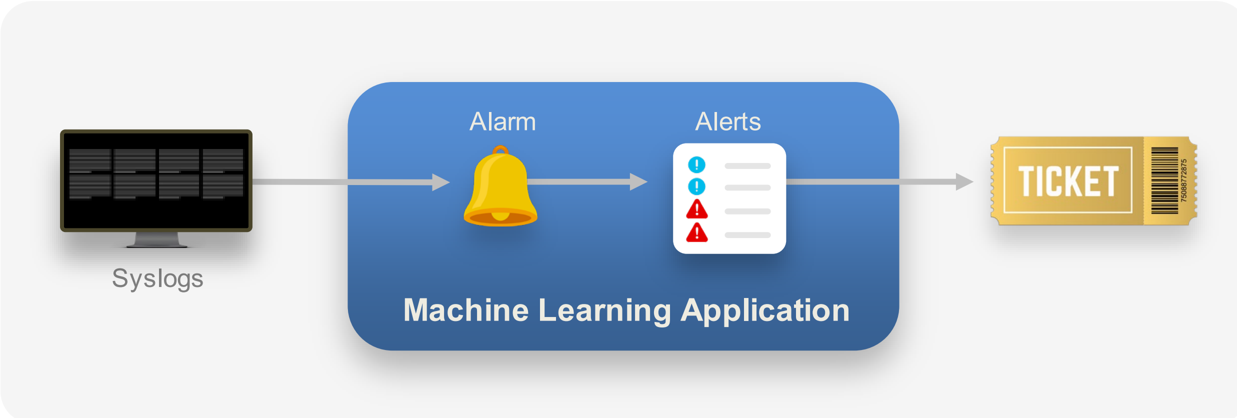

Figure 4 - Predicting equipment failures from syslog data.

Present Mode of Operations

1. Syslogs are generated by node equipment.
2. Syslogs are automatically collected, supplemented and made available for manual review by the operations team.
3. The Operations team associates ongoing incident tickets with the logs. Due to variations in time of review, interpretations of logs, frequent updates, etc. there is inconsistency in when and how these logs are associated with the corresponding incident tickets.
4. Technicians attempt intervention/non-intervention to resolve the issues. Tickets are closed at a varying numbers of days later.

Future Mode of Operations

1. --- Same as 1 of PMO ---
2. At configurable intervals, the Operations team receives a list of ‘At Risk’ equipment ids with risk score and risk drivers for the associated incident.
3. Staff takes remote action on ‘At Risk’ equipment ids to proactively solve issues before the incidents materialize.

Operations Benefit: Maintance action on ‘At Risk’ equipment can be proactively taken in advance of those equipment leading to an incident.

Business Value: Incidents which impair services and hence customers’ experiences are prevented. Such adverse contributions to churn and NPS are reduced.

2. Data

The data consisted of system generated equipment syslogs and the incidents created (manually) by the operations team. Thematically, the data encompassed:

- Syslogs – timestamp, equipment ID, log type, log entry (machine generated text)
- Incidents – timestamp, element ID, incident type, on-site flag (includes text fields)

It should be noted that not all incidents are reflected in the log data. Hence, we did not expect to be able to predict all incidents; indeed, we will see this in the results reported below.

For the results reported, we utilized data from a telecom operator.

3. Machine Learning Approach

Much as in the prior use case, we created an analytics solution based on Supervised Machine Learning whose key modelling element was an advanced classifier. To prove the concept, we constructed a training pipeline in which the syslog's fields were transformed into features reflecting the lexicals extracted from the text in the log entries along with their incident history. To aid in the system noise reduction, only logs which persisted for greater than a configured time period were used. Furthermore, spectral clustering was used for further reduction. Two binary labels were assigned to each training record; one based on whether an incident occurred in the subsequent 24-hour period and the other depicting whether an intervention was taken. Two different classifiers were trained, one for each label. The first classifier reflected classes of equipment organized by the probability of their members to fail. Predictions were made, at each hour, for each equipment logged during the period 15-75 minutes prior. The output was a daily prediction list for 'At Risk' equipment and whether the nature of the failure likely required an on-site visit or not.

4. Results

A daily list of equipment which are likely to result in an incident within the subsequent 3 days was predicted. On average it was found that approximately 4% of the equipment account for about 75% of the total incidents. In particular, we found that approximately 39% of the non-intervention incidents could be predicted at least 12 hours before the incident would have occurred. Furthermore, with just this limited information, the model was also able to predict, at 77% confidence, 31% of the incidents for which no on-site intervention was needed; see Figure 5.

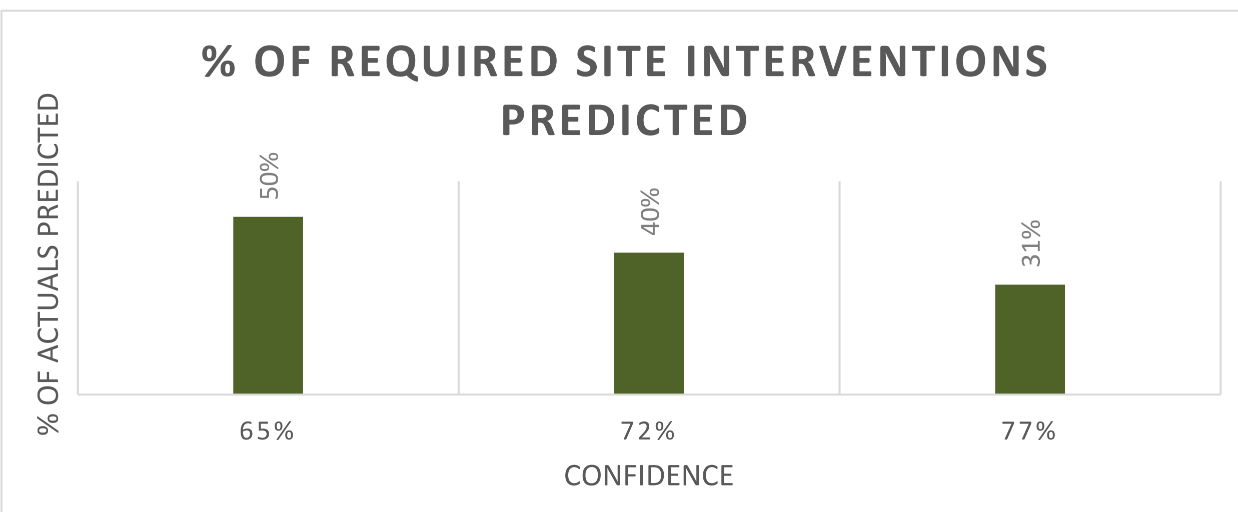

Figure 5 – The percentage of actual site interventions predicted by the ML model run at various levels of confidence.

5. Challenges Addressed

Familiar operations analytics challenges presented themselves in this project and were part of the reason we adopted the ML approach described above. Noteworthy amongst them were:

- Inconsistency in the manual incident ticketing process.
- Significant and varying time lags between the occurrences of the logs and the supposed incidents.

Conclusion

Studies have shown that the reduction of churn and increase in NPS are driven, in part, by improvements in QoE. We have described two equipment failure use cases which, in our work with customers, we have found to (a) impact customer's experiences and (b) were feasible given their data and operator processes. Given that no two operator's ecosystems are the same, mileage will vary.

Beyond the use cases discussed here, we have also found a number of analytics use cases with QoE impact and for which data is often available, including

- Prediction of incidents from alarms and thus prioritizing to which to attend.
- Facilitating faster service restoration through diagnostic steps such as discovering the root issue driving incidents.
- Discovering issues, such as alarms, which do not negatively influence QoE so that attention can be focused on the QoE impacting ones.

All of these use cases can be delivered via the application of machine learning driven by real-time alarms, KPIs and Syslog data streams from devices.

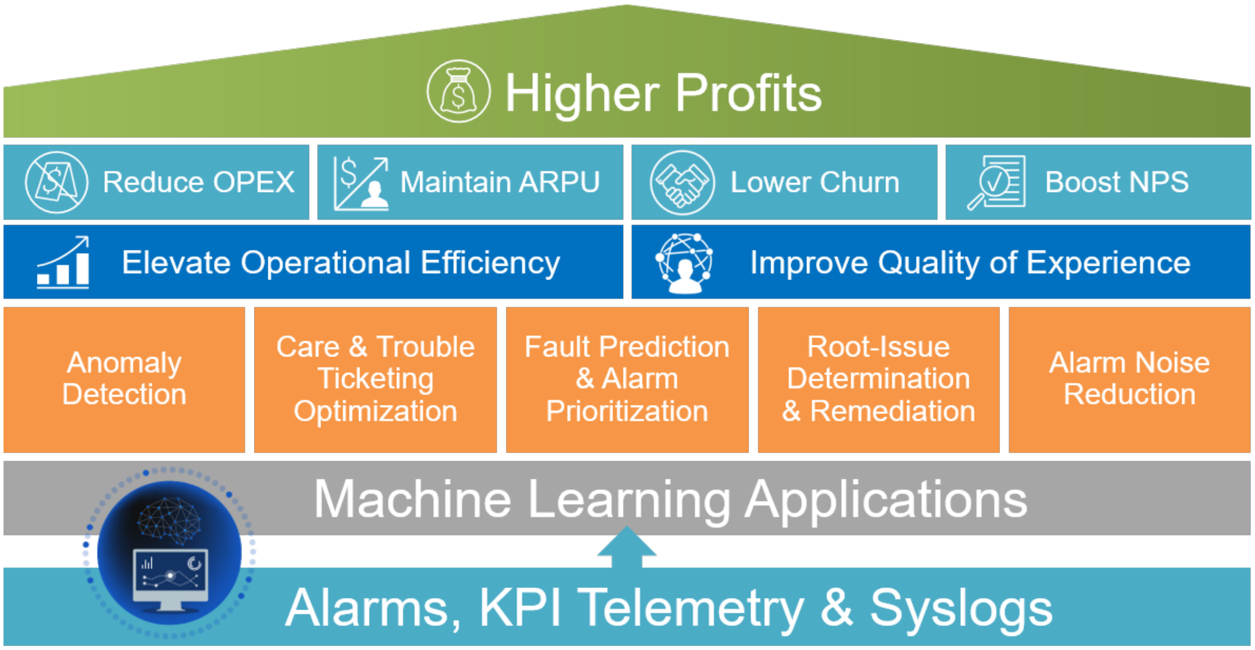

Figure 6 - Driving profits via machine learning on state data from network devices.

Abbreviations

AI	Artificial Intelligence
CPE	Customer Premises Equipment
HFC	Hybrid Fiber-Coax
KPI	Key Performance Indicator
KQI	Key Quality Indicator
MAC	Media Access Control
ML	Machine Learning
MSO	Multiple Systems Operator
NPS	Net Promoter Score
PMO	Present Mode of Operations
QoE	Quality of Experience

Bibliograph & References

Incognitio. (2016). *2016 BROADBAND CONSUMER QoE SURVEY REPORT.* Incognito Software Systems Inc. Retrieved from Incognitio.com.

Ovum. (2017, February). *Ensuring high quality broadband applications: a necessity for consumers and business critical for service providers.* Retrieved from https://ovum.informa.com/~/media/informa-shop-window/tmt/files/whitepapers/ensuring-high-quality-broadband-applications-feb-17-extended-version.pdf