**`INCIDENT PREDICTION**

**Objective:**

Predicting forthcoming incident tickets to prevent SLA breaches and maintain customer satisfaction by giving better services.

**Overview of Analysis:**

Predicting incidents using ML algorithms (DTR, RFR and KNNR) based on past 16 months data.

This contains 4 phases:

1. Data preprocessing (Cleaning and deriving new variables).
2. Training models with past incidents.
3. Validating model performance by comparing last 10 days predicted incidents and actual incidents.
4. Predicting incidents for next 5 days.

**Data details:**

Table name : **INC\_2019**

Columns used : Six Columns used

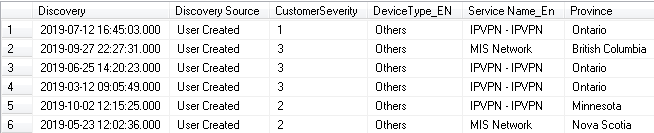
|  |
| --- |
| **Discovery** |
| **Discovery source** |
| **CustomerSeverity** |
| **Device Type\_EN** |
| **Service Name\_En** |
| **Province** |

Data period : 01-Jan-2019 **to** 06-Apr-2020

No. of. Records:  **7547**

Data base : Usecases (167 server)

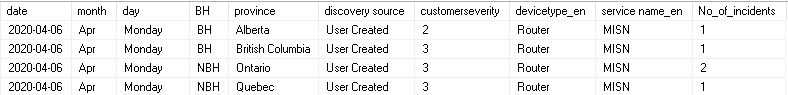
Sample data :



**Data preparation:**

1. Removed “Null” values and made column levels in proper format (Removing special characters, and upper case letters and so on)
2. Binning of Province column. (20 % of data has so many provinces so Model won’t pick pattern )
3. Binning of Device type column. ( 90 % data contains only Switch and Router, so rest are binned)
4. Derived Date, Day, Month and Business Hours variables from **Discovery** column.
5. Derived No\_of\_Incidents column by grouping all the variables.

Sample data for training models:



**Model Building:**

Here we are going to predict continuous data and we have used regression models for training and gave an ensemble output.

Target : **No\_of\_incidents**

Predictors : Month, Day, BH, Province, Discovery source, Customer severity,

Device type\_en and service name\_en.

Models used:

1. Decision Tree Regression
2. Random Forest Regression
3. KNN Regression (K = 4)

Note: sk-learn library was used and default parameters have applied.

Model Performance: (Trained and validated with 100% data)

|  |  |  |
| --- | --- | --- |
| **Model Name** | **R-Square** | **RMSE** |
| DTR | 0.89 | 1.44 |
| RFR | 0.83 | 1.57 |
| KNNR | 0.69 | 2.14 |

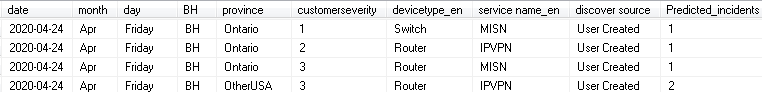
Formula for ensemble output:

Final output = (2\*DTR+KNNR+RFR)/4

**Prediction:**

Currently we are predicting incidents for five days and it will be automatically updated in IncidentPrediction\_N5 table in 167 server.

Sample output:



**Validation:**

The model performance will continuously monitored by validating the last 10 days predicted output and actual incident arrival. If the variation between actual and prediction is high, we need to retrain the model.

Scripts and documents path:

C:\Users\Kumaresan\Desktop\Incident prediction