Software Requirements Specification

for

Traffic prediction and anomaly detection in wireless cellular networks using call detail records

Version 1.0 approved

Prepared by Sameer Khan

Abhishek Kumar

Mayur Kharbas

Rahul Gupta

21-09-2018

# Introduction and Requirement Analysis

## Purpose

The intended software to be developed is a data analytical model which will be provided in the form of a web application. Its main objective is to predict cellular traffic in wireless mobile networks and anomaly detection to improve Quality of experience for mobile users using call detail records. This document provides an overview, scope and functional and non-functional requirements of the proposed product.

## What is the problem?

a) Huge load on existing network due to explosive growth of call and data consumption in recent years. This causes network congestion and increases call drop rates.

b) Lack of existing infrastructure. In India one cell tower manages above 400 customers as opposed to global average of 200-300. Thus resource (bandwidth) allocation becomes an extremely vital process in cellular network operation.

c) Sudden spike in usage leads to congestion and network deterioration. Early detection is key to prevent revenue loss and customer churn rate.

## Scope

The web application proposed is aimed for providing solutions to telecom operators. It will help them gain an insight into their consumer base and analyze customer usage pattern and behaviour to come up with efficient sales strategy and optimize operations.

# Functional Specification

## Product Functions

* Predicting cellular traffic of users based on ground truth data
* To detect anomaly in cellular network. Anomalies result in Network outage or are a result of network outage. Anomaly could be a sudden spike in usage leading to network congestion or an unusually low usage due to poor network
* Providing various infographics from data analysis to provide actionable insights to users

## User Classes and Characteristics

* **Telecom operators:**
  + - **Sales department:** Based on the traffic predicted, the sales department can come up with usage plans for customers to achieve high Return on Investment.
    - **Operation department:** Early detection of network outages based on anomaly detection. High usage areas and time can be determined and bandwidth can be allocated accordingly.
    - **R&D department:** The analysis provided will be useful to the R&D department.

## Design and Implementation Constraints

* Call detail records of consumers are protected under the IT Act of the Government, thus the data is only available to Telecom operators and Law Enforcement officials.
* Data provided needs to be preprocessed to obtain the desired output.

## User Documentation

After the completion of the product, user manual will be provided.

## Assumptions and Dependencies

* Data must be in specified format
* Spatiotemporal data is required
* It is assumed that the Telecom Operator maintains the Call Detail Records of its users

## Traffic Prediction (Scenario)

2.6.1 Description and Priority

The model will output the cellular traffic it predicts using the data fed into the system. This feature is of the highest priority since it is the main objective of the product.

2.6.2 Stimulus/Response Sequences

The user will get a detailed analysis of the call detail record submitted. This includes traffic prediction and anomaly detection which will help the Sales and Marketing as well as Operation department of the Telecom operator. After submitting the data the user will get the results in the form of graphs and other infographics to help gain actionable insights.

2.6.3 Functional Requirements

REQ-1: Call Detail Records

The dataset should be an aggregated data of call details in an area like a city or a neighbourhood.

REQ-2: The dataset should be in csv format.

REQ-3: The data should be spatiotemporal representing a geographical location.

Common Mistakes/ Errors: Input file not in appropriate format.

Expected Messages: File format incorrect.

Open Issues: Limitation on the size of the input file not yet decided.

# External Interface Requirements

## User Interfaces

A web application where user will upload call detail record and they will get various output in form of graphs, histograms and other infographics.

A web application with supported browser is required on the user’s side:

* Google Chrome 45+
* Mozilla Firefox 45+
* Microsoft Edge
* Apple Safari

It will not support Internet Explorer.

## Communications Interfaces

* Web application run on a web browser.
* Communication standards used are FTP and HTTP

# Other Nonfunctional Requirements

## Performance Requirements

The time taken for the results provided should be proportional to the size of the input dataset. Expected time of completion should be provided to the user for ease of use.

## Safety Requirements

The call detail records of the users are protected by the government under Section 81 of the IT Act. Thus the web application must ensure that this data is not compromised.

## Security Requirements

The web application must use https protocol to ensure data security. The IT act in the Indian Constitution protects the use of Call Detail Records, thus it is paramount that the data must not be compromised.

## Software Quality Attributes

The application must provide correct infographics based on the analysis conducted on the dataset given. It must also provide accurate error feedback if the dataset provided is incompatible. The output should also be displayed in a timely manner depending on the size of the input size.

# Technical Specification

## Algorithms

1) K means clustering for anomaly detection

K-means can be applied among “total activity and activity hours” to find the usage pattern with respect to the activity hours.

By using this clustering mechanism, we can find the clusters making more traffic to the telecom network in the measure of total activity.

2) Arima model for cellular traffic prediction

ARIMA stands for Auto-Regressive Integrated Moving Averages.

It can be used for forecasting a stationary time series data.

## Dataset:

Dataset is taken from Telecom Italia.

The Spatio-temporal data was made public by Telecom Italia as a part of their Open Big data Initiative.

The data is geo-referenced and is processed from CDRs of their subscribers

For each grid, following information is present -

1) Grid ID,

2) Time stamp (in milliseconds) of 10 minutes duration,

3) Country code,

4) Activity in terms of received SMS,

5) Activity in terms of sent SMS,

6) Activity in terms of inbound calls,

7) Activity in terms of outbound calls, and

8) Internet usage.