

Introduction and Methodology

We are currently developing the algorithm with the use of city data from New Delhi, India. We continue to approach Aid organisations across the globe for data from conflict locations to use our model for refugees.

Currently, we select routes based on a prediction of danger levels at different locations. Rather than considering the concentration of crime in each area, we further make predictions based on a regression model that considers how recent each crime was.

Our algorithm has been divided into the parts that follow.

Data Collection and Organisation

View our base dataset here -

https://drive.google.com/file/d/10ZAaC0gUp8bV2g52_W0jDXws5Shy0_Xn/view?usp=sharing

We have organised, cleaned, and extended the dataset specifically for the purposes of our algorithm.

Each crime datapoint consists of its map coordinates, type, and year.

Available crimes have been divided into the following types -

- Murder
- Rape
- Robbery
- Theft

The dataset ranges from 2017 to 2021.

Region Classifier

Our map has been divided into regions of 1 sq km each, partitioning datapoints based on their coordinates. All crimes within a region have been considered in reaching a region-wise danger score for each year

Danger Scoring

The danger level for each region (D_i) has been calculated using the following equation* -

$$D_i = 14 * M_i + 7 * R_i + 5 * B_i + 0.5 * T_i$$

M_i - Number of Murders in that year

R_i - Number of Rapes in that year

B_i - Number of Robberies in that year

T_i - Number of Thefts in that year

Crime	Crime Weight
Murder	14
Rape	7
Robbery	5
Theft	0.5

*The crime weights have been assumed with thorough analysis based on minimum sentences for each crime type according to the Indian Penal Code.

Danger Prediction & Route Scoring

Each year's danger score is used through a linear regression model, to predict the current danger score for each region, based on past trend.

These scores are finally uses in selecting the optimal (safest) route with the lowest Route Danger Score (R_i)

$$R_i = \sum D_i * M_i$$

M_i = Number of metres of the route lying within that particular region

The route with the lowest score is displayed.

(If two routes have the same score, the one taking less time is shown)