Multi-Criteria Vulnerability Estimation Technique for Edges in a Network Graph

A Project Report

submitted by

MANOGNYA PARAM KOOLATH

in partial fulfilment of the requirements
for the award of the degree of

BACHELOR OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

June 2021

THESIS CERTIFICATE

This is to certify that the thesis titled **Multi-Criteria Vulnerability Estimation Technique for Edges in a Network Graph**, submitted by **Manognya Param Koolath**, to the Indian Institute of Technology, Madras, for the award of the degree of **Bachelor of Technology** , is a bonafide record of the research work done by her under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Usha Mohan
Research Guide
Professor
Dept. of Management Studies
IIT-Madras, 600 036

Prof. Kaushik Mitra
Research Co-Guide
Assistant Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Place: Chennai

Date: 18th June 2021

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere and heartfelt gratitude to my advisor **Dr. Usha Mohan**, Department of Management Studies, for the continuous support and guidance throughout the course of my B. Tech project, and for her patience, motivation, enthusiasm and immense knowledge.

My sincere thanks also goes to **Dr. Kaushik Mitra**, Department of Electrical Engineering for co-guiding me on this project, and for his time.

Besides my advisors, I wish to thank **Sweety Hansuwa**, Ph.D. student, Department of Management Studies for her encouragement, insightful comments and discussions.

I would like to thank my parents, and the institute faculty members for motivating me and being supportive throughout the work.

I also take this opportunity to thank my friends who have helped and co-operated throughout the course of this project with stimulating discussions and conversations.

ABSTRACT

KEYWORDS: Flood; AHP; TOPSIS; Regression; Clustering; Risk; Vulnerability

This project is aimed at designing and quantifying vulnerability of different streets and localities. The available vulnerability measures only consider the level of inundation from past experiences. But the actual risk/vulnerability would depend on a variety of other factors like the amount of impact the disaster has, resources at hand (government's) to save affected lives and property, and other environmental factors. We come up with a way to take in all these factors into consideration and to give us a vulnerability score, in particular for streets. We have trained the model using data available for Chennai, however, it is possible to reuse this method to other localities and districts as well.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

IITM	Indian Institute of Technology, Madras
OR	Operations Research
DOM	Disaster Operations Management
AHP	Analytic Hierarchy Process
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
NIS	Negative Ideal Solution
PIS	Positive Ideal Solution

CHAPTER 1

Introduction

Disasters are any unforeseen serious disruption which can overwhelm the systems in place and involves human, environmental and economic impacts that exceed the society's ability to cope on its own. Most disasters cannot be prevented, but their effects can be mitigated to a great extend as explained in Salamati and Kulatunga (2017). Disaster management aims at avoiding or even reducing the effects in case of a disaster, and at achieving a speedy recovery. Due to the uncertainty of disaster occurrence including various location-based scenarios which can arise, probability of occurrence, the complication in estimating the demand and supply, and the difficulty in identifying the resources available, operations management has become very popular in disaster management. The use of various tools and methods such as OR in disasters is referred to as Disaster Operations Management (DOM), and it has become very popular in the past few years.

There are four main stages or phases in the DOM life cycle, according to FEMA (2004). These are:

1. Mitigation
2. Preparedness
3. Response
4. Recovery

Mitigation includes those measures taken to avoid a disaster from happening, reduce its chance of happening, or diminish its wrecking effects. For example, vulnerability and risk assessment, construction of barriers, building protocols, etc come under the mitigation phase.

Preparedness refers to those plots and plans made to deal and respond in case of a disaster. This stage helps save lives by facilitating quick rescue operations. By procuring and positioning resources, response time during a disaster could be significantly lowered. Acquiring needed resources, equipment, vehicles, constructing shelter homes, recruiting and training a crew, etc come under the preparedness phase.

Response phase is how we respond immediately to a disaster. During this phase, all measures to keep affected people safe are adopted. It includes evacuation of vulnerable people and animals, search and rescue operations, supplying emergency medical and food kits, etc.

The recovery phase is the long term plans and methods to go back to the normal situation. Removal of debris, reconstruction of damaged infrastructure, financial assistance schemes, etc come under the recovery phase.

For any disaster, the first two phases, mitigation and preparedness, are very important to carry out an efficient disaster response and recovery, and thereby to protect and safeguard lives and property. We can observe from the past few years that to ascertain our goal of safety and low impact in Chennai, the preparedness and disaster management needs to be significantly enhanced.

Disasters seem to be those uncontrollable problems that test the ability of societies and countries to effectively safeguard their people and infrastructure, and to reduce the impacts on human lives and property loss, and to recover soon. The randomness in impacts and problems, and the uniqueness of events demand dynamic, real-time, effective and cost efficient solutions, thus making disaster management very fitting to be solved using OR tools and techniques, as detailed by Altay and Green (2006). Different OR methods have been tried out earlier as described by Hoyos *et al.* (2015) and Bayram (2016). Let's consider the location allocation problem: Given a geographical map of a region, how do we optimally allocate new facilities? As a preparation for any disaster, there are relief centers managed by state governments in our country. These relief centers are not disaster-particular as they are pretty generic and could be used during times of any disaster. For example, the nearest relief center itself might have been affected by the disaster. Now let's consider the problem where we need to allocate new relief centers to satisfy the needs of all disaster-affected (predicted) areas. It involves allocating relief centers to satisfy the demand and needs of vulnerable areas. Location allocation problems map areas to relief centers based on demand, supply and other factors. Since the problem should be solved at a planning stage, it would help to incorporate vulnerabilities of both the shelter locations and different streets. Hence we model the entire geographic area as a network graph, $G(V, E)$ where V (nodes) is the set of relief centers and demand locations and E is the set of 'edges', or the streets that connect the nodes.

The vulnerability of a street varies due to multiple reasons like the type of disaster, population of the area, disaster specific factors, available shelters in case of evacuation, etc. It is hence important to know the vulnerability of a particular location towards a particular disaster. Earthquakes and floods are the common disasters studied across the world, and since Chennai is prone to floods, we will focus more on floods as the type of disaster, unless explicitly specified otherwise.

1.1 Problem Statement

For the location allocation problem, we need vulnerability to perform any sort of optimization, and it should be measurable. It is therefore important to have a metric to quantify vulnerability for a given locality (can be reused for relief center locations as well). Similarly, to map localities with relief centers, we need the vulnerability of edges (streets) inter-connecting them.

There are multiple factors to consider while coming up with a vulnerability score for a region or a street, and can hence be formulated as a multi-criteria decision making problem. In this project, we came up with a generic method to calculate flood-vulnerability scores, and evaluate and test it for streets in the Chennai district of Tamil Nadu.

1.2 Available Data

The main source of data is the City Disaster Management Plan 2018 book published by the Greater Chennai Corporation which has details about different zones, wards, streets, relief centers and 2015 floods.

Chennai is divided into 15 zones and each zone is divided into multiple wards, making it 200 in total. We use 172 relief centers and 184 streets spanning the entire Chennai for this project.

The input dataset was a precompiled file with pairwise distances between those 184 streets and 172 relief centers.

Chapter 2 discusses in detail the methods we have used to come up with a vulnerability score, and in chapter 3, there is detailed explanation about how exactly we implemented each of it, and the results obtained.

CHAPTER 2

Methodology

Since the idea is to come up with a vulnerability score from different criteria, this would be a multi-criteria decision making (MCDM) problem. Hence we have used various MCDM methodologies to tackle the problem at hand:

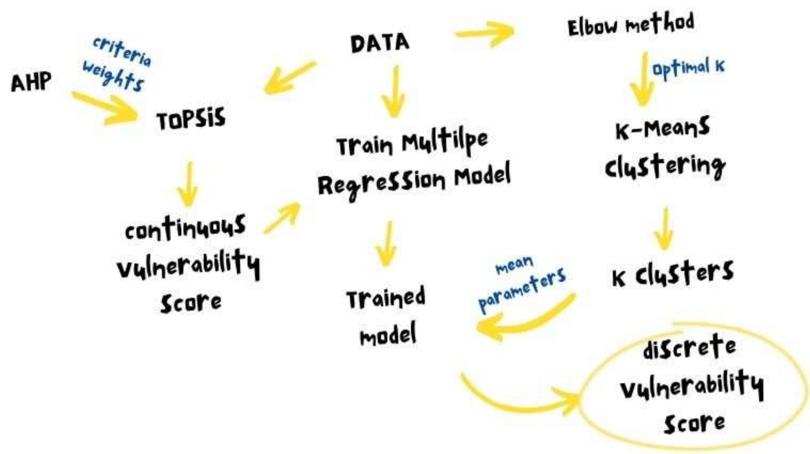

Figure 2.1: Methodology: Summary

2.1 Analytic Hierarchy Process

The Analytical Hierarchy Process (AHP) is a multi-criteria decision making method developed by Saaty (1990). It combines math and psychology to reach a numeric criteria weights for each criteria from pairwise comparison values. Rather than recommending a "universally correct" set of weights, it takes the user's intuitions and preferences into consideration, to get the values which best suits the problem at hand.

Let's say we have N criteria for which we have to obtain weights using the AHP algorithm, the following steps are followed:

1. We have to construct a pair-wise comparison matrix, C of dimensions NxN, where rows and columns represent each chosen criteria. $C_{i,j}$ is given a relative importance score of criteria i over criteria j. Hence, it is intuitive that $C_{i,i} = 1 \forall i \in \{1..N\}$ and $C_{i,j} = \frac{1}{C_{j,i}} \forall i, j \in \{1..N\}$.

   As directed by Saaty (1990), the relative importance score is:
   - 1 for equal importance
   - 3 for moderate importance
   - 5 for strong importance
   - 7 for very strong importance
   - 9 for extreme importance
   - 2,4,6,8 for intermediate importance
   - Inverse values accordingly

2. The next step is to calculate the normalize pairwise matrix, NPM. For each element $C_{i,j}$ in C, it is calculated as $N_{i,j} = \frac{C_{i,j}}{\sum_{i=1}^{N} C_{i,j}}$.

3. The criteria weights for each criteria $i \in \{1, ..N\}$ is given as $W_i = \frac{\sum_{j=1}^{N} NPM_{i,j}}{N}$.

4. Now that we have criteria weights, we have to calculate the consistency ratio (CR) with the obtained weights. We can accept the weights if CR < 0.1. To calculate CR:
   (a) Calculating weighted pairwise matrix WPM_{NxN}: $WPM_{i,j} = C_{i,j} * W_j$.
   (b) Calculating weighted sum for each criteria $i \in \{1, ..N\}$: $WS_i = \sum_{j=1}^{N} WPM_{i,j}$.
   (c) Next we calculate the ratio(R) of weighted sum to the calculated criteria weight for each criteria $i \in \{1, ..N\}$. $R_i = \frac{WS_i}{W_i}$.
   (d) λ_{max} is calculated by taking the average of all these N ratios obtained in the previous step.
   (e) Consistency index CI is calculated as $CI = \frac{\lambda_{max} - N}{N-1}$.
   (f) Finally consistency ratio, CR, is given by dividing CI by the random index for that particular N. Random index is the consistency index of the randomly generated pairwise comparison matrix.

2.2 Technique for Order of Preference by Similarity to Ideal Solution

This technique, abbreviated as TOPSIS is another multicriteria decision making method developed by García-Cascales and Lamata (2012). It is based on the logic that the chosen criteria should have high geometric distance from the negative ideal solution (NIS) and should be closer to the positive ideal solution (PIS). Since most criteria are of incongruous dimensions, normalization is essential before calculating the distances. We can develop a score based on these distances, which gives us a priority ordering based on the score.

2.3 Multiple Linear Regression

Linear regression is a statistical method to model a linear relation between reasons (independent variables) and results (dependent variable), as described by de A. Lima Neto *et al.* (2004), Uyanık and Güler (2013) and Helwig (2017). In linear regression, we mathematically model the dependent variable as a linear dependency of the independent variables and try to evaluate the suitable coefficients to reduce the error in prediction.

The model has the form:

$$y_i = b_0 + \sum_{j=1}^{p} b_j x_{ij} + e_i \tag{2.1}$$

for $i \in \{1, ..., n\}$ where

- $y_i \in \mathbf{R}$ is the response for the i-th data point.
- $b_0 \in \mathbf{R}$ is the regression intercept.
- $b_j \in \mathbf{R}$ is the j-th predictor's regression slope.
- $x_{ij} \in \mathbf{R}$ is the j-th predictor for i-th data point.
- e_i is a gaussian error term.

In multiple regression, $p > 1$. The idea is to estimate the unknown constants, $b_i \forall i \in \{0, 1, ..., p\}$. One common approach is to minimize ordinary least squares error, to estimate these constants. After estimation, the dependent variable is predicted as :

$$y = b_0 + \sum_{j=1}^{p} b_j x_j + e_i \tag{2.2}$$

2.4 K-Means Clustering

K-Means clustering is an unsupervised clustering algorithm which is quite popular, as explained by Mannor *et al.* (2011). The objective of this is to group data points that are similar and discover patterns. We define a number k which is the number of cluster centroids we'll have at the end of clustering. Every data point is allocated to one of these k clusters, and the idea is to ensure centroid stability of each cluster.

2.4.1 Elbow method

The elbow method is a way to determine the optimal number of clusters in a given dataset to be clustered. The idea is to plot 'Distortion' against candidate values of K. Distortion is calculated as the average of sum of squared distances from each data point to its respective cluster center. The optimal K is picked as the candidate K at the elbow of the curve (The point where the curve changes its slope visibly from high to low or from low to high). The method is explained in detail by Marutho *et al.* (2018) and Umargono *et al.* (2019).

In the next chapter, we'll go through how each of these methods were executed and the results obtained.

CHAPTER 3

Execution and Results

An AHP-TOPSIS model has been introduced in Jena and Pradhan (2020) to study and improve the earthquake risk assessment. We apply a similar method to improve vulnerability evaluation of floods.

3.1 Identifying Parameters

Many factors can contribute to losses during a disaster, the major factors being environmental and due to population and occupancy. We have determined six major factors which can contribute to making a region more or less vulnerable in terms of floods. As we work with streets mainly, we have the following parameters with respect to any street:

- P1: Weighted sum of capacities of relief centers within a distance of 1 km from the street.
- P2: Weighted sum of capacities of relief centers at a distance less than 3 km, but greater than or equal to 1 km from the street.
- P3: Population of the ward where the street is located.
- P4: Number of water bodies (pond, lakes, etc) in the same ward as the street.
- P5: Sum of lengths of canals and rivers in the same zone (in km)
- P6: Score based on water inundation in 2015 floods.

For P1 and P2, weighted sum is considered to make sure that the location of the relief center is not very vulnerable to floods. The weight is based on the previous (2015) water inundation score so that if the relief center had high inundation, the weightage of its capacity will be low.

Note that P4, P5, P6 are natural and environmental factors which would affect the flood vulnerability of a street. However, P1, P2, P3 are those factors which can be used to determine the loss incurred, as supply-demand would be analogous to these parameters. For P3, if we could get the streetwise population, that score would give a more precise estimate, however since the data was unavailable for Chennai, we are using ward-wise population as an estimate in this work.

3.2 Data Collection, Compilation and Processing

The data used is mainly obtained from Coorporation (2018).

3.2.1 P6

In the Coorporation (2018), there are four levels of inundation specified and localities are classified into one among these four. We give a corresponding score for each:

- Above 5 feet: Score is 7
- 3 - 5 feet: Score is 5
- 2 - 3 feet: Score is 3
- Less than 2 feet: Score is 1

Note that this score can give an idea of flood vulnerability based only on the level of inundation during 2015 floods. For each street, we consider the ward's inundation to be the same as that of that street. If more localities within a ward had different inundation levels, to avoid the effect of outliers in any street's data, we take an average of all inundation scores in that ward. Whenever wardwise data was not available, we used zonewise data.

3.2.2 P1 and P2

We have the capacities of all the 172 relief centers from the Chennai Government's document. We also have the inundation score from P6. Note that this score will be between 1 and 7 where 1 represents low inundation and 7 represents the maximum. The weight for relief center capacity is calculated using:

$$w = \frac{6.7 - 0.7IS}{6} \tag{3.1}$$

where w is the weight for a location with ward inundation score of IS. Note that the weight comes out to be 1 if $IS = 1$ and 0.3 if $IS = 7$. Now we have the weight and capacities of each relief center. We also have the pairwise distance matrix between each street and relief center. From these, we can easily calculate P1 and P2 using a python code.

3.2.3 P3 and P4

The population of all wards, and the number of water bodies in each ward are directly available in the document Coorporation (2018).

3.2.4 P5

The lengths of all canals and rivers along with their lengths is available in the document Coorporation (2018). We added the lengths in each zones to get zonewise values of P5.

3.3 Weight Calculation using AHP

3.3.1 Pairwise Comparison Matrix

As mentioned in the previous section about AHP, we start by defining a pairwise comparison matrix.

	P1	**P2**	**P3**	**P4**	**P5**	**P6**
P1	1	3	0.7	2	3	2
P2	0.3333333333	1	0.3333333333	0.5	1	0.5
P3	1.428571429	3	1	2	3	2
P4	0.5	2	0.5	1	2	0.5
P5	0.3333333333	1	0.3333333333	0.5	1	0.3333333333
P6	0.5	2	0.5	2	3	1
Sum	4.095238095	12	3.366666667	8	13	6.333333333

Table 3.1: AHP: Pairwise Comparison Matrix

Those cell values in light grey are defined by us and the rest follows due to the properties of pairwise comparison matrices. The logic behind the ones defined by us are:

Group 1: P1, P2, and P3

Since P1 is for relief centers within 1 km and P2 factors in those relief centers upto 3 km distance, it is only appropriate to allocate a higher priority to P1 than P2. It's hence logical to allocate P1 a priority 3 times that of P2. P3 is a measure of demand and P2 and P1 would measure the supply. Hence, it makes sense that P1 has 0.7 times the priority of P3, and P3 has a priority 3 times more than that of P2.

Group 2: P4, P5, and P6

Note that P4, P5, and P6 are environmental factors outside human control. The number of water bodies (P4) definitely affects floods as the chances of them filling up during rains are high. The previous level of inundation (P6) gives us a measure of elevation of that particular area. Also, length of canals in a zone (P5) is important, but however, since it is not very specific to the streets being considered, it should have a lower priority than P4 and P5. Hence it makes sense for P4 to have twice the priority of P5, and half the priority of P6. It is also right for P6 to have 3 times the priority of P5.

Group 1 and Group 2

Since we want to have a vulnerability score which measures impact to lives, property and society, it is appropriate to assign a higher priority to P1, P2, and P3 (Group 1) when compared to P4, P5, and P6 (Group 2). Hence we assign P1 to have twice the priority of P4 and P6, and thrice the priority of P5. P2 has a priority equal to P5 as both are not very specific to a particular street. P4 and P6 is assigned to have a priority double that of P2. Just like P1, we assign P3 to have twice the priority of P4 and P6, and thrice the priority of P5.

In order to accept these pairwise priorities, we have to make sure that the consistency ratio after calculating criteria weights is acceptable.

3.3.2 Normalized Pairwise Matrix and Criteria Weights

The next step is to calculate the normalized pairwise matrix (explained in section 2.1.1), and subsequently, the criteria weights. The normalized pairwise matrix obtained for our problem is in Table 3.2, and the criteria weights in Table 3.3.

	P1	P2	P3	P4	P5	P6
P1	0.2442	0.25	0.2079	0.25	0.2308	0.3158
P2	0.0814	0.0833	0.099	0.0625	0.0769	0.0789
P3	0.3488	0.25	0.297	0.25	0.2308	0.3158
P4	0.1221	0.1667	0.1485	0.125	0.1538	0.0789
P5	0.0814	0.0833	0.099	0.0625	0.0769	0.0526
P6	0.1221	0.1667	0.1485	0.25	0.2308	0.1579

Table 3.2: AHP: Normalized Pairwise Matrix

	Criteria Weights
P1	0.2498
P2	0.0804
P3	0.2821
P4	0.1325
P5	0.076
P6	0.1793

Table 3.3: AHP: Criteria Weights

3.3.3 Consistency Ratio

Now that we have the criteria weights, we have to calculate the consistency of the assigned pairwise priorities (explained in section 2.1.1).

	P1	P2	P3	P4	P5	P6	Weighted sum
P1	0.2498	0.2411	0.1974	0.265	0.2279	0.3586	1.5398
P2	0.0833	0.0804	0.094	0.0663	0.076	0.0897	0.4895
P3	0.3568	0.2411	0.2821	0.265	0.2279	0.3586	1.7315
P4	0.1249	0.1607	0.141	0.1325	0.1519	0.0897	0.8007
P5	0.0833	0.0804	0.094	0.0663	0.076	0.0598	0.4596
P6	0.1249	0.1607	0.141	0.265	0.2279	0.1793	1.0989

Table 3.4: AHP: Weighted Pairwise matrix, and weighted sums

Applying calculations to table 3.5,

$$\lambda_{max} = 6.1028 \tag{3.2}$$

	Weighted sum	Criteria Weights	Ratio
P1	1.5398	0.2498	6.1649
P2	0.4895	0.0804	6.0922
P3	1.7315	0.2821	6.1386
P4	0.8007	0.1325	6.0427
P5	0.4596	0.076	6.0505
P6	1.0989	0.1793	6.1279

Table 3.5: AHP: Weighted sums and ratios to criteria weights

$$CI = 0.0206 \tag{3.3}$$

$$RI = 1.24 \tag{3.4}$$

$$CR = 0.0166 \tag{3.5}$$

Since $CR < 0.1$, our calculations and weights are consistent. Hence we accept the criteria weights as per Table 3.3.

3.4 TOPSIS using weights

For TOPSIS, we use the criteria weights which we obtained from AHP. We have compiled data as per section 3.2. This results in a matrix, X_{MXN} where each row represents data for a particular street and parameters are organized column-wise. In our case, $M = 184$ and $N = 6$ since we have 184 streets across Chennai under consideration. The first step is to calculate the normalized decision matrix:

$$\overline{X_{ij}} = \frac{X_{ij} * W_j}{\sqrt{\sum_{j=1}^{n} X_{ij}^2}} \quad \forall i \in \{1, ..., M\}, \quad j \in \{1, ..., 6\} \tag{3.6}$$

where W_j is the criteria weight calculated using AHP for j-th criteria.

The next step is to calculate the ideal best and ideal worst for each criteria from values in the normalized decision matrix. Since in our case, we want a high score if vulnerability is high.

If j-th criteria is a beneficial criteria, then:

$$X_j^+ = \max_i \{\overline{X_{ij}}\}, \quad X_j^- = \min_i \{\overline{X_{ij}}\}, \quad i = 1, 2, \ldots, 184 \tag{3.7}$$

else, j-th criteria is a cost criteria, in which case:

$$X_j^+ = \min_i \{\overline{X_{ij}}\}, \quad X_j^- = \max_i \{\overline{X_{ij}}\}, \quad i = 1, 2, \ldots, 184 \tag{3.8}$$

Accordingly, our ideal best and worst values for all criteria are in Table 3.6.

Ideal best (X_j^+)	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**
Ideal best	0	0	0.0498	0.0465	0.0111	0.032
Ideal worst (X_j^-)	0.0694	0.0172	0.0021	0	0.0012	0.0046

Table 3.6: TOPSIS: Ideal Best and Ideal Worst values for different criteria

The next step is to calculate Euclidean distance for each street from ideal best and ideal worst data points.

Euclidean distance from ideal best:

$$S_i^+ = \{\sum_{j=1}^{6} (\overline{X_{ij}} - X_j^+)^2\}^{0.5} \quad \forall i \in \{0, ..., 184\} \tag{3.9}$$

Euclidean distance from ideal worst:

$$S_i^- = \{\sum_{j=1}^{6} (\overline{X_{ij}} - X_j^-)^2\}^{0.5} \quad \forall i \in \{0, ..., 184\} \tag{3.10}$$

The next step is to calculate performance score of the vulnerability for each street, which could be used as the vulnerability score of that street. It is given by:

$$V_i = \frac{S_i^-}{S_i^- + S_i^+} \tag{3.11}$$

The intuition here is that vulnerability is high if distance from ideal worst is low and ideal best is high.

3.5 Multiple Regression

One issue with TOPSIS is that as we have more data points, we'll have to calculate the ideal best and worst every time and calculate the performance score. This could be eliminated by training a regression model using the available scores calculated using TOPSIS. In which case, if we have a new streets with parameters, we can just estimate its vulnerability score using the trained regression model.

3.5.1 Data

We have data of 184 streets: The 6 parameters are the independent variables and the TOPSIS vulnerability score is the dependent variable.

3.5.2 Training

Since we have only 184 data points, our ability to train complex models is limited and hence we're using a linear regression model. The *sklearn* library was used for training and testing. The dataset was split into training and testing set in the ratio 4:1. The loss function used was root mean squared error.

3.5.3 Results

We achieved an accuracy of 98.77%. The predicted vulnerability scores after training are as shown in Fig 3.1.

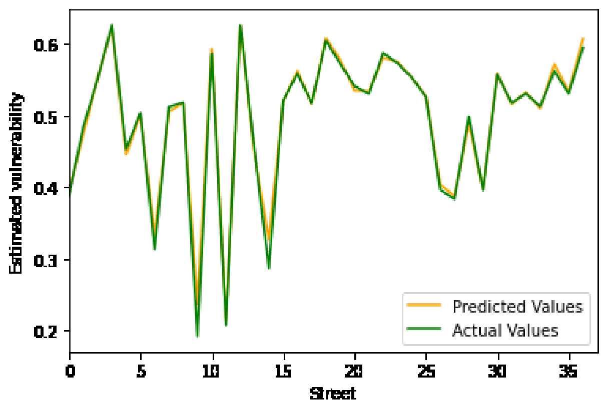

Figure 3.1: Linear Regression: Estimated vulnerability after training

We can see that the predictions of our regression model is pretty close to our TOPSIS values.

3.6 K-Means Clustering

The previous method gives us a continuous vulnerability score. If we want to have discrete vulnerability scores, clustering is the way to go (since it's unsupervised). The idea is to split the 184 streets into K clusters, and each cluster would be given a vulnerability score based on the prediction from our regression model.

3.6.1 Elbow Method for Optimal K

As explained in section 2.1.4, elbow method gives us the optimal value of K. On plotting distortion against candidate values of K, we get the following graph:

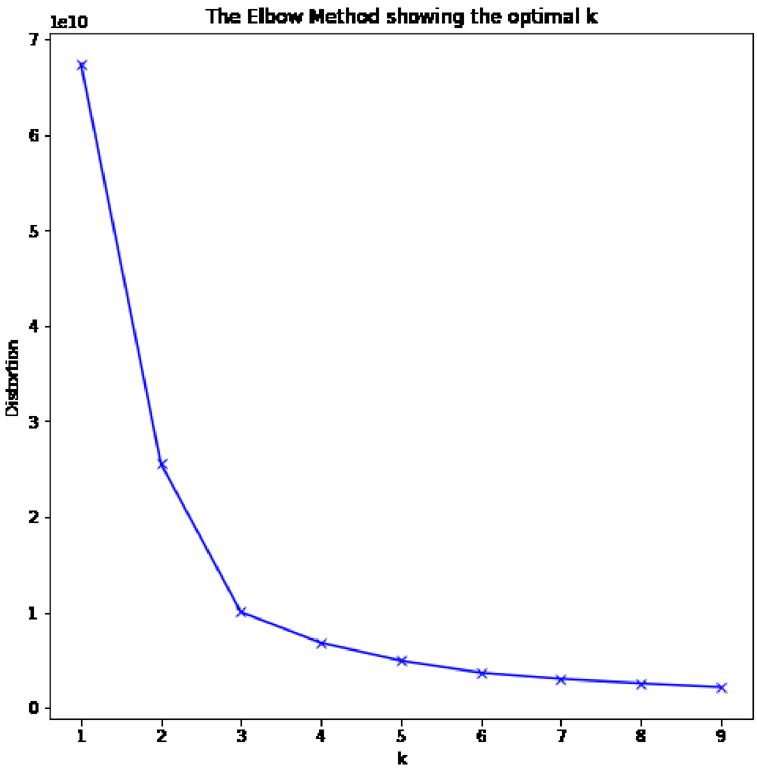

Figure 3.2: Clustering: Elbow method

It is hence clear that K=3 is the optimal number of clusters.

3.6.2 Clustering

We use the *sklearn* library to perform clustering. We obtained 3 clusters and their mean parameters as:

Cluster Id	**Central values**						**Predicted Vulnerability Score**
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	
0	745.949	3888.451	41110.019	0.657	12.691	1.884	0.497
1	922.021	2233.988	82468.100	1.05	8.935	4.472	0.574
2	463.807	1418.292	20614.196	2.643	12.374	2.161	0.515

Table 3.7: Clustering: Mean parameters

3.7 Hybrid Approach for discrete vulnerability score

The vulnerability is predicted for the central parameters of each cluster using the regression model we trained earlier. We have 3 vulnerability values as shown in Table 3.7. They are ranked in ascending order and assigned low vulnerability, medium vulnerability and highly vulnerable clusters. All streets are classified into one of these three based on the cluster they belong.

CHAPTER 4

Results and Conclusions

This work showed us that as many as 20 streets out of 184 in Chennai are highly vulnerable, and 56 are at a medium vulnerability level to floods (Table 4.1). Given that, Chennai has a decent history of floods, these should be the streets of focus to the government. Officials can increase relief centers in their vicinity, where the vulnerability is low and try to move these streets to a different vulnerability category. The government should take a look into the critical condition of infrastructure and populated areas where vulnerability reduction can be improved.

Category	Count of Streets
High Vulnerability	20
Medium Vulnerability	56
Low Vulnerability	108

Table 4.1: Summary of Results

There are a few limitations found in this research which could affect the accuracy of the proposed model. Lack of street-specific data like population, and inundation might have reduced the accuracy of the proposed model.

The method used here provides us with practical information on risk assessment of floods in Chennai. This method could be extended to any other street or locality in any other area.

APPENDIX A

Complete Results from Clustering

The final results after clustering, showing vulnerability levels of the 184 streets considered is shown in the table below:

Street	Vulnerability level	Street	Vulnerability level
Street1	Medium Vulnerability	Street93	Low Vulnerability
Street2	Medium Vulnerability	Street94	Low Vulnerability
Street3	Medium Vulnerability	Street95	Low Vulnerability
Street4	Medium Vulnerability	Street96	Low Vulnerability
Street5	Medium Vulnerability	Street97	Low Vulnerability
Street6	Medium Vulnerability	Street98	Low Vulnerability
Street7	Medium Vulnerability	Street99	Low Vulnerability
Street8	Medium Vulnerability	Street100	Low Vulnerability
Street9	Medium Vulnerability	Street101	Low Vulnerability
Street10	Medium Vulnerability	Street102	Low Vulnerability
Street11	Medium Vulnerability	Street103	Low Vulnerability
Street12	Medium Vulnerability	Street104	Low Vulnerability
Street13	Medium Vulnerability	Street105	Low Vulnerability
Street14	Medium Vulnerability	Street106	Low Vulnerability
Street15	Medium Vulnerability	Street107	Low Vulnerability
Street16	Medium Vulnerability	Street108	Low Vulnerability
Street17	Medium Vulnerability	Street109	Low Vulnerability
Street18	Low Vulnerability	Street110	Low Vulnerability
Street19	Low Vulnerability	Street111	Low Vulnerability
Street20	Low Vulnerability	Street112	Low Vulnerability
Street21	Medium Vulnerability	Street113	Low Vulnerability
Street22	Medium Vulnerability	Street114	Low Vulnerability
Street23	Medium Vulnerability	Street115	Low Vulnerability
Street24	Medium Vulnerability	Street116	Medium Vulnerability

Street25	Medium Vulnerability	Street117	Medium Vulnerability
Street26	Medium Vulnerability	Street118	Low Vulnerability
Street27	Medium Vulnerability	Street119	Low Vulnerability
Street28	Low Vulnerability	Street120	Medium Vulnerability
Street29	Low Vulnerability	Street121	Medium Vulnerability
Street30	Low Vulnerability	Street122	Medium Vulnerability
Street31	Low Vulnerability	Street123	Medium Vulnerability
Street32	High Vulnerability	Street124	Low Vulnerability
Street33	Low Vulnerability	Street125	Low Vulnerability
Street34	Low Vulnerability	Street126	Low Vulnerability
Street35	Low Vulnerability	Street127	Low Vulnerability
Street36	Low Vulnerability	Street128	Low Vulnerability
Street37	Low Vulnerability	Street129	Low Vulnerability
Street38	Low Vulnerability	Street130	Low Vulnerability
Street39	Low Vulnerability	Street131	Low Vulnerability
Street40	Low Vulnerability	Street132	Low Vulnerability
Street41	Low Vulnerability	Street133	Low Vulnerability
Street42	Low Vulnerability	Street134	Medium Vulnerability
Street43	Low Vulnerability	Street135	Medium Vulnerability
Street44	Low Vulnerability	Street136	Medium Vulnerability
Street45	Low Vulnerability	Street137	Medium Vulnerability
Street46	Low Vulnerability	Street138	Medium Vulnerability
Street47	Low Vulnerability	Street139	Medium Vulnerability
Street48	Low Vulnerability	Street140	Medium Vulnerability
Street49	Low Vulnerability	Street141	Medium Vulnerability
Street50	Low Vulnerability	Street142	Medium Vulnerability
Street51	Low Vulnerability	Street143	Medium Vulnerability
Street52	Low Vulnerability	Street144	High Vulnerability
Street53	Low Vulnerability	Street145	High Vulnerability
Street54	Low Vulnerability	Street146	High Vulnerability
Street55	Low Vulnerability	Street147	High Vulnerability
Street56	Low Vulnerability	Street148	High Vulnerability
Street57	Low Vulnerability	Street149	High Vulnerability

Street58	High Vulnerability	Street150	High Vulnerability
Street59	Low Vulnerability	Street151	High Vulnerability
Street60	Low Vulnerability	Street152	High Vulnerability
Street61	Low Vulnerability	Street153	High Vulnerability
Street62	Low Vulnerability	Street154	High Vulnerability
Street63	Low Vulnerability	Street155	High Vulnerability
Street64	Low Vulnerability	Street156	High Vulnerability
Street65	Low Vulnerability	Street157	High Vulnerability
Street66	Low Vulnerability	Street158	High Vulnerability
Street67	Low Vulnerability	Street159	High Vulnerability
Street68	Low Vulnerability	Street160	High Vulnerability
Street69	Low Vulnerability	Street161	High Vulnerability
Street70	Low Vulnerability	Street162	Low Vulnerability
Street71	Low Vulnerability	Street163	Low Vulnerability
Street72	Low Vulnerability	Street164	Low Vulnerability
Street73	Low Vulnerability	Street165	Low Vulnerability
Street74	Low Vulnerability	Street166	Low Vulnerability
Street75	Low Vulnerability	Street167	Low Vulnerability
Street76	Low Vulnerability	Street168	Medium Vulnerability
Street77	Low Vulnerability	Street169	Medium Vulnerability
Street78	Low Vulnerability	Street170	Low Vulnerability
Street79	Medium Vulnerability	Street171	Low Vulnerability
Street80	Medium Vulnerability	Street172	Medium Vulnerability
Street81	Medium Vulnerability	Street173	Low Vulnerability
Street82	Low Vulnerability	Street174	Low Vulnerability
Street83	Low Vulnerability	Street175	Low Vulnerability
Street84	Low Vulnerability	Street176	Medium Vulnerability
Street85	Low Vulnerability	Street177	Medium Vulnerability
Street86	Low Vulnerability	Street178	Medium Vulnerability
Street87	Low Vulnerability	Street179	Medium Vulnerability
Street88	Low Vulnerability	Street180	Medium Vulnerability
Street89	Medium Vulnerability	Street181	Medium Vulnerability
Street90	Low Vulnerability	Street182	Medium Vulnerability

Street91	Low Vulnerability	Street183	Medium Vulnerability
Street92	Low Vulnerability	Street184	Medium Vulnerability

Table A.1: Discrete Vulnerability Scores

APPENDIX B

Code and Supplementary Data

The compiled and processed data, calculations and predictions can be found at Data (2021). The code which was used can be found at Code (2021).

Both are added to the references section as well.

REFERENCES

1. **Altay, N.** and **W. G. Green** (2006). Or/ms research in disaster operations management. *European Journal of Operational Research*, **175**. URL http://works.bepress.com/nezih_altay/3/.

2. **Bayram, V.** (2016). Optimization models for large scale network evacuation planning and management: A literature review. *Surveys in Operations Research and Management Science*, **21**(2), 63–84. ISSN 1876-7354. URL https://www.sciencedirect.com/science/article/pii/S1876735416300125.

3. **Code** (2021). Btp. URL https://colab.research.google.com/drive/1Qz0XlYztungALHJCTaq-piiOtksnuVcA?usp=sharing.

4. **Coorporation, C.** (2018). City disaster management plan. URL https://www.chennaicorporation.gov.in/images/CDMP%20Book%20Wrapper%20Full%20Book%20(%20English).pdf.

5. **Data** (2021). Btp. URL https://docs.google.com/spreadsheets/d/1zwmXgeu_UWs89Hnoa6bYO-ylLgO10p5dPxLPCdVU8Rg/edit?usp=sharing.

6. **de A. Lima Neto, E.**, **F. A. T. de Carvalho**, and **C. P. Tenorio**, Univariate and multivariate linear regression methods to predict interval-valued features. *In Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2004, 526–537. URL https://doi.org/10.1007/978-3-540-30549-1_46.

7. **FEMA** (2004). The four phases of disaster. fema – emergency management institute. URL https://training.fema.gov/emiweb/downloads/is111_unit%204.pdf.

8. **García-Cascales, M. S.** and **M. T. Lamata** (2012). On rank reversal and topsis method. *Mathematical and Computer Modelling*, **56**(5), 123–132. ISSN 0895-7177. URL https://www.sciencedirect.com/science/article/pii/S0895717711007850.

9. **Helwig, N. E.** (2017). Multivariate linear regression. URL http://users.stat.umn.edu/~helwig/notes/mvlr-Notes.pdf.

10. **Hoyos, M. C.**, **R. S. Morales**, and **R. Akhavan-Tabatabaei** (2015). Or models with stochastic components in disaster operations management: A literature survey. *Computers Industrial Engineering*, **82**, 183–197. ISSN 0360-8352. URL https://www.sciencedirect.com/science/article/pii/S0360835214004136.

11. **Jena, R.** and **B. Pradhan** (2020). Integrated ann-cross-validation and ahp-topsis model to improve earthquake risk assessment. *International Journal of Disaster Risk Reduction*, **50**, 101723. ISSN 2212-4209. URL https://www.sciencedirect.com/science/article/pii/S2212420919317273.

12. **Mannor, S.**, **X. Jin**, **J. Han**, **X. Jin**, **J. Han**, **X. Jin**, **J. Han**, and **X. Zhang**, K-means clustering. *In Encyclopedia of Machine Learning*. Springer US, 2011, 563–564. URL `https://doi.org/10.1007/978-0-387-30164-8_425`.

13. **Marutho, D.**, **S. Hendra Handaka**, **E. Wijaya**, and **Muljono**, The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. *In 2018 International Seminar on Application for Technology of Information and Communication*. 2018.

14. **Saaty, T. L.** (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, **48**(1), 9–26. ISSN 0377-2217. URL `https://www.sciencedirect.com/science/article/pii/037722179090057I`. Desicion making by the analytic hierarchy process: Theory and applications.

15. **Salamati, P.** and **U. Kulatunga**, The importance of disaster management impact of natural disasters on hospitals. 2017.

16. **Umargono, E.**, **J. E. Suseno**, and **V. G. S. K.**, K-means clustering optimization using the elbow method and early centroid determination based-on mean and median. *In Proceedings of the International Conferences on Information System and Technology - Volume 1: CONRIST,*. INSTICC, SciTePress, 2019. ISBN 978-989-758-453-4.

17. **Uyanık, G. K.** and **N. Güler** (2013). A study on multiple linear regression analysis. *Procedia - Social and Behavioral Sciences*, **106**, 234–240. ISSN 1877-0428. URL `https://www.sciencedirect.com/science/article/pii/S1877042813046429`. 4th International Conference on New Horizons in Education.