SCTE® CABLE-TEC EXPO®
FAST FORWARD 2021

ATLANTA, GA
OCTOBER 11-14

SCTE
a subsidiary of CableLabs®

UNLEASH THE POWER OF LIMITLESS CONNECTIVITY

2021 Fall Technical Forum
SCTE® • NCTA • CABLELABS®

Operational Transformation

Right Technician at the Right Time

Anastasia Vishnyakova

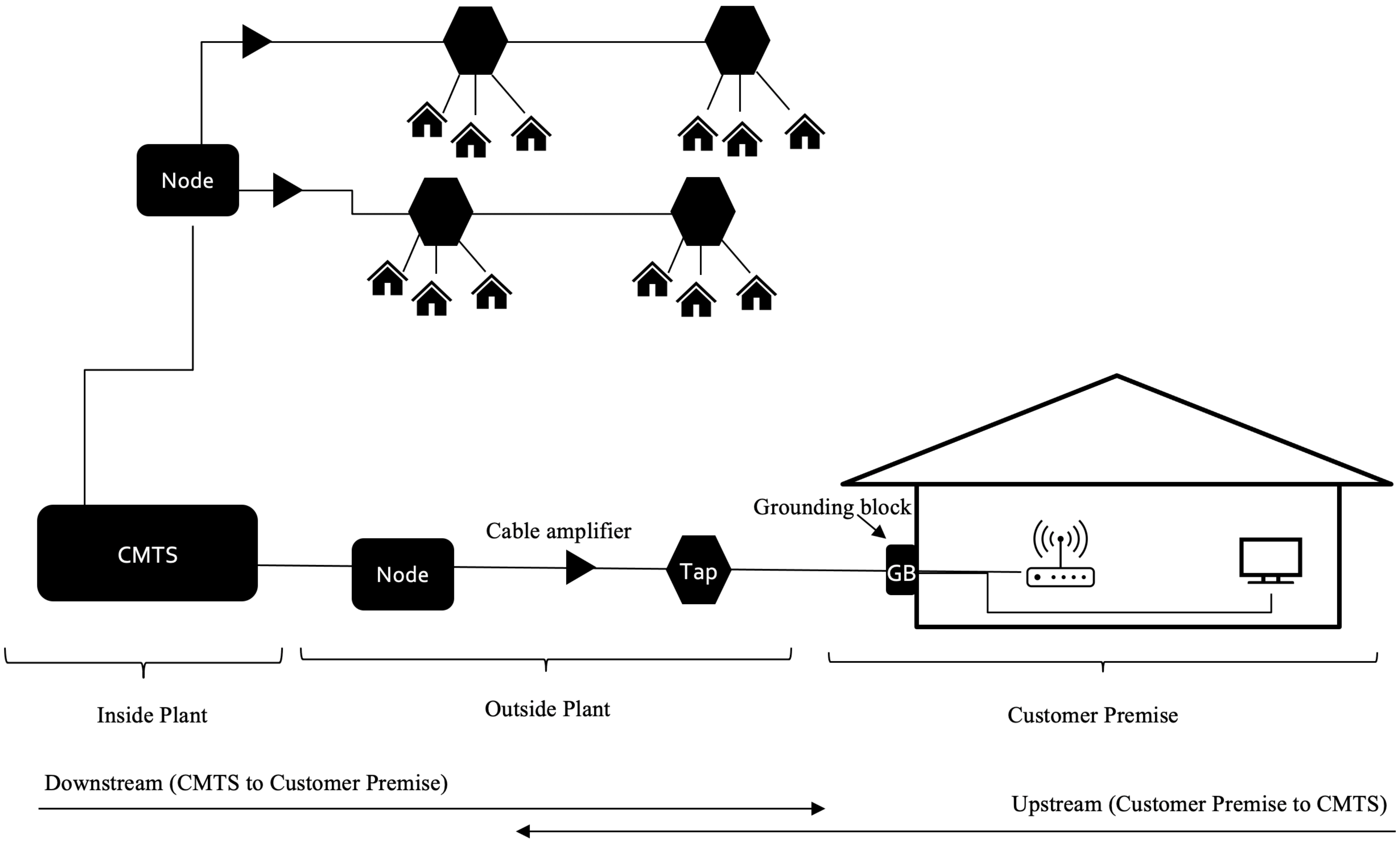
Machine Learning Engineer
Comcast

SCTE
a subsidiary of CableLabs

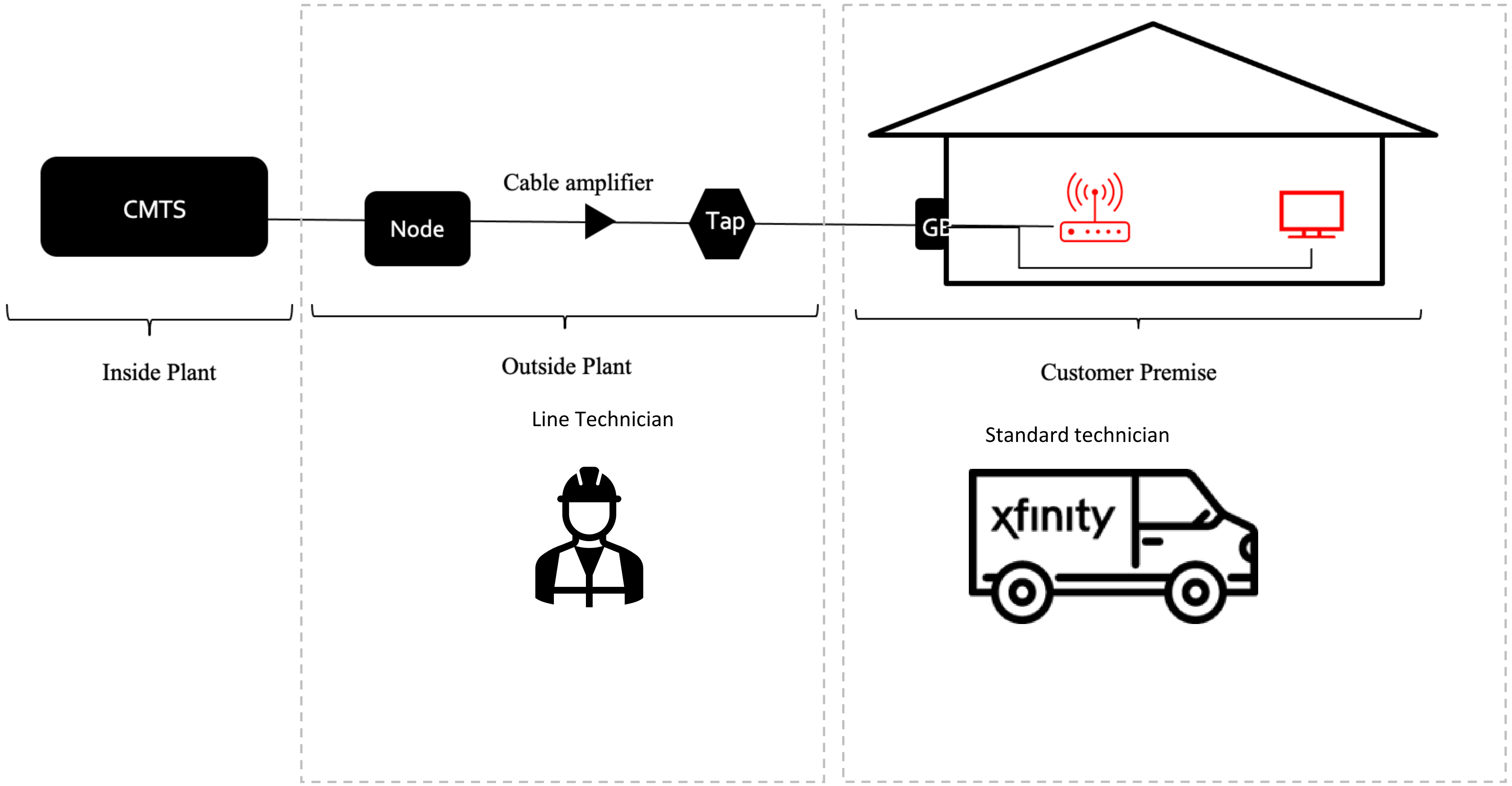
2021 Fall Technical Forum
SCTE • NCTA • CABLELABS

SCTE CABLE-TEC EXPO FAST FORWARD 2021
VIRTUAL EXPERIENCE
OCTOBER 11-14

Plant is a Complex System with Many Components



Troubleshooting problems takes different skills



Right Technician at the Right Time

- 3% of technician visits end up in line tech referrals

Machine learning model can identify when line technician is needed

Right Technician at the Right Time

$ Save cost of sending both technicians

☺ Service customers better

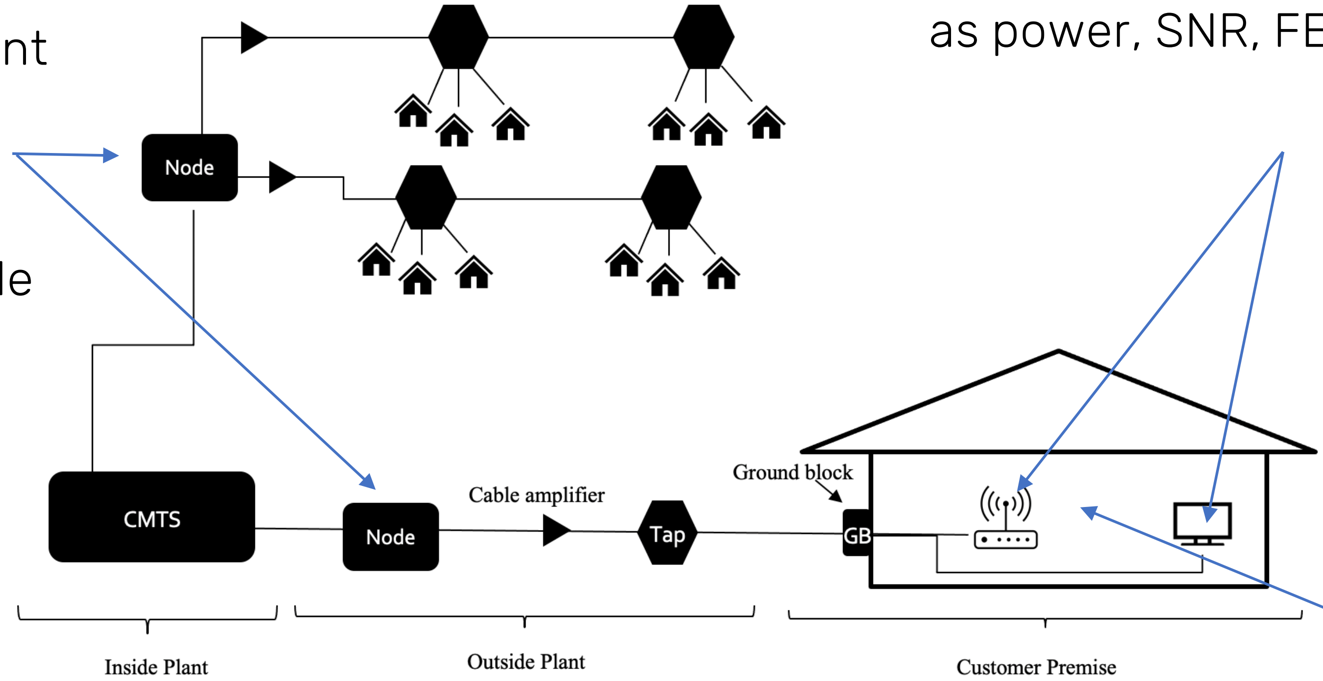
Remote DOCSIS Telemetry Data

<u>Node-Level Network Degradation Impairment Analysis Tool</u>

- Consumes account and node-level data
- Identifies events impacting multiple accounts
- Reports outages and plant issues

<u>Proactive Network Maintenance Software</u>

- Polls devices several times per day
- Reports upstream, downstream telemetry such as power, SNR, FEC, ICFR, Timeouts, etc.



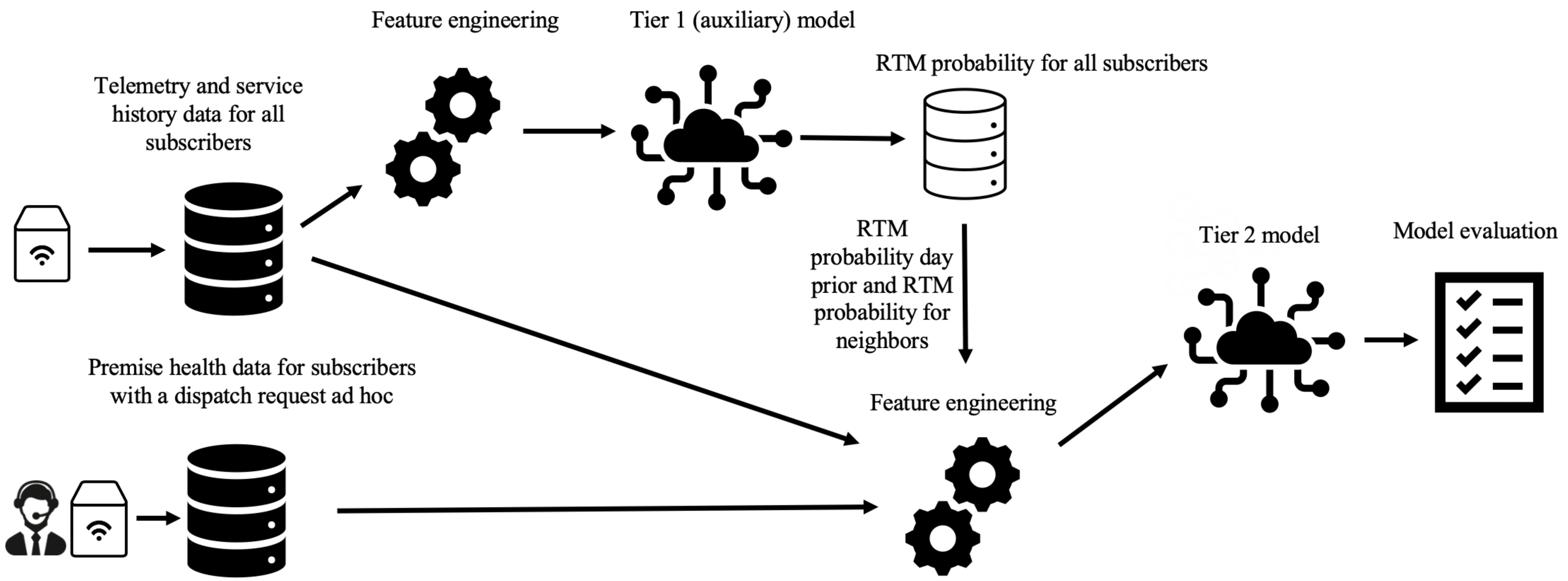
<u>Account and Device Network Analysis Degradation Tool</u>

- Consumes data from pollers and identifies impairments.
- Checks for isolated device, whole home, and neighborhood impairments

RTM Machine Learning Model

- XGBoost Classification Model
- Trained on the universe of subscribers with trucks
- Line technician visit confirmed as “valid” is target label
- Model weights to address class imbalance
- Tuned to use 500 estimators with max depth 2
- Over 400 input features
- Two-tier model approach (model stacking)

Model Training Process

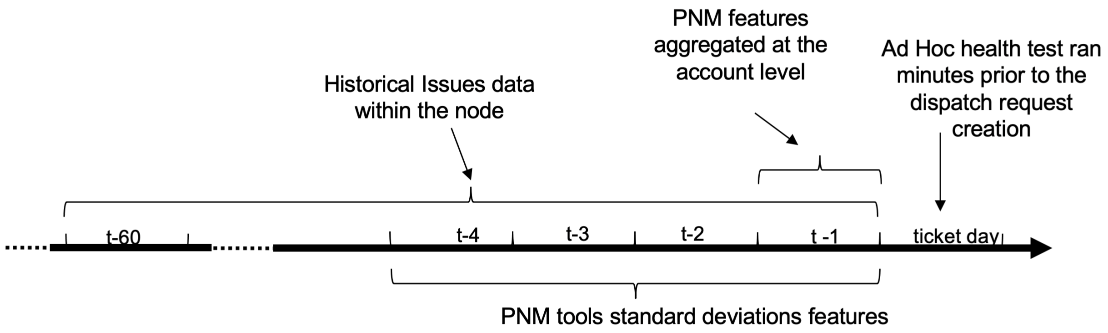


Feature Engineering

- Features from PNM poller software, device/account-level analysis tool, and node analysis tool are transformed to aggregate multiple polls to daily account-levels using min/max and avg aggregations.

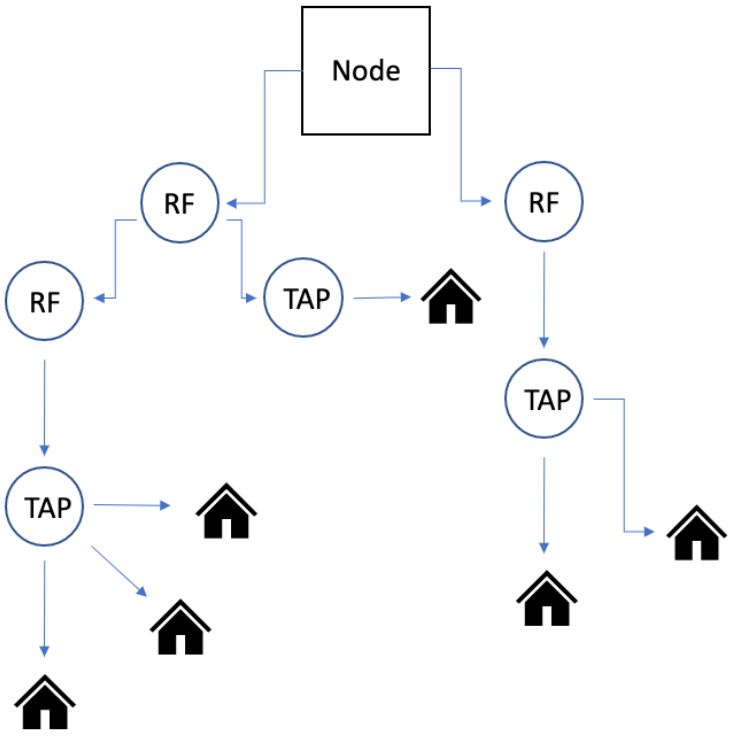
- Standard deviations are examined across 4 days for selected feature

- Cumulative measures are transformed to absolute values (for counters), data are normalized (for rates calculations such as FEC)



Topology Aggregations

- Directional data representing paths from node to subscribers' homes
- Some graphs are short, others are deep
- Feature aggregations:
  - tap (parent) level
  - across amplifiers supporting the customer
  - comparisons between neighbors

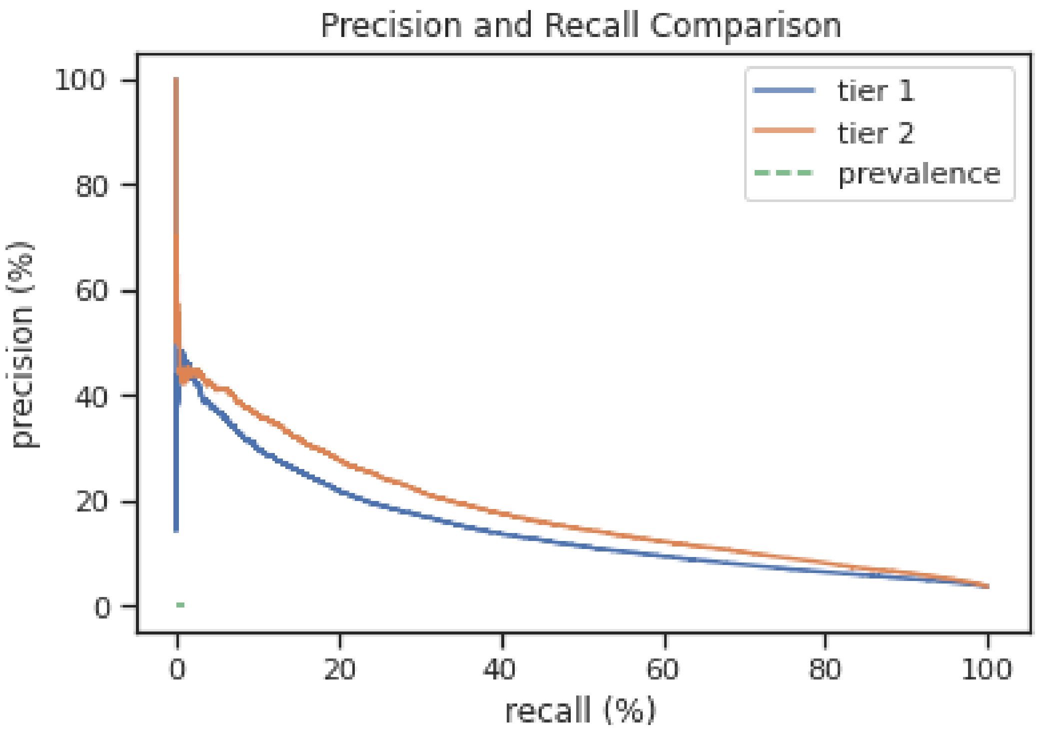


SCTE. CABLE-TEC EXPO
FAST FORWARD 2021
ATLANTA, GA > OCTOBER 11-14

Model evaluation

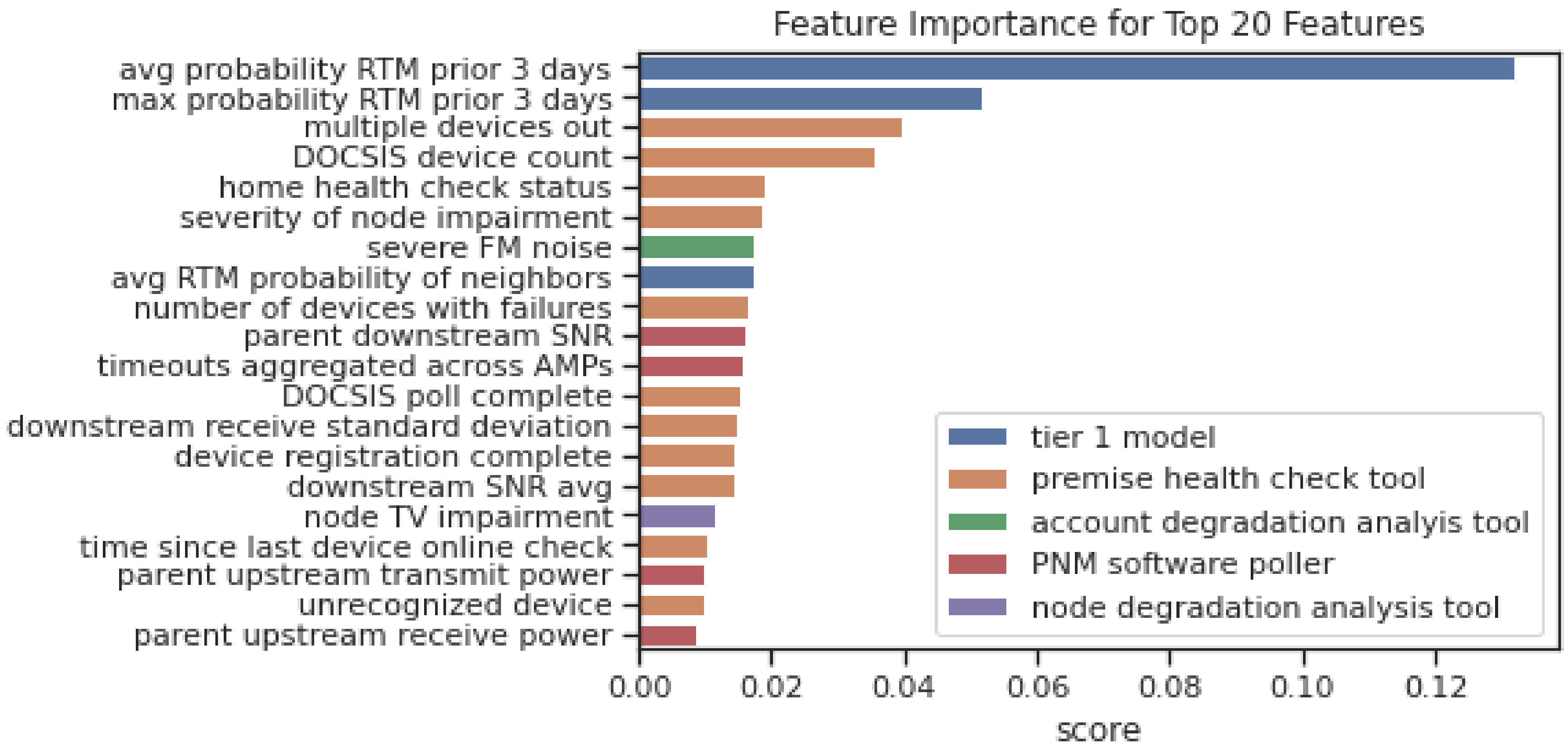
Model performs 13x better than random guess

- We evaluated the model on the out-of-sample test data split on time
- At 5% recall, tier 1 model precision is 37.2% precision, tier 2 – 41%
- Tier 2 model performs 11 times better compared to the random guess (average prevalence at 3%)

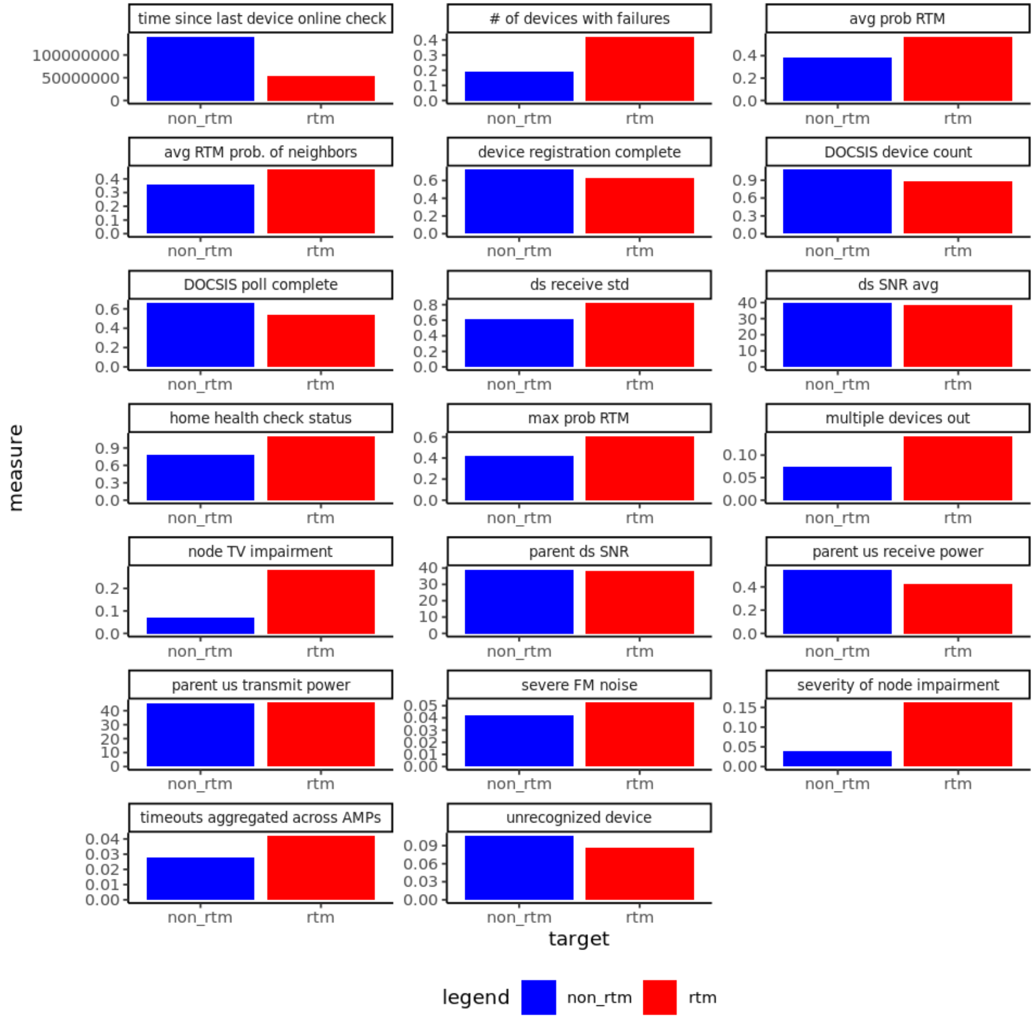


 | expo.scte.org

Top Importance Features



Feature Exploration



- We explore directionality of unscaled features for top 20 influential features
- Device failures, prior RTM probability, high downstream standard deviation, high number of impairments, for example, are positively correlated with RTM outcome.
- SNR, complete registrations, complete polls, are negatively correlated with RTM outcome

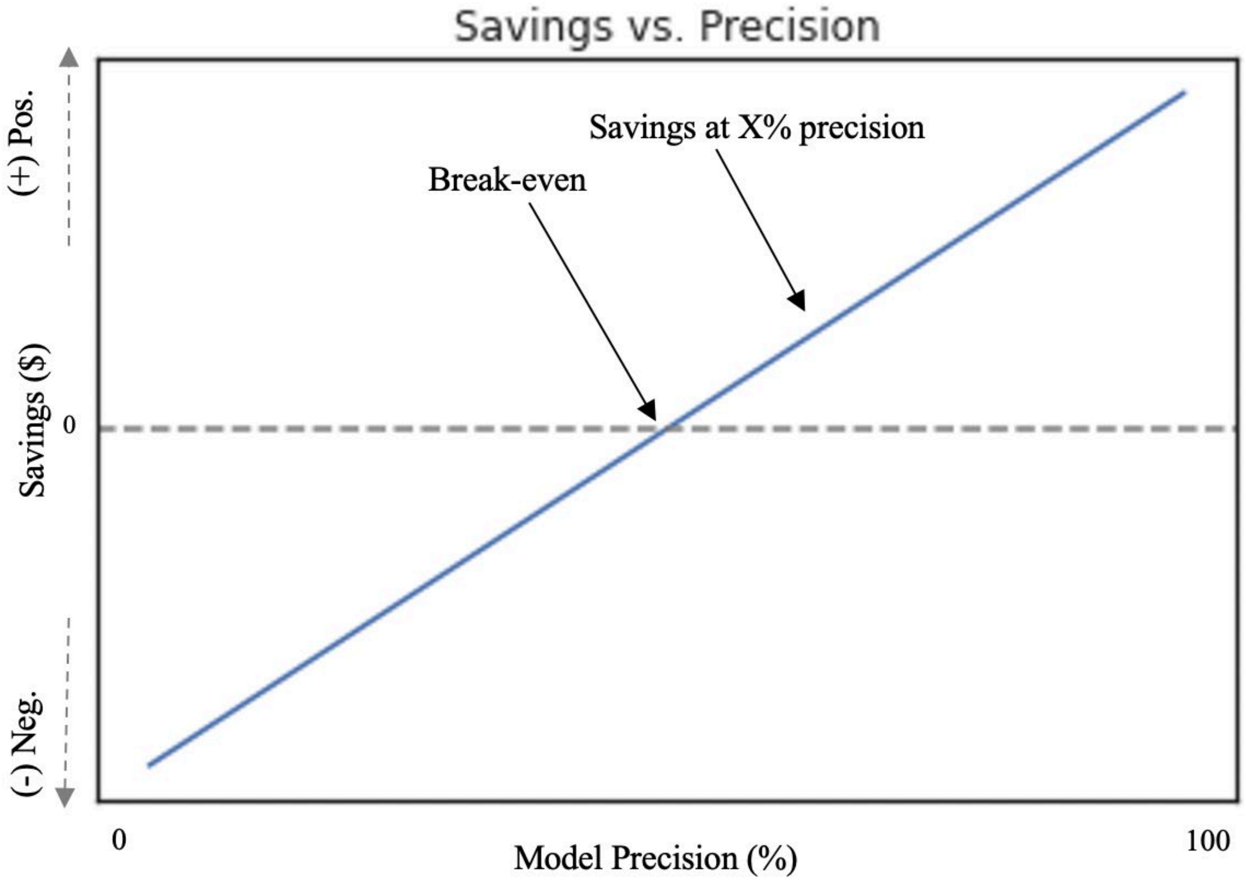
F*P-L*1-Y*1-P = Snet

Is the Model Good Enough to Deploy?

- Cost analysis can help to define minimum precision required to incur a net zero savings relative to BAU.
- Customer-level (only customer with a truck) and neighborhood level approaches (additional the impact to the neighborhood)
- Denote
  - Cost of fulfillment truck as F and cost of line truck as L, where L > F
  - Operations and research cost as R
  - Cost of True Positive: (L+F)-F -R = F - R
  - Net Savings assuming cancelation rate Y
- Savings = $Saving_{TP} - \mathrm{Cost}_{\mathrm{FP}}$

F*P-L*1-Y*1-P = Snet

Is the Model Good Enough to Deploy?



Net Savings:

$$\underbrace{\left((F - R) * P\right)}_{\text{TP}} - \underbrace{\left(L * (1 - Y) - R\right) * (1 - P)}_{\text{FP}} = S_{net}$$

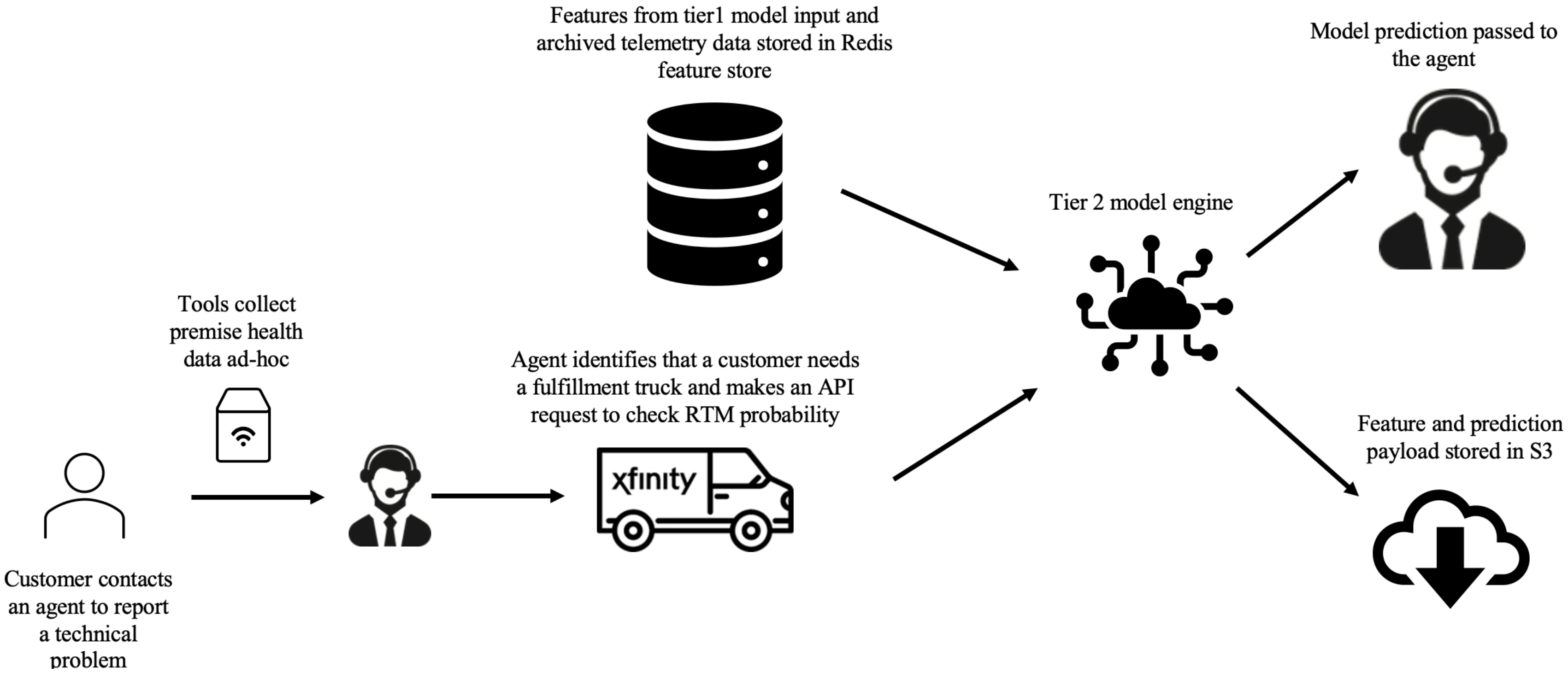
Precision to break even:

$$P_{even} = \frac{L(1 - Y) - R}{F - 2R + L(1 - Y)}$$

“Soft” Model Deployment

- POC with a ”Human in the Loop”
- We engaged 4 regions to participate in the trial
- RTM candidates sent using AWS SES and evaluated by SME off-site
- Debrief with SME helped us to discover additional data sources, improve feature engineering, and identify false positive cases

Model Deployment



Conclusions

Machine learning is key to the operational transformation

- Telemetry measurements can help us identify impairments
- Machine learning is key to prioritizing issues
- RTM model will help us move from reactive to proactive approach
- Issues resolved in one, not two visits
- Avoiding neighborhood impact of impaired service, additional calls and trucks
- Tune the model to achieve precision to deliver net benefit to the system
- Right Technician at the Right Time

Co-Authors

SCTE. CABLE-TEC EXPO
FAST FORWARD 2021
ATLANTA, GA > OCTOBER 11-14

Thank you to the colleagues who collaborated with me on the paper

- Rama Mahajanam
- Mike O'Dell
- May Merkle-Tan
- Catherine Hay
- Lisa Pham

SCTE® CABLE-TEC EXPO® FAST FORWARD 2021

ATLANTA, GA
OCTOBER 11-14

SCTE
a subsidiary of CableLabs®

Thank You!

Anastasia Vishnyakova

Machine Learning Engineer
Comcast
Anastasia_Vishnyakova@comcast.com