



Case Study – Weather Effect on Trouble Tickets

Spring & River

Problem Statement:

Telecom related troubles are logged regularly by customers for installed devices and connection in their premises. Work order to address the problem and sort out by technician is managed by workforce management based on prediction in prior from the weather forecast report of 7 days in ahead.

One major problem in prediction of trouble tickets due to weather is extreme weather/climate conditions. In house prediction model cannot reach to the number of trouble tickets for such a scenario.

Goal of this project was to predict/match the trouble tickets number for extreme weather/climate conditions as well as for normal weather days. Company is using random forest algorithm to predict the trouble ticket number. The challenge is to reach to the near prediction of trouble ticket numbers where company's model is failed to match. This will serve in prior workforce management planning and schedule to face the scenario of extreme weather condition in time.

Dataset:

The dataset given was log of trouble tickets due to weather conditions and weather forecast of last three years collected from different weather stations across wire centre group.

Pre-processing and Exploratory Data Analysis (EDA):

After initial observation of two data sets, missing value was treated and not related information are removed. Dimension of data is rationally reduced to focus on relevant features as suggested by the package Boruta. Strong features are selected with number as 16. Weather data is merged with trouble ticket for each day on each wire centre under the wire centre group. By distribution plot of features, outliers are removed. Weather variables are closely examined to find the relationship with the prediction. Extreme weather days are identified and data value are explored to find any possible reason for high trouble ticket numbers.

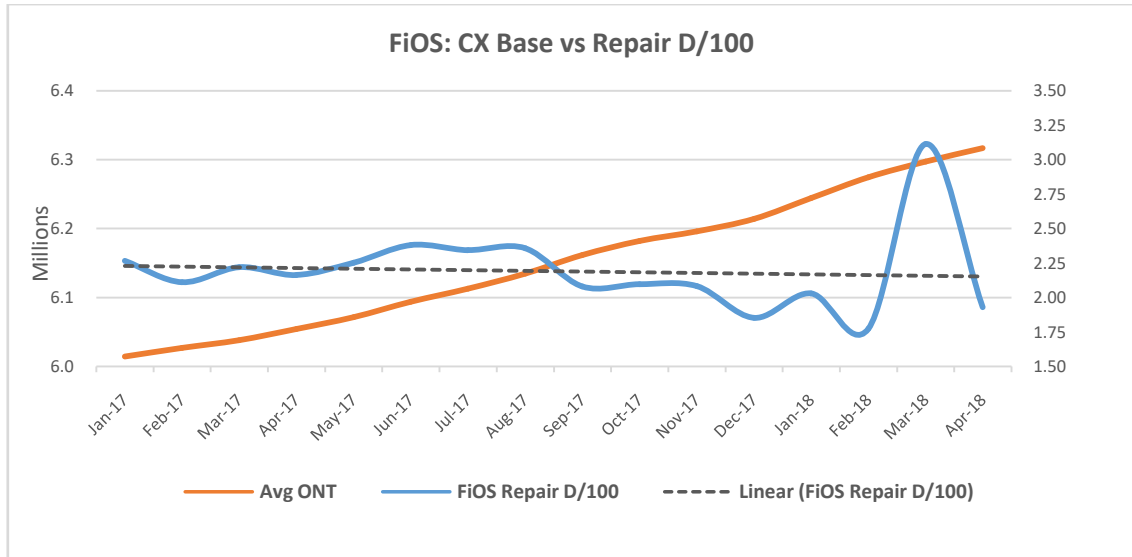
Train Model:

The clean data has been split into train and test with 70:30 ratio. We build a multi linear regression model with moving average of weather variables. The adjusted R^2 has shown result for normal weather date vs. extreme weather date as 57% vs. 51% for fibre connection and 73% vs 68% for copper connection respectively.

This result indicates that we need to find out some other important variable which has relation with the number of trouble tickets might have so we need to re-engineering feature for better prediction.

Feature Engineering:

It is observed that no of trouble tickets in different wire centres within the same global region under same weather forecast condition are varying. The reason is directly related to the no of relevant fibre and copper connections.



We have added this feature in the data set and have used GBM with fine tuning parameter for regression prediction and found the more accurate prediction for both weather and extreme weather date.

Accuracy reached to 82% for weather and 79% for extreme weather condition dates. This can be compared with the result of existing in-house model near to 50%.

Conclusion:

Weather model for telecom trouble tickets have been explored in machine learning frame work. The linear model suffers from low Adjusted R-squared which has been enhanced by introducing new features like the number connection of fibre and copper in the data set and using GBM model.

For better result seasonal change of weather with respective geographical area from past data needs to be recorded within the training data set to train the model for better prediction of extreme weather condition. Subsequent effect of extreme weather condition has to be also to be taken care to predict more trouble tickets though those dates are normal weather condition as per weather forecast data.