This document contains approaches for determining the important features which can impact a given target feature.

Find Variables That Impact Jitter Variable

From the given dataset, we are interested in finding features that impacts jitter variable.
We explored two approaches for finding features which impact variable jitter.

1. **Correlation based approach**
   In this approach, we compute correlation between jitter feature and other features. We find those variables which have highest correlation with jitter feature. features with highest correlation tend to impact the jitter variable. Constraints of this approach is that it could capture linear dependencies only, but not higher order dependencies.

   Features that affect jitter most
   wirelink_dl_mbps, wan_dl, num_streams, download, wirelink_up_mbps, tput_theta0, avg_noise, pq_airuse, pq_requested, chanutil_intf

2. **XG(Extreme Gradient Boosting) Boost based approach**
   For finding features which impact jitter variable, a xg boost based regression model is build. Feature importance measures the amount of variance explained by a particular feature while building regression model for jitter.

   Features that affect jitter most

   Use of Unscaled variables :
   rx_bytes, wan_dl, epoch, tput_theta2, mcs_drops, avg_util_tx, tput_uerr, rx_rate, rt, upload

   Use of Scaled Variables :

   rx_bytes, wan_dl, epoch, tput_theta2, mcs_drops, rx_rate, tput_uerr, pq_dropped, rt, upload

Find features which are impacted by rssi

In order to find features which are impacted by rssi, we can compute the correlation of features with rssi, either positively or negatively. We can utilize the correlation values to select the features which are impacted by rssi. Choose those features which have highest correlation with rssi. These are the features which are impacted by rssi.