Simplifying Field Operations Using Machine Learning

Applications of Machine Learning to Multiple System Operators (MSOs)

A Technical Paper prepared for SCTE/ISBE by

Sanjay Dorairaj
Senior Director, Comcast Innovation Labs
Comcast Corporation
1050 Enterprise Way, Sunnyvale, CA 94089
215-900-2349
Sanjay_Dorairaj@cable.comcast.com

Chris Bastian
CTO, SCTE
140 Philips Road, Exton, PA 19341-1318
cbastian@scte.org

Bernard Burg
Sr Manager, Data Science, Comcast Applied Ai
Comcast Corporation
1050 Enterprise Way, Sunnyvale, CA 94089
Phone: 408.900.8575
Email: Bernard_Burg@cable.comcast.co

Nicholas Pinckernell
Data Scientist
Comcast Corporation
1899 Wynkoop #550, Denver, CO 80202

Table of Contents

Title **Page Number**

Introduction 4
1. Maching Learning Overview 5
2. Applications of Machine Learning for MSOs 7
   2.1. Supervised versus Unsupervised Learning 9
   2.2. Training set, Test set and Hyperparameters 10
   2.3. Feature engineering 10
   2.4. Ensemble approaches 11
   2.5. Online versus offline algorithms 11
3. Operational Efficiency Improvements Using Machine Learning 11
   3.1. Machine Learning – A New Operations Paradigm 12
4. Typical Development Methodology 12
5. Case Study: Spectral Impairment Detection 14
6. Design Approaches 15
   6.1. Mathematical Modeling 15
      6.1.1. Feature Detection 17
      6.1.2. Example of impairment detection: Tilt and roll-off 18
      6.1.3. Results 19
   6.2. ML Models: Towards an Optimal Solution 20

Conclusion 21

Abbreviations 23

Bibliography & References 23

List of Figures

Title **Page Number**

Figure 1 - From the Stone Age to the Age of Big Data and Machine Learning 4

Figure 2 - Cambrian Explosion of Data (Source: Patrick Cheesman) 5

Figure 3 - Data Science Venn Diagram 7

Figure 4 - Data Requirements for Supervised and Unsupervised Algorithms 10

Figure 5 - Machine Learning Workstreams 13

Figure 6 - Machine Learning Process (Source: Crisp Industry Standard Process for Data Mining - CRISP DM) 14

Figure 7 - A few RF spectrum impairment samples 15

Figure 8 - Spectral Impairments and Their Shapes 16

Figure 9 - Spectrum, Channels and Features 17

Figure 10 - Feature Detection 18

Figure 11 - Tilt and Roll-off detection 18

Figure 12 - Suckout, Notch, Foreign Carrier and Resonant Peaking Classification 19

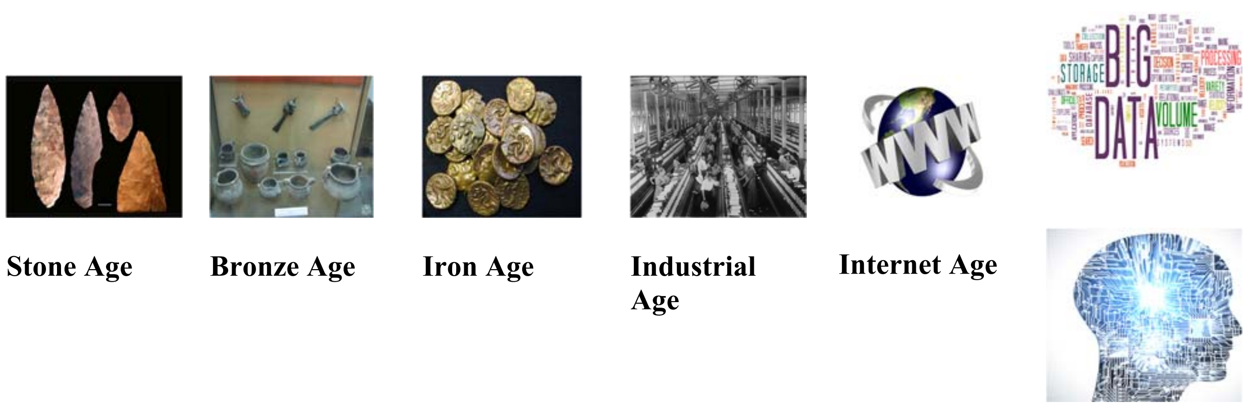
Figure 13 - Prediction Example – Wave, Tilt and Suckout 19
Figure 14 - Comcast Labeling Machine Solution Overview 20
Figure 15 - Comcast Labeling Machine – Labeling Impairments 21

List of Tables

Title	Page Number
Table 1 - Data Modeling vs Machine Learning approach	6
Table 2 - Machine Learning algorithms	8

Introduction

Every so often in the history of our evolution, humans discover something so important that it propels us into a new plane of technological and intellectual superiority. Over two million years ago, the **Stone Age** helped us build tools that established us as the dominant species on this planet. Much later, the **Bronze Age** (circa 3500 BC) and the **Iron Age** (circa 1200 BC) catapulted us to new levels of technological sophistication through the introduction of coin-based currencies, faster means of transport, durable manufacturing and construction and numerous other developments. This laid the foundation for the **Industrial Age** (circa 1700 AD), which ushered in the age of mechanized agriculture, mass transportation and electronic communication. The invention of the computer and the internet in the later parts of the 20th century heralded the dawn of the **Internet Age**. Individuals anywhere on the globe could now communicate and exchange information with one another. And much like Ray Kurzweil's Law of Accelerating Returns [1], the Internet Age is hardly over. Now, we find ourselves at the cusp of two back to back, tightly coupled events that are also bound to be of equally great historical significance - the **Age of Big Data** and the **Age of Machine Learning**.



Stone Age **Bronze Age** **Iron Age** **Industrial Age** **Internet Age**

Age of Big Data and Machine Learning/AI

Figure 1 - From the Stone Age to the Age of Big Data and Machine Learning

The explosion in data aka "Big Data", is a direct result of the exponential improvements in computing power and storage, with similar decreases in their cost [2]. This fueled an abundance of both personal and organizational data. The chart, below, provides a dramatic portrayal of the rapid growth of data over just one decade. Despite all of this data, the insights that we were able to generate has been limited by decades-old statistical and mathematical techniques and there wasn't much innovation in this field. The advent of Machine Learning has propelled us forward, by offering techniques that transform the big data into a veritable gold mine of valuable insights.

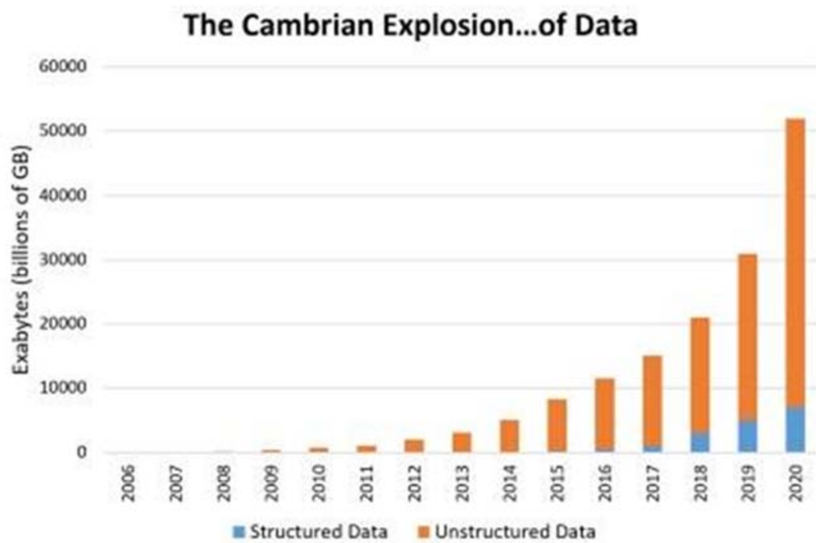


Figure 2 - Cambrian Explosion of Data (Source: Patrick Cheesman)

This paper is about machine learning - its definition and its applications. It especially examines the relevance of machine learning from the perspective of the cable's multiple system operators (MSOs). While there have been some attempts in technical and trade literature to pinpoint the benefits of machine learning to cable service operators, there has not yet been a holistic treatment of the subject, to our knowledge. This paper is an attempt to fill that gap.

1. Maching Learning Overview

Definitions of machine learning tend to compare it with traditional statistical methods. Leo Breiman, one of the pioneers and early evangelizers of machine learning, talked about the two cultures of statistical modeling - the data modeling culture and the algorithmic culture [3]. In the **data modeling approach**, which could be compared to traditional statistical approaches, the model assumes an underlying stochastic process. Inferences are made using techniques such as linear and logistic regression. Sample sizes are determined based on concepts founded in probability and inferential statistics and generally tend to be a tiny portion of the population size. **Machine learning** or the **algorithmic approach,** on the other hand, does not assume the existence of a well-defined process to the underlying data. Instead, it treats the model as a black box. Machine Learning algorithms such as neural networks and decision trees try to decipher the underlying patterns in the data using methods similar to that of maximum likelihood estimation. These algorithms typically require large amounts of data to yield good predictions.

Table 1 - Data Modeling vs Machine Learning approach

	Data Modeling Approach	Machine Learning Approach
Sample data requirements	Low	High
Constraints	Several	Few
Validation	Goodness of fit, residual examination	Performance on an independent test data set
Multiple variable prediction accuracy	Low	High
Data Interpretation characteristics	Linear or curvilinear patterns that can be approximated as functions	Complex non-linear patterns

Traditional inferential statistics has found its niche in several areas such as predicting an election outcome or predicting the effects of a new medication on a population. They perform well when the number of predictor variables are low. As the number of predictor variables increase, these models tend to break down. This is because of the large number of constraints these models are required to satisfy to yield valid predictions [4]. As the number and diversity of predictor variables increase, it becomes more and more difficult for these constraints to be met. On the other hand, machine learning algorithms are capable of dealing with complex processes and millions of predictor variables. The key requirement for machine learning to be successful is a data-rich environment, and the explosion of data in organizations today has proven to be instrumental in the increasing popularity and success of machine learning.

What role does machine learning play in Artificial Intelligence (AI)? AI is an overarching term that encapsulates all attempts to instrumentalize technology with the ability to think and act independently, much like humans do. It refers not only to the software and algorithms that renders this capability but the hardware and control systems as well. Machine Learning can be viewed as the subset of AI technologies that deals with pattern recognition.

A crucial advantage that humans have over existing computing platforms is our ability to make inferences from a complex set of input events. For example, our eyes are sophisticated enough to visually process information in three-dimensional space and recognize objects and emotions with little difficulty. Another example is our ability to look at a multi-variable time-series chart and immediately identify the anomalies present. The intelligence that enables us to excel at these tasks can be traced down to our uncanny ability to leverage our historical knowledge to perform real-time pattern matching. Machine learning and its derivative technology – Deep Learning, render computing platforms with pattern matching skills. In some cases, they are far superior to humans because they can process numerous parameters and complex underlying processes in an almost unbounded manner, limited only by computing and storage costs. In addition to the strong reliance on mathematics and statistics, machine learning is also strongly tied to software development, since the amount of data that it needs to be successful requires the use of state of the art software development methodologies.

The below diagram succinctly captures the overlapping areas of knowledge that data science comprises of - computer science, mathematics, statistics and domain expertise. The *unicorns* in the middle refer to those data scientists who possess the rare combination of all these skills.

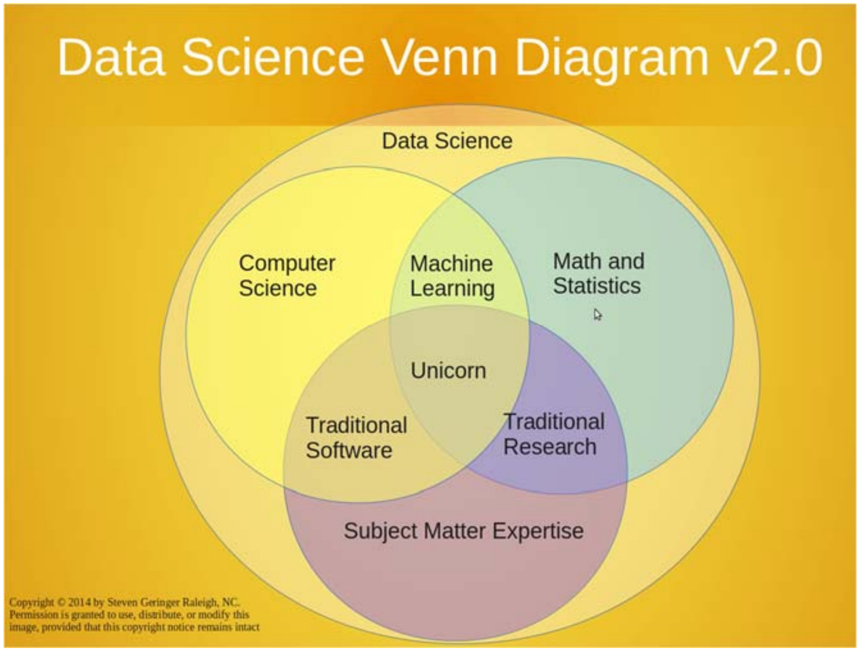


Figure 3 - Data Science Venn Diagram

2. Applications of Machine Learning for MSOs

In this section, we define general classes of machine learning algorithms and discuss how these classes of algorithms can add value to service providers.

The general classes of machine learning algorithms

1. Classifiers
2. Clustering Algorithms
3. Recommender Systems
4. Anomaly Detection Algorithms
5. Linear Regression

Classifiers are used to discern similarities among sets of data and assign them to categories based on their similarity. Examples of classification could be identifying objects in a video frame, identifying the underlying sentiment in a customer service message – happy, upset or neutral, or, associating a log message from a set-top to a specific error class. The technologies powering classifiers range from the simple - decision trees and random forest, to the very complex - deep neural networks. The choice of technologies used are typically functions of the level of complexity and the number of features in the underlying data. Image classification has been shown to benefit greatly by technologies derived from neutral networks such as convolutional neural networks (CNNs).

Clustering algorithms group similar data into clusters. They are typically used to group data that share similar characteristics or to look for significant deviations in data. For example, clustering algorithms could be used to look at smart home data and create user profiles based on shared behavioral characteristics – for example, early risers, late risers and so on. Clustering algorithms range from the simple such as K-Means clustering to more advanced algorithms such as agglomerative hierarchical clustering that may require additional tuning for optimal performance.

Recommender systems are a class of algorithms that make user or product recommendations based on historical usage or behavioral data. They can be used to suggest movies to users based on what those users have watched in the past or based on what users with similar viewing habits may have watched. For example, if two users like *Star Wars* and one of the users has watched *Dark Matter*, another sci-fi series, then the recommender would suggest Dark Matter to the other user. In a similar way, they can also be used to recommend products that users would like to purchase. In the case of customer service, they can recommend actions that the customer service representative can take in each situation based on past actions. A popular method of building these recommendations is using an algorithm called Collaborative Filters.

Anomaly Detection algorithms are similar to classification algorithms except that they typically only deal with cases where there just two classes of data exist and where one class occurs with an extremely low frequency. If the anomalies are relatively large, then clustering algorithms can be used; however, if anomalies are very few, joint probabilistic methods to model the rare events are more appropriate. Anomaly detection can be used to look for events such as billing fraud and device errors in cases where device failure is rare.

Linear Regression algorithms are used to make predictions about continuous variables. An example could be predicting customer churn rate or predicting bandwidth utilization. Linear Regression and Classifier algorithms share similar characteristics with respect to the technologies that are used. Where they differ is while classifier algorithms are designed to maximize the separation between dissimilar data points to allow for classes to be determined, linear regression algorithms interpret results in a continuous manner. One other point to note is that ML-based linear regression models are typically interpolative, traditional statistical linear regressions models are both interpolative and extrapolative. This only points to usage and does not imply that the traditional model is superior to the ML-model in cases where extrapolation is required.

Table 2 summarizes the above discussion.

Table 2 - Machine Learning algorithms

Class of Algorithms	Description	Technology Examples	Applications
Classifiers	Assigns data to categories based on similarity to other data.	Random Forest and Neural Networks	Sentiment Analysis, Image Classification
Clustering Algorithms	Groups similar data into clusters	K-Means, Hierarchical Clustering	User profiles and anomaly detection
Recommender Systems	Make recommendations based on historical data	Collaborative Filtering	Product recommendations
Anomaly Detection	Identify rare events	Joint Probabilistic modeling	Billing fraud detection
Linear Regression	Predict values for continuous variables	Linear Regression	Churn rate prediction

Discussed below are a few machine learning concepts and ideas that are also important to the successful application of machine learning.

2.1. Supervised versus Unsupervised Learning

As mentioned earlier, machine learning algorithms seek to find underlying patterns in data and mathematical ways of representing those patterns. The mathematical representation is referred to as a **model**. This search for patterns leads to two broad classes of machine learning – **supervised** and **unsupervised** learning. Given any set of data, if a machine learning algorithm is asked to determine underlying patterns in an autonomous manner, then that form of machine learning is known as unsupervised learning. Examples of unsupervised learning are (1) building consumer behavior profiles from customer call data and (2) classifying defect data into groups based on similarity between defects. Supervised learning is guided learning. In this case, the data also includes a parameter known as a **label** that captures the system response to a given set of input parameters. In determining customer churn for example, the available data will include the number of issues seen by a customer on a set-top on any given day. These are the predictor variables. In addition to the predictor variables, supervised algorithms require a label that could indicate whether the customer tried to cancel service that day. Supervised algorithms can be used to build a model that can predict the probability of a customer cancellation from the predictor and response variables. These algorithms have been shown to be effective in improving customer diagnostics, optimizing call centers, increasing the efficiency of truck rolls, and pro-active network healing.

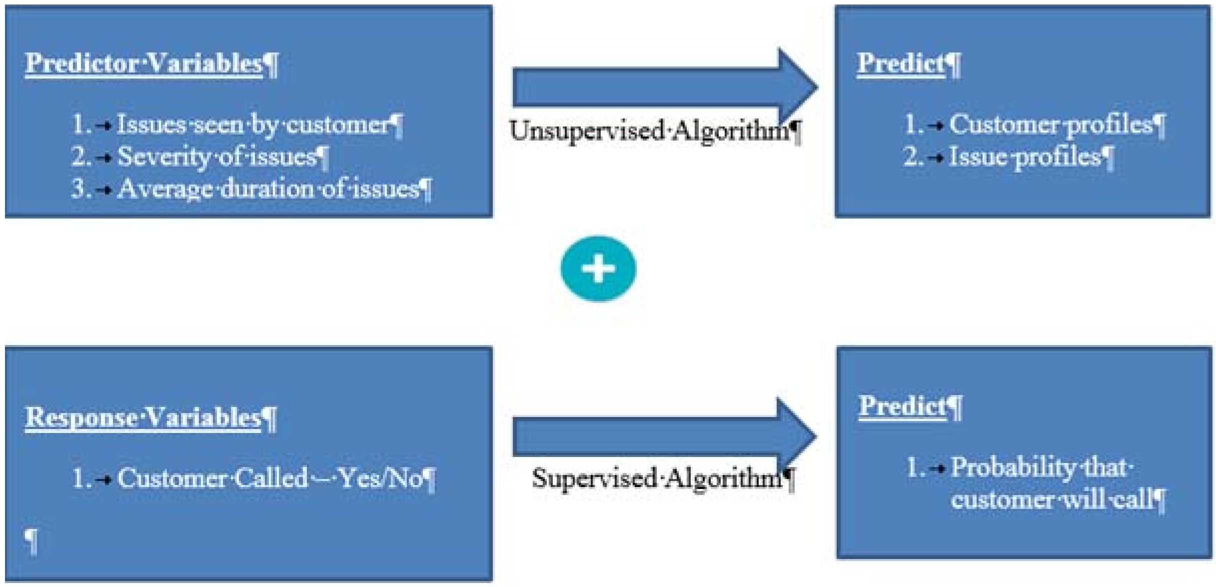


Figure 4 - Data Requirements for Supervised and Unsupervised Algorithms

2.2. Training set, Test set and Hyperparameters

Data used to build machine learning models is typically broken down into subsets - the **training data** and the **test data**. Training data is used to train the algorithm and allow it to build a model for the underlying data. Typically, the algorithm contains a number of tunable parameters, called **hyperparameters,** that are used to optimize the performance of the model. For example, when trying to use a clustering algorithm to build customer profiles, one of the hyperparameters is the number of clusters. In our examination of the resulting clusters from such an algorithm, we may notice that a certain cluster count yields a more optimal set of clusters than another cluster count. In a similar way, other hyperparameters can also be tuned till an optimal model is obtained. A key success factor for a machine learning model is to ensure that the training set and the test set are kept completely separate. This ensures the absence of any kind of bias during model generation. For this reason, hyperparameter tuning is not done using the test set, but rather, the training data is subdivided into a training set and a cross-validation set, and the cross-validation set is used to validate hyperparameters.

2.3. Feature engineering

There are two types of parameters that come into play with machine learning. The first type is referred to as a predictor variable and the second type is referred to as a response variable. Predictor variables are variables that are used to make predictions and response variables are the prediction. In image recognition, for example, pixels in an image are the predictor variables and the predicted class (cat, dog, flower etcetera) is the response variable. Similarly, when predicting the likelihood of a customer call, predictor variables could include the state of the set-top box modem and state of the infrastructure. In this case, whether the customer called, given the set of predictor variables, would be the response variable.

The selection of predictor variables is a crucial part of machine learning since the quality of the predictor variables ultimately determines the quality of the prediction. Predictor variables are also referred to as **features.** Feature selection is in itself a complex process and it has spawned a whole separate branch of machine learning called **feature engineering.** Feature engineering usually involves two types of activities (1) Reducing the set of all possible features into a set of features suitable that are better predictors of the output class and (2) Transforming or extending the set of available features with new features that are more suitable for the particular machine learning task. A popular method to transform one set of parameters into a smaller set of better predictor variables is called Principal Components Analysis (PCA).

2.4. Ensemble approaches

Often, when doing machine learning, the algorithms taken separately do not yield the best results. However, when combined with other machine learning algorithms or even multiple instances of the same algorithm, the quality of the results tends to improve, this is referred to as the **ensemble approach** to machine learning and this method is quickly gaining popularity in the machine learning community. The random forest algorithm is such as example. Sometimes, results from different algorithms such as random forest and support vector model (SVM), may be combined to yield a better classifier. Software tools include features that offer the programmatic selection of the best ensemble models through trial and error. A successful demonstration of the ensemble approach is the Netflix Prize which went to a team of machine learning engineers that developed the best algorithm using a similar ensemble approach [5].

2.5. Online versus offline algorithms

In certain cases, machine learning models may need to be built in real-time or **online** mode. For example, recommender systems need to process incoming events in real-time and provide recommendations based on the current state of the system. In this case, the model will need to be updated in real-time to ensure that the recommendations are up to date. In cases such as anomaly detection however, it may not be necessary to build a real-time model and an **offline** model is sufficient. In this case, models are built when data is available and refreshed with lesser frequency, perhaps on the order of weeks or months.

Depending on the type of application, an online or an offline model may be required. Not all machine learning algorithms work in an online model, so therefore, if choosing the online learning route, it is important than an algorithm that supports online learning is selected.

3. Operational Efficiency Improvements Using Machine Learning

As discussed above, there are several applications to machine learning. Some of the applications such as recommender systems, campaign management systems, market analysis and so on are revenue generating. Other applications have to do with cost optimization. These include customer call prediction, churn prediction, fault prediction, capacity planning and so on. In this section, we focus on the potential for machine learning to improve the operational efficiency of an organization.

Listed below are a set of reasons establishing how machine learning can help with operational efficiency goals.

- Cable system operators have a lot of data sources (understatement!) with valuable information about the state of the system
- These data sources are currently used only for basic- to medium-level analytics tasks, such as relative frequency comparison, difference computations and advanced visualizations.
- Predictive analytics using machine learning can help flag customer service issues in advance, presenting operators an opportunity to fix them before they disaffect service

- Machine learning tools can also be used to perform root cause analysis to identify underlying issues and recommend remediation actions
- When ML insight is deployed in development and field tools, it helps drive down call volume and truck rolls, thereby decreasing operational costs related to these activities

3.1. Machine Learning – A New Operations Paradigm

Machine Learning is a new paradigm of operations. This is especially true for field technicians who stand to benefit the most from this tool. Field technicians are used to certainty. When a DOCSIS monitoring device is plugged into an outlet, the expectation is that the spectral signature that they see is exactly what is present. The same goes for other measures such as signal loss, signal-to-noise ratio, signal levels and so on. This is a deterministic paradigm where what is reported is exactly as it is.

Machine Learning solutions are different. They do not provide answers that are a 100 percent guaranteed to be true. What you get is an answer and a probability associated with that answer being true. For example, in the case of a spectral impairment, the machine learning solution may say that there is a 95 percent chance of the signal containing a wave impairment. How should the field technician or the network operations center react to a probabilistic result? There are known methods of handling uncertainty and these are all based on an application's aversion to false positives.

Evaluating performance of machine learning models involves balancing cost reduction, customer satisfaction and model complexity. A large volume of repair calls implies that **small improvements can yield sizeable cost saving.** Consider the below example

- 1 million repair calls a month at a hypothetical $10 per call implies a monthly cost of $10m per month.
- A **1 percent reduction** results in a 100-thousand-dollar monthly saving and an annual saving of approximately **1.2 million dollars**

Machine learning also provides a means for tuning the model to yield a desired false positive rate. Reducing the number of false positives would however drive down the number of true positives, so there is a tradeoff that must be made. The examples below show two use cases

- A destructive self-healing action such as a reboot would require higher precision; we should therefore minimize false positives to reduce disruption to the customer
- A non-intrusive self-healing action such as a billing change would allow for lower precision, as false positives influence the overall result in the same manner as false negatives.

Similarly, in the case of the spectral impairment detection case study considered in this paper, the machine learning algorithm would assign similar probabilities to each of impairment that it detects and the field operations and the network operations center should have a strategy to deal with this information in a meaningful manner.

4. Typical Development Methodology

The development and deployment of machine learning within an organization typically takes place as two parallel, though connected, workstreams. The **first workstream** is more centered on the modeling effort. The **second workstream** focuses on ensuring that there is a path to deployment for the models being developed. The two efforts are viewed as happening concurrently because of the complex nature of deploying a machine learning solution in a cable system operator's production environment.

Figure 5 shows the two workstreams and Figure 6 shows a high-level view of the machine learning model lifecycle.

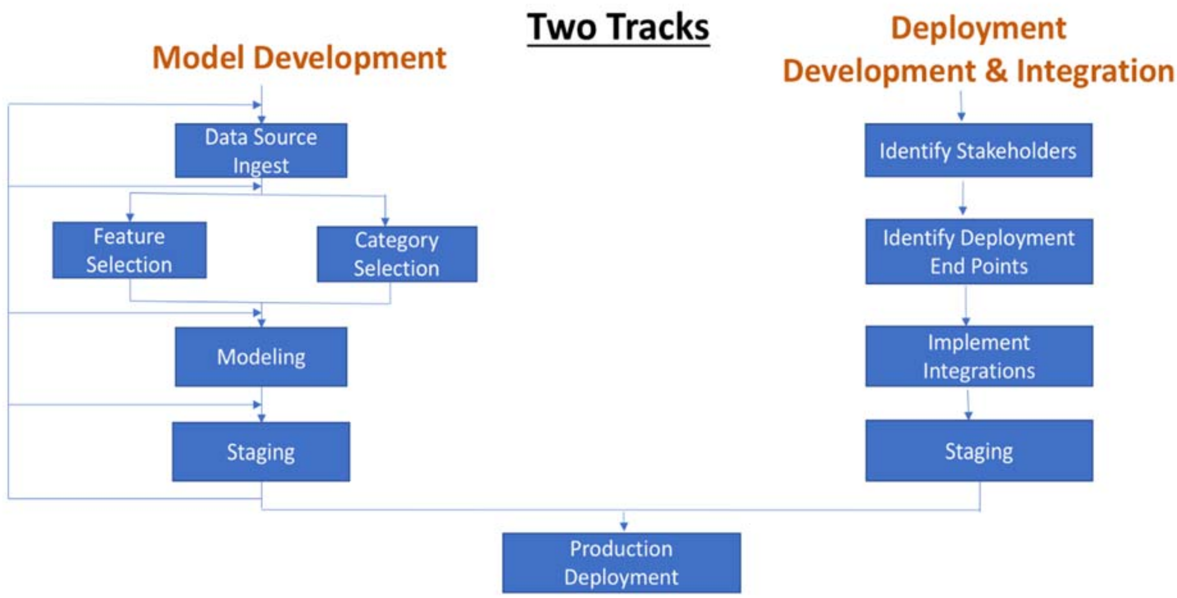


Figure 5 - Machine Learning Workstreams

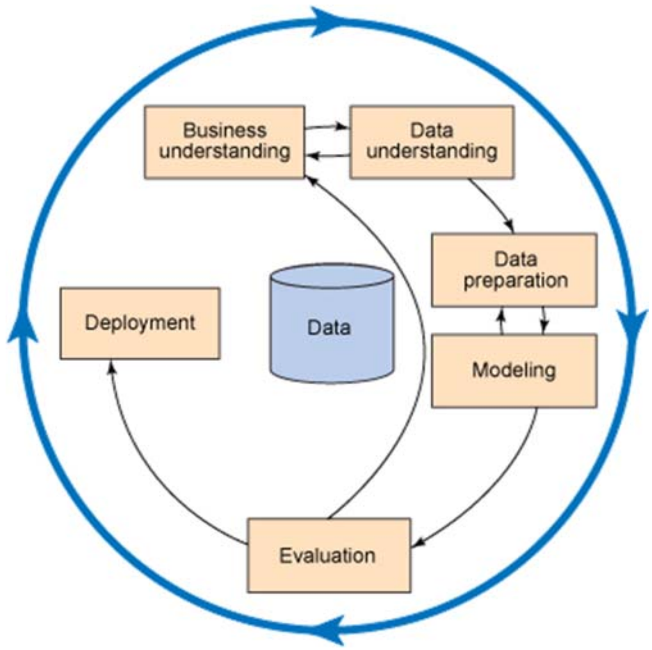


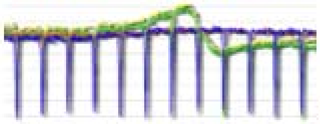
Figure 6 - Machine Learning Process (Source: Crisp Industry Standard Process for Data Mining -CRISP DM)

5. Case Study: Spectral Impairment Detection

Cable operators monitor the use of the spectrum for every device (e.g. cable modem). Such measurements give a state of the communication between the network infrastructure and the device.

The goal of this method is to automatically characterize these spectra by labeling all their impairments. This is instrumental to: 1) Assess the performance of the RF spectrum, 2) Consider variation over time and temperature; 3) Standardize automation & detection or anomalies, and 4) Remove subjectivity and manual interpretation by technicians.

Experts have identified 15 impairments for which automatic detection would bring a competitive advantage. Each of these impairments exhibits an identified cause, and is linked to a repair action that improves the performance of the RF spectrum. For example:


is a *resonance peaking* caused by an amplifier problem


is an *adjacency/alignment* caused by a head-end problem


is *a suck-out* caused by an in-home problem

Figure 7 - A few RF spectrum impairment samples

The 15 plant-related impairments ripe for detection and subsequent correction include: Suck-outs, Notches, Tilt (and direction), Ripples / Waves, Off-Air Ingress, Foreign carriers, Wideband / Edison, Roll-off, Resonance / Peaking, Filters, Leveling, Adjacency / Alignment, Power Summary, Distortion / Intermod, and Pilot-to-Channel ratio.

6. Design Approaches

The accuracy of the spectral impairment detector currently in production is low, with only 5 impairments being detected. The new impairment detection described is significantly more accurate, targeting the detection of 10 of the 15 known impairments.

To enhance the accuracy of spectral impairments interpretation two methods are being pursued. Each of them will result in a much higher impairment classification accuracy.

Mathematical modeling: Spectral data is modeled through traditional signal processing methods, extracting features characterizing each of the 15 impairments in a direct, static mapping.

ML models: An ML algorithm learns dynamic mappings between the features extracted by the mathematical model and the impairments. As such, ML uncovers optimal solutions, fine-tuning each feature to its best use within a context. This comes at the cost of labeling huge quantities of data to perform the supervised learning. Addressing this issue, the team built a labeling engine to crowd-source labeling within Comcast.

6.1. Mathematical Modeling

Each of the impairments described in the figures below are detected by a corresponding set of features.

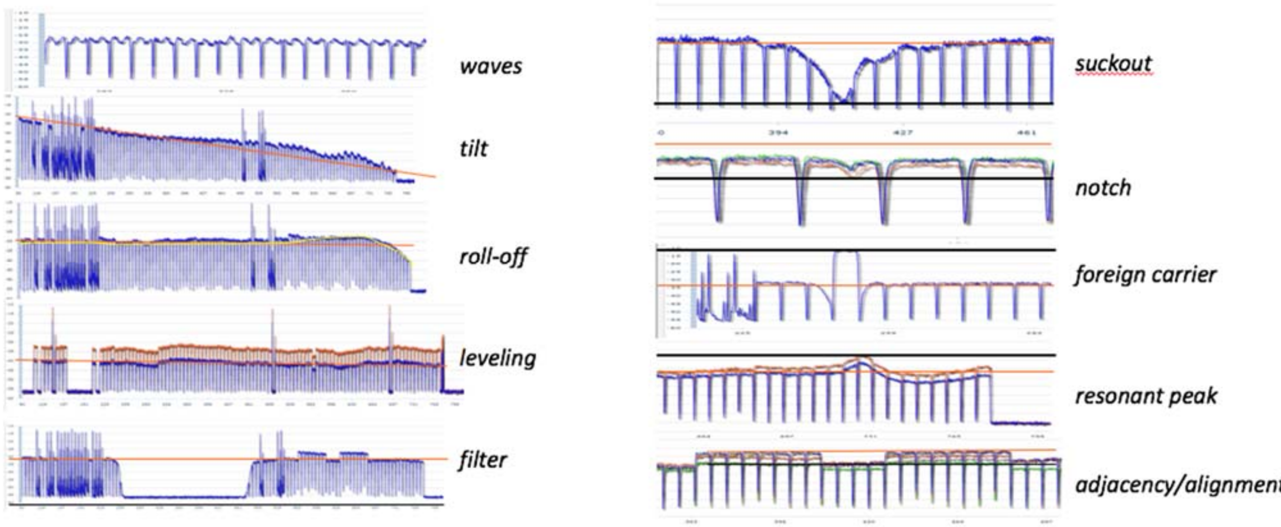


Figure 8 - Spectral Impairments and Their Shapes

Some of the impairments, like roll-off, filters or suck-outs are very impactful to end customers, even preventing them from accessing some channels. Other impairments, like waves, off-air-ingress or tilt slow transmissions down. All of these impairments are linked to known causes. Their diagnostic is key to the performance of Comcast's operations, and is of particular use to field technicians, because it allows them to pinpoint the cause of poor performance, or installation malfunctions.

The proposed approach is based on noise-resistant feature detection. Two data representations are used in parallel. The spectrum representation uses the complete spectrum, with a sampling at 117 kHz. The channel representation characterizes each TV channel which corresponds to a 6 MHz sampling. Channel representation is used and well understood by technicians.

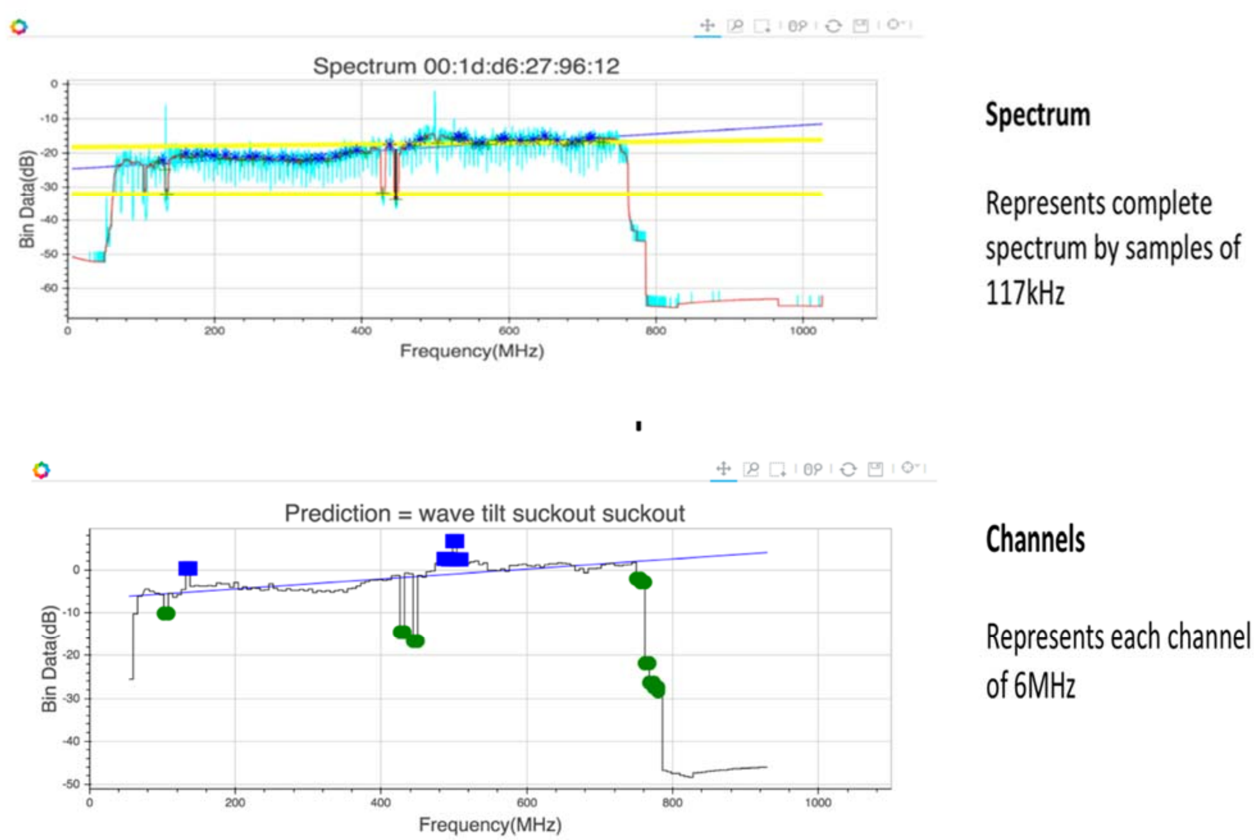


Figure 9 - Spectrum, Channels and Features

The features are independent from each other and are oblivious to the frequency at which they appear, and their combination allows detection of impairments. Impairment detection methods are also independent from each other, allowing their results to be combined. Thus, this overall detection method allows fine tuning both features and impairments, independently. This flexibility permits the introduction of new features, as well as new methods for impairment detection, without affecting existing detection methods.

6.1.1. Feature Detection

The program extracts similar features for both channel and spectral representations.

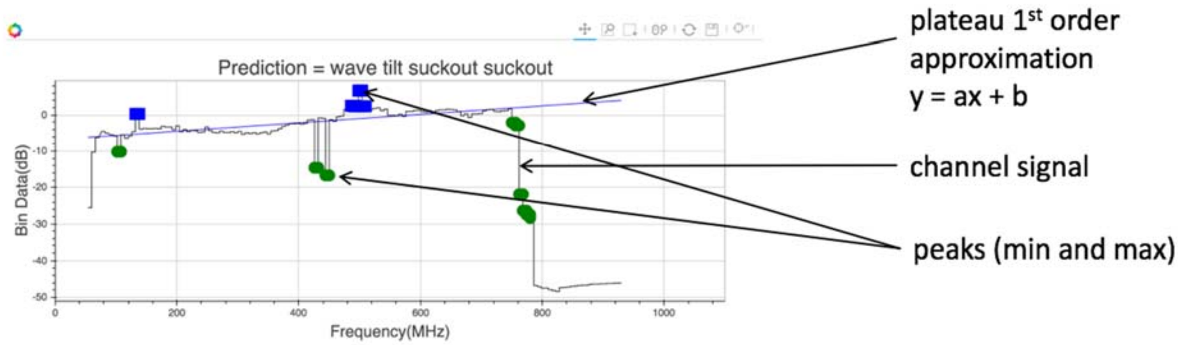


Figure 10 - Feature Detection

A plateau, with frequencies between 120MHz and 750MHz, is considered for the analysis. A linear approximation of this plateau offers stable features y = ax + b to assess the flatness of the plateau and its height. A similar approach is undertaken with higher degree approximations of the same plateau. From this plateau, positive and negative peaks are detected. The shape around these peaks is an important feature, as some of the peaks are formed by single channels, whereas others have parabolic shapes --illustrating that many channels are affected simultaneously.

6.1.2. Example of impairment detection: Tilt and roll-off

A tilt is well approximated by a linear signal. The roll-off, in contrast, shows a fast decrease of the signal amplitude at high frequencies, and is better modeled as a parabola.

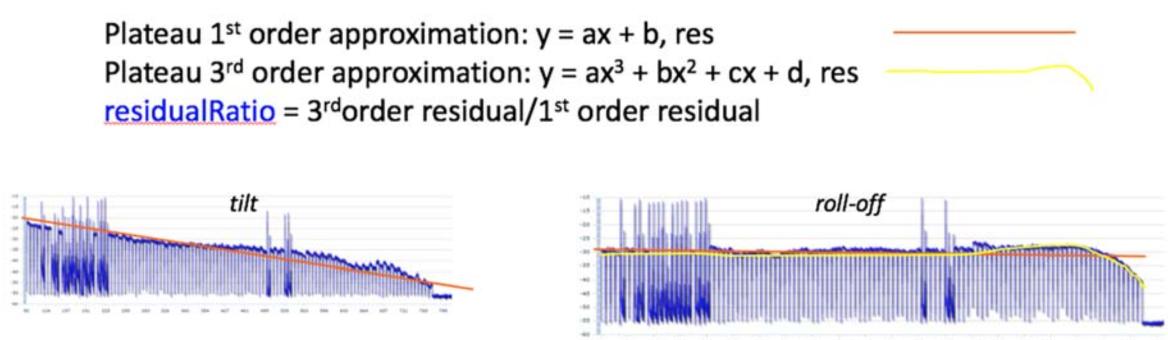


Figure 11 - Tilt and Roll-off detection

One solution differentiating a tilt from a roll-off is to make the ratio between the residuals of the 3rd order approximation and the 1st order approximation. If the ratio is near to 1, the impairment is a tilt since no fast decrease of the signal amplitude at high frequencies was detected.

Example of impairment detection: Suckout, notch, foreign carrier, resonant peak

From a feature extraction point of view, suckout and notch impairments are seen as signal dips, whereas foreign carrier and resonant peaks represent crests in the signal amplitude.

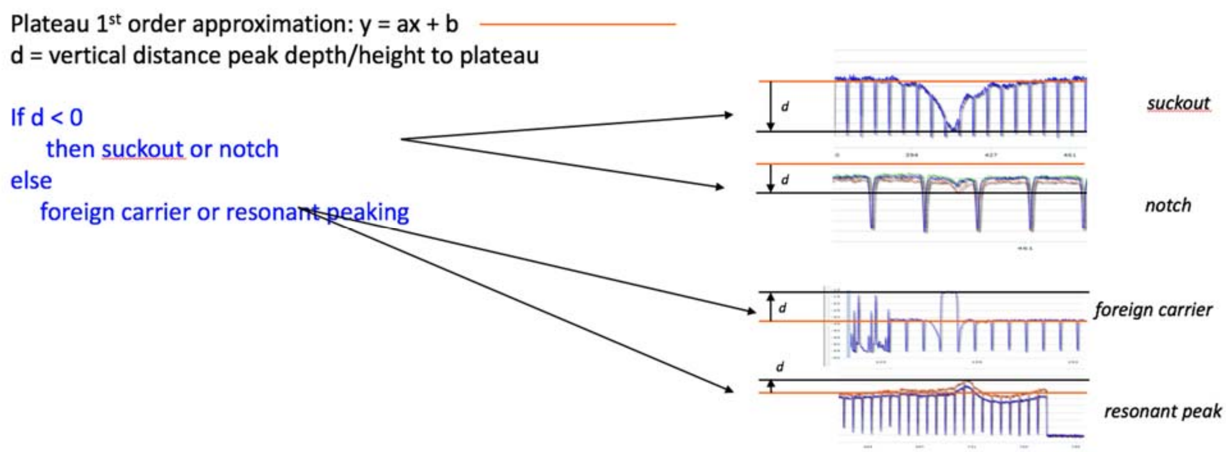


Figure 12 - Suckout, Notch, Foreign Carrier and Resonant Peaking Classification

The suckout is a large dip spanning several channels, whereas the notch is a tiny dip that cannot be seen in the channel view.

The foreign carrier is a sharp, single channel peak in the signal, whereas the resonant peak is a shallow peak spanning across several channels.

6.1.3. Results

The presented method returns a complete impairment diagnostic including all impairment instances detected on a spectral signal.

In Figure 13, the presented method discovers a combination of wave, tilt, suckout at 429MHz, and suck-out at 445MHz. In the figure, suckouts are annotated with a red cross-circle:

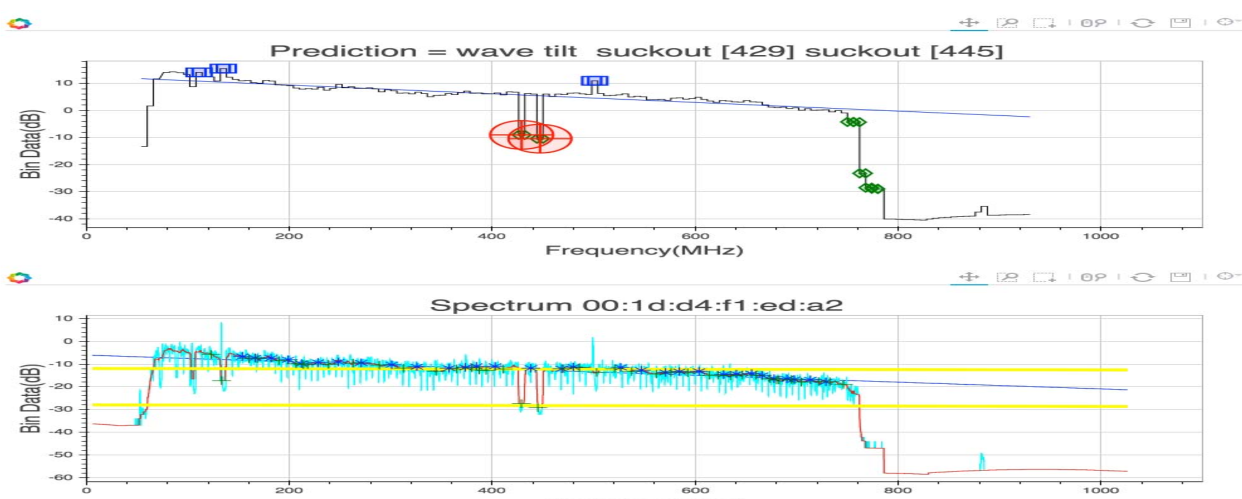


Figure 13 - Prediction Example – Wave, Tilt and Suckout

One of the advantages of this mathematical modeling is its simplicity: Each of the impairments is linked to a few features extracted via signal processing. The fine tuning of these features warrants some experimentation and skills, e.g. for defining the threshold making the difference between the tilt and the roll-off. These parameters are largely independent and can be fine-tuned independently. But static tuning might not be the optimal solution.

6.2. ML Models: Towards an Optimal Solution

ML models can be used to bring mathematical models into a new dimension. Instead of having a finely tuned mathematical model working with static parameters -- like the threshold making the difference between the tilt and the roll-off -- imagine having an ML algorithm that *dynamically* fine tunes these parameters, according to the expected output. The great advantage of ML is that it uses algorithms such as linear regression and classifications to determine the best parameter settings. ML is capable of optimizing thousands of parameters that are far beyond the capabilities of what humans can fine-tune.

However, ML comes at a huge cost in this setting. ML works best in supervised learning, so, data needs be labeled. The features that were treated independently in the mathematical model are now mixed together, leading to a combinatorial explosion. Labeling 15 features into 6 buckets (e.g. none, tiny, small, medium, large, huge) leads to 15^6 possibilities = 11 million to hit each bucket at least once. This domain is most probably sparsely populated; however, this simple calculation shows that labeling data is a daunting task, way beyond human capabilities. At a first glance, hundreds of thousands of labeled data could and should be generated.

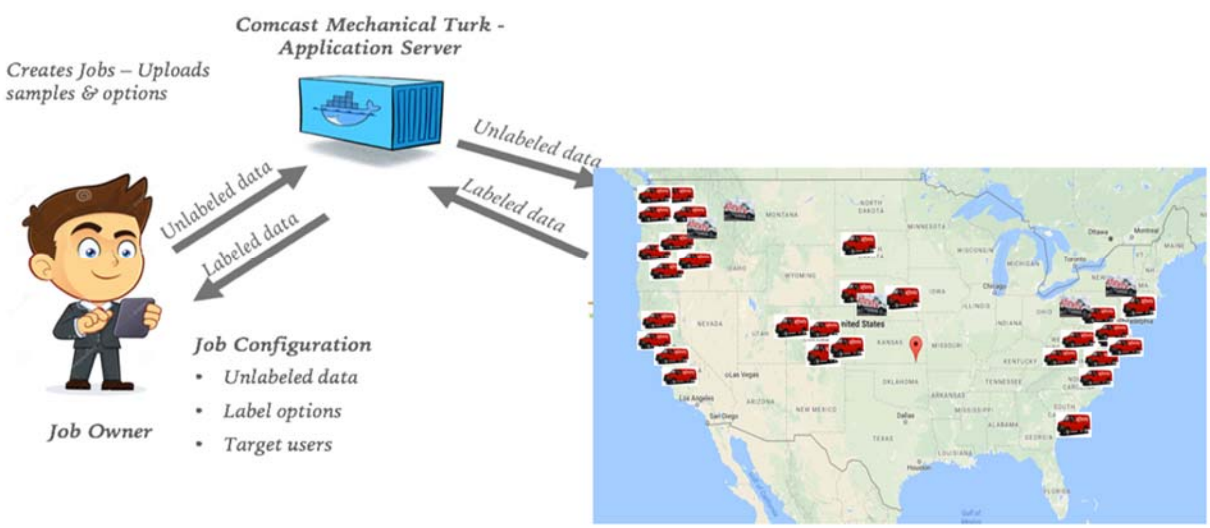


Figure 14 - Comcast Labeling Machine Solution Overview

The good news is, Comcast and other, like-minded MSOs have thousands of experts in the field capable of labeling this data. These are our industry's technicians. The idea is to farm the labeling task out through a "Turk mechanism" [6]. In this case, the unlabeled data is provided to the Comcast Mechanical Turk server that crowd-sources the data to be labeled to the technicians. At any time, an underutilized technician can access data and label it, through an API accessed by an app on their device.

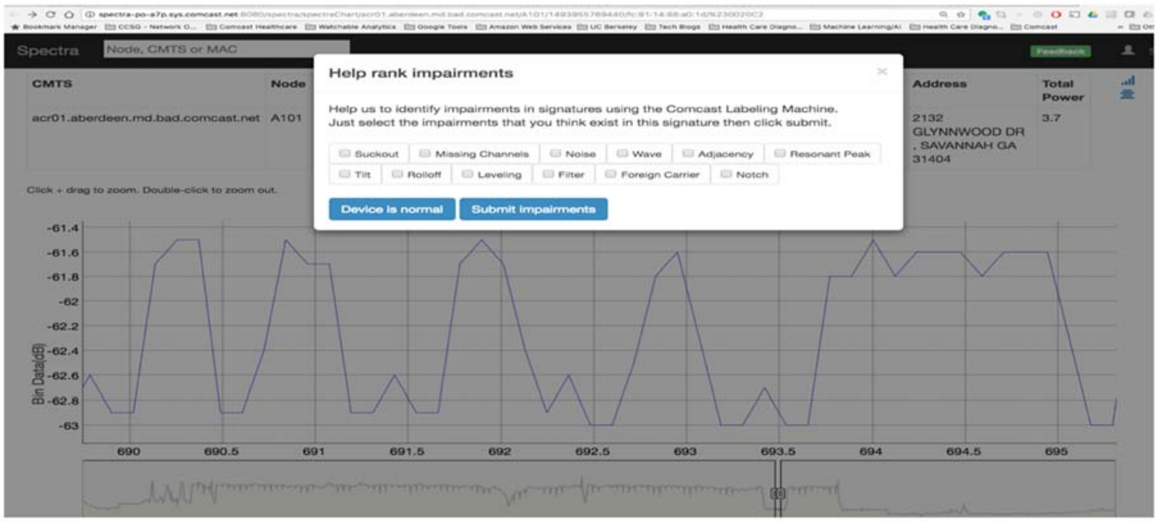


Figure 15 - Comcast Labeling Machine – Labeling Impairments

The graphic interface shown above allows a quick annotation, prior to sending the data back to the labeling Turk. The same data can be labeled independently by several technicians to improve the label quality through a vote or an averaging process. The collected data can be reviewed at the labeling engine prior to sending it to the ML algorithm.

Conclusion

Comcast and similar service provider companies have a lot to gain through the applications of machine learning, especially in the area of improving operational efficiencies.

A key takeaway from this paper are the concepts of machine learning as they relate to multiple service operators. Especially important is the new paradigm under which machine learning operates – that of probabilistic expectations and the move away from determinism.

"The best is the enemy of the good".

The search for answers that are a hundred percent guaranteed to be true can stifle our ability to be successful because it makes us resistant to innovative approaches to problem solving, such as machine learning. Machine learning provides us not just an answer but also the probability associated with the outcome. We, as cable operators, should begin to appreciate the value of such results and have processes that can educate folks on how to use this information. Only then, can we reap the true benefits of machine learning.

This paper also looked at how spectral impairments in the RF spectrum can be predicted using two approaches, one based on straight-forward mathematical modeling and another based on machine learning. Mathematical modeling is similar to a rule-based approach where patterns in the RF spectrum are predicted based on how well they fit certain mathematical functions. The mathematical functions are built using one or more observations. The main drawback of the mathematical model is its inability to scale to accommodate a larger set of representative spectral impairment patterns. Machine learning trains numerous sets of labeled spectral impairment observations and uses this method to build a model for spectral impairment detection. It can also leverage the feature selection work done using the mathematical model.

Given the vast amount of training data that the machine learning model has seen, it is able to better discern subtle differences in spectral waveforms and consequently leads to better predictions. In addition, it is much more maintainable than the mathematical modeling approach since learning to identify a new spectral impairment is simply a matter of adding the new spectral impairment data to the training set and rebuilding the model. The presence of labeled data or an easy method to label the data would further simplify the machine learning approach. The mathematical modeling approach on the other hand is harder to maintain because it requires an expert to generate new functions to recognize a new impairment.

Both the mathematical model and the machine learning model can nicely coexist, with the predictions that they each make serving to contribute to reinforce the overall prediction accuracy.

Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Network – A type of neutral net that has been very successful in image classification
DOCSIS	Data Over Cable System Interface Specification – The protocol used to carry data traffic over cable infrastructure
ML	machine learning
MSO	Multiple Service Operators – typically refers to cable providers
PCA	Principal Component Analysis – A type of machine learning algorithm used for feature transformation
RF	radio frequency
SVM	Support Vector Model – A type of machine learning algorithm used for classification

Bibliography & References

The Singularity is Near, By Ray Kurzweil, ISBN - 9780715635612

Declining storage costs, mkomo.com, URL: http://www.mkomo.com/cost-per-gigabyte

Statistical Modeling: The Two Cultures, Leo Breiman , URL: https://projecteuclid.org/euclid.ss/1009213726

Gauss-Markov assumptions for linear regression models, URL: https://en.wikipedia.org/wiki/Gauss%E2%80%A93Markov_theorem

The Netflix Prize, URL: http://www.netflixprize.com/

Amazon Mechanical Turk, URL: https://www.mturk.com/mturk/welcome