# Stage 1 Calculation of SDI

## Tile description

Each tile has a unique id and the position of the tile and its subsequent child tiles is fixed on the surface. We have (Lat, Lon) co-ordinates of the parent tile at level 5. For level 6 and 7 we have average co-ordinates of the devices present in that tile.

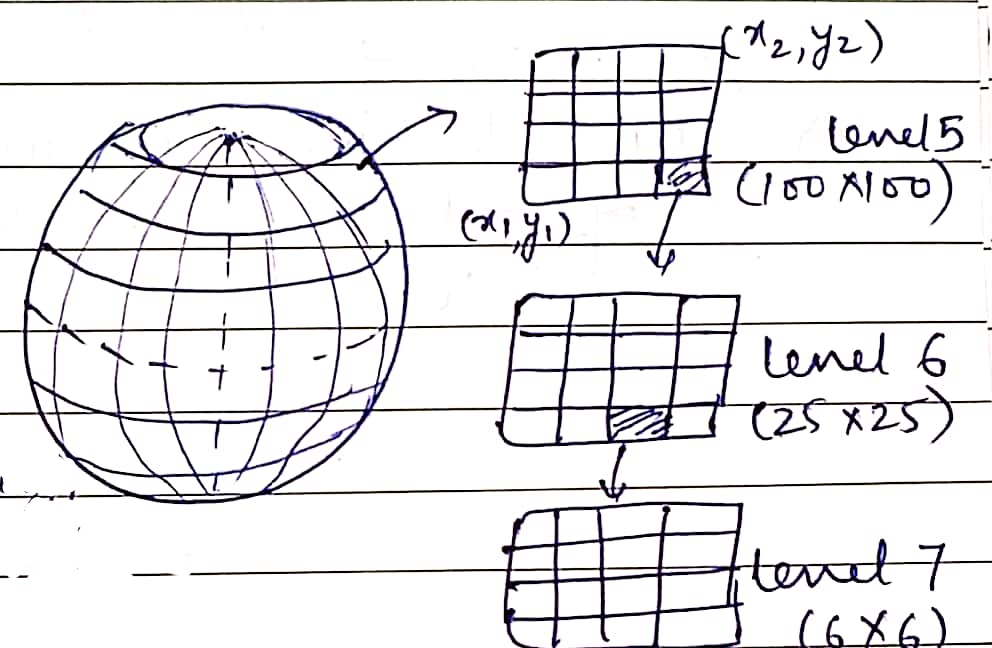


Figure 1 Tile description at 3 levels

The data we have is for each unique *tile id*, number of devices and number of new devices at a time-interval of 15 minutes daily for level 6 and level 7. Below is the example of Tile data:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Timestamp** | **Tile id** | **# devices** | **# new devices** | **Average Lat** | **Average Lon** | **Ping** | **Average HPE** | **Approx Median HPE** |
| 2020-03-02 00:30:00 | T1 | 5 | 3 | X1 | Y1 | … | … | … |
| 2020-03-03 00:30:00 | T1 | 4 | 1 | X2 | Y2 | … | … | … |
| 2020-03-03 00:30:00 | T1 | 3 | 2 | X3 | Y3 | … | … | … |
| 2020-03-02  00:30:00 | T2 | 1 | 0 | X4 | Y4 | … | … | … |
| 2020-03-03 00:30:00 | T2 | 2 | 1 | X5 | Y5 | … | … | … |
| 2020-03-04  00:30:00 | T2 | 1 | 0 | X6 | Y6 | … | … | … |

Figure 2 Sample tile data

\*Tile id is fixed by co-ordinates

## SDI Calculation

* To calculate SDI based on the historical data provided, we are going to use the following features:
  + For each tile and each time period, derive the distribution of historical distances of clusters using average coordinates of the clusters (tiles at level 7)
  + Zone/Category of the tile (c)
  + Effective area of the tile (a)
  + # of new devices (n)
  + # of devices (d)

Hence, SDI can be calculated at Tile level basis as:

SDI = ; *σ represents the Variance of the distance*

i.e**. SDI (Tile T, time t) =**

Assumption: Movement of people inside a tile is driven by certain purpose.

* SDI will be calculated hierarchically for the child tiles (for eg. level 7) and it will be used to calculate SDI for the parent tile (for eg. Level 6).

## Step 1

The calculation of distribution depends on the distances calculated in a Tile at given time interval (15 min).

For example, at time t1 for different days, in each tile, coordinates of devices are (x1, y1), (x2, y2) and so on. We'll take average of coordinates which is X (refer to Fig 3) and calculate distances of all the devices from X. The distribution of distances will represent the concentration of people at any time interval in tile T.

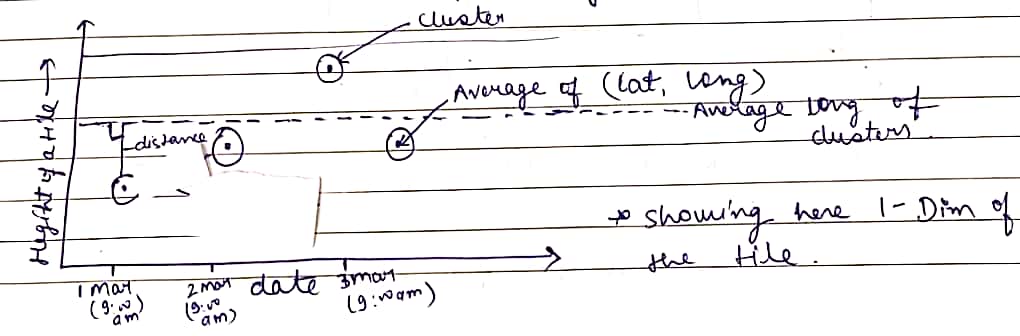


Figure 3 Distance calculation for a set of Devices

## Step 2

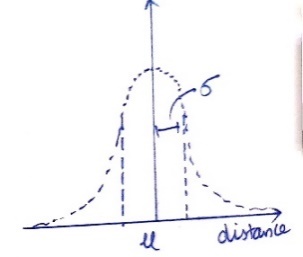
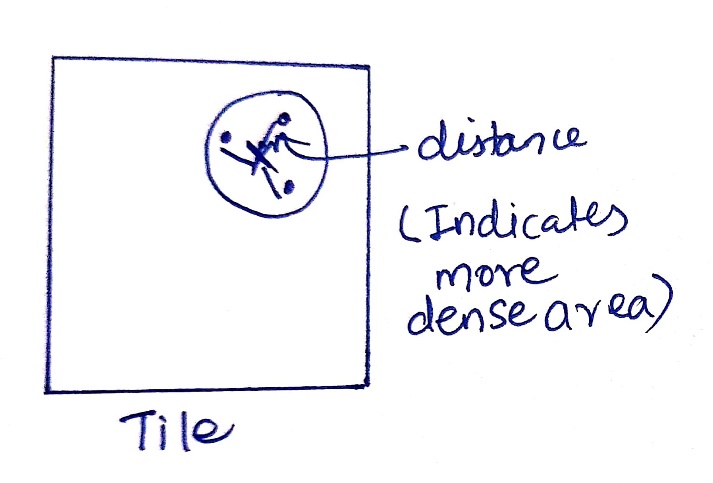


Figure 4 Example of Hierarchical tile level distribution

* Plot average coordinates of each clusters/devices
* Average of these clusters can be respresented as X
* Calculate distance between each cluster with the average coordinates
* Plot distance (d1, d2,……, dn) in order to get the variance among clusters
* Lower the variance (σ) of the distance indicates low value of SDI



## Stage 2 Forecasting SDI

* From the previous step we'll calculate SDI for historical data for a given tile at a given time-interval. This helps us to create time-series like structure of historical SDI for 15m time slots/tile. Advanced Machine learning techniques such as neural network model can be used to predict SDI on rolling basis.

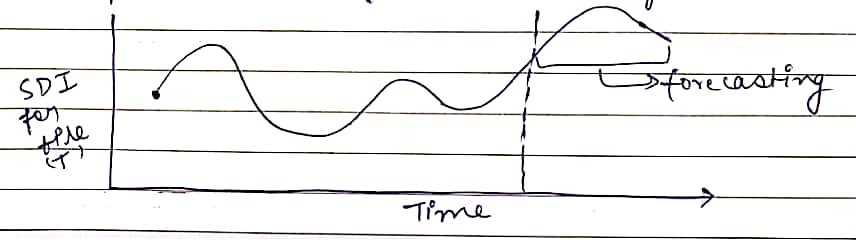


Figure 5 SDI Trend with respect to time

Training parameters required to forecast SDI:

1. Attributes of the tile
   * Zone/Category of the tile
   * Effective surface area of the Tile
2. Historical SDI data calculated in previous steps for each tile
3. Hierarchical SDI