### Proposal Objective:

In this project, we intend to process complain data from BCRM and identify possible root cause and bucket them under predefined category.

This proposal is aimed to automate the whole process of root cause analysis. We intend to create a data flow pipeline which will automate the process right from source till the visualization and categorization of root cause analysis.

We will use python programming to apply text analytics algorithms on SN remarks data and create insights using secondary analysis. Python will also be used for data inspection, data quality check and data profiling.

The system will also create word frequency dictionary based on SN remarks which will help in identifying the keywords related to RCA. The reporting system will identify outliers and help further investigation of RCA.

### Benefits

* Automation – with much lesser manual intervention
* Resource dependency and timeline dependency
* Enhance reporting mechanism
* Dynamic keyword dictionary
* Reduce mapping gaps on root cause analysis

### Estimated Time to Complete Root Cause Analysis

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **MAJOR ACTIVITIES** | **ESTIMATED DURATION FOR CURRENT APPROACH (MONTHLY)** | **ACTUAL MAN-DAY FOR CURRENT APPROACH** | **ESTIMATED DURATION FOR PROPOSED SOLUTIONS (DAYS)** | **ESTIMATED MAN-DAY FOR PROPOSED SOLUTIONS** | **ESTIMATED MAN-DAY PERCENTAGE REDUCTION** | ASSUMPTIONS |
| **FETCH SOURCE DATA** | 15 | 45 | 5 | 15 | 67% | * Each data source should have a pre-defined format |
| **CONSOLIDATE AND PROCESS RAW DATA** | 15 | 45 | 5 | 15 | 67% | * Computational resources should be provided by TNB * Computational server environment set up is done at TNB side |
| **ANALYZING ROOT CAUSE** | 15 | 60 | 5 | 10 | 83% | * SN Remarks data should be updated on week/monthly basis * Keywords for tagging should be updated by TNB |
| **SECONDARY ANALYSIS AND VALIDATION RCA** | 30 | 60 | 7 | 21 | 65% | * Pre-defined rules are provided by SMEs (TNB) |
| **VISUALIZATION AND REPORTING** | 30 | 30 | 5 | 10 | 67% | * Layout formats and graphs to be defined by TNB * Any additional graph or regular work may change the estimated time for the whole project |

Below table shows brief overview of steps needed in the process:

|  |  |  |  |
| --- | --- | --- | --- |
| **Step 1** |  | Fetch Source Data | * Extraction of complain data * BCRM team will extract all the billing data and SN remarks for all service request and complaints for root cause analysis. * BCRM team will extract the latest Meter Installation Tampering (MIT) date, Reversal data and Meter Installation Date to be incorporate with the root cause analysis. |
| **Step 2** |  | Consolidate and process Raw Data | * Cleansing and formatting data as per the input requirement of the script. * Consolidate and build relational model. |
| **Step 3** |  | Analyzing Root Cause | * Text analytics is used to analyze unstructured text by identifying root cause based on the list of keywords. * Apply text analytics on SN Remarks to identifying the root cause. The text analytics algorithm will categorize the SN Remarks into multiple root causes by capturing the pre-defined keywords. * Below are the root causes that we will focus on for the text analytics: - * Meter Faulty * Incorrect Reading * Estimated Bill * Long Billing Period * Meter Crossing * Meter Tampering * Customer Consumption Behavior |
| **Step 4** |  | Secondary Analysis and Validation RCA | * Root cause verification will be performed by using the secondary analysis. |
| **Step 5** |  | Visualization and Reporting | * Driver tree report. * Slicing and dicing of billing data on consumption pattern. |

A screenshot of a cell phone

Description automatically generated

Data Flow Diagram of the proposed solution for RCA

## Fetch Data Source

* It is required to freeze the total count of service request and complaints on monthly basis.
* BCRM will extract all the billing information and SN remarks for the fixed service requests and complaints of the month.
* Status report on all service request and complaints are required for root cause validation. Once the status is completed, it should fall under a root cause and not further investigation.
* BCRM will extract the latest Meter Installation Tampering (MIT) data, Reversal Data and Meter Installation Date data to be incorporate with our analysis.

## Consolidated and Process Raw Data

* Cleansing and formatting data based on requirements.
* Consolidate all the relevant files and identify the association between them.

# Analyzing Root Cause

* Build text analytics algorithm to categorize all the service request and complaints into multiple root causes based the list of keywords.
* Adopting the list of keywords help to determine 7 different root cause: -

1. Meter Faulty
2. Incorrect Reading
3. Estimated Bill
4. Long Billing Period
5. Meter Crossing
6. Meter Installation Tampering
7. Customer Consumption Behavior

* A main excel sheet will be created for the list of keywords. The list of keywords can be updated any time as keywords from the main excel sheet will be synchronize with the text analytics algorithm.
* Those uncaptured data from text analytics will be bucketed into further investigation.

### Secondary Analysis and Validation RCA

* Captured data from text analytics will be validated by using secondary analysis: -

1. Long billing period – Check billing period (i.e. billing period more than 31 days)
2. Estimated bill – Compare billing type for previous month and current month (i.e. Estimate/Estimate and Estimate/Actual)

* Captured data with no billing information will be sent to rangers’ team to investigate the root cause in the billing system.
* Uncaptured data from text analytics will be categorized as further investigation.
* The root cause analysis validation data from BCRM and Rangers will be the dominant file for the whole analysis output. Below is the overall hierarchy for input files that we consider in our root cause analysis: -

1. BCRM validated root cause
2. Ariff’s analysis
3. Text Analytics output
4. Secondary Analysis output

### Visualization and Reporting

* Two main reports will be designed using visualization tool reporting: -

1. Driver tree report
2. Consumption pattern based on service request and complaints