**Clustering the Effects of Disasters on Social Unrest**

Jimmy Erickson

**Abstract**

**1. Introduction**

Understanding the role of a disaster and modeling its effects on social unrest is a complicated process that requires a multi-faceted approach. After a disaster occurs, a government’s ability to efficiently go in and provide aid to the affected area is essential to the survival and recovery of the population affected by the disaster. When a government’s actions towards disaster recovery or prevention are not perceived as efficient or effective, a disaster could directly trigger social unrest (Olson, 1998). That is, social unrest activities such as protests could follow immediately after the occurrence of a disaster; this was observed in Beirut, August 2020. An explosion of poorly-stored material resulted in massive destruction, as well as wide-spread demonstrations that led to a number of officials stepping down (BBC, 2020). On the other hand, while a disaster itself might not trigger social unrest in the short term, there are situations where a disaster or a series of disasters could alter various socio-economic factors in the affected population as fuels for grievances and dissatisfaction against certain policy or government (Becker, 2016). With a growing number of disasters in the world today, understanding different aspects of them is growing increasingly important. The effects of disasters are farther-reaching than simply what buildings they destroy. Understanding the features that go into the social effect of a disaster can help in future research into understanding the social effects of disasters. During our research into the topic of disasters and social unrest, there were no papers specifically focused on the social effect of a disaster as either a fuel or trigger for social unrest.Our solution to satisfy this need is to study possible relations between disasters and social unrest through clustering analysis. We will be utilizing data about disaster events, data about social unrest events, and other additional datasets that are found to be helpful. This paper will not be focusing on innovation in the field of Computer Science. Instead, we will be focusing on studying the relationship between disasters and social unrest while finding new ways to apply data science techniques to our proposed problem. We will be demonstrating fairly simple data science techniques in order to analyze and model the data, as well as writing scripts in order to assist in managing the data, but our initial plan does not include computer science innovation.This paper’s main goal is to contribute to the field of disaster research by approaching disasters from the context of unrest. And to use clustering analysis to study these events so as to support our research and give a starting point for future research of this topic. We will be generating datasets that are a combination of multiple datasets including GDELT, DesInventar, and other datasets such as infrastructure, demographic wealth inequality, and other datasets that are determined to be relevant to this research.

Overview of the rest of the paper/thesis

***I assume we have to wait until more of the paper is written in order to do the preview part***

Questions:

1. Do we write this with the knowledge we have at the beginning of the project or at the end of the project (i.e. mentioning the datasets we use)

**2. Related Work and Background**

**3. Methodology**

**NEED AN OVERVIEW OF THE SUBSECTIONS**

**3.1. Data Sources**

The data sources we are using are the DesInventar and GDELT databases. These databases were chosen because they filled various criteria that made them optimal for our research.

The DesInventar database is a project sponsored by the United Nations to track disaster events around the world and assemble a highly accurate dataset of these disasters. DesInventar partners with universities in each region they are tracking in order to ensure accurate data with minimal undercounting or overcounting. The DesInventar dataset we have for India includes the three states of Uttarakhand, Tamil Nadu, and Orissa for the year range 1985 - 2012. The attributes provided for events in these regions include the following: state, district, block, date, type, sub-type, cause, source, deaths, injuries, and many additional metrics we will not be using for the purpose of this project.

The GDELT database is a project to map all significant events in the world, including social unrest events for 1985-2012, which are the events we will be using from this source and appending to our events from DesInventar. Our GDELT dataset has the following fields: date, source (original actor), target (target actor), CAMEOcode (defines the exact type of event), NumEvents (if there are multiple smaller events that act as one), latitude and longitude.

We did not use ACCLED or ICEWS. The reasoning for this is that the DesInventar database has data consistently until 2012, when their partnerships in India expired. This caused a problem because the ACCLED protest dataset had dates starting in 2012.

**3.2. Data Preparation**

The first step in preparing the data was to combine the three separate datasets we had for DesInventar, which were the three states of India that were supported by this project. Those states are Orissa (Odisha), Tamil Nadu, and Uttarakhand. The DesInventar events are fairly evenly spread out within the states, though not evenly spread out between the states with Orissa having 9,081 events, Tamil Nadu having 15,535 events, and Uttarakhand having 3,879 events. To continue preparing the DesInventar data, we set about getting a latitude and longitude value for each of the DesInventar entries. Since we were given the state and district in which a disaster occurred, we fed that information into the Bing maps API which then returned latitude and longitude values for each of our disasters. For combining them, we went about the process of removing excess columns that we did not want to test out initially in our various modeling methods that are discussed in Section **3.3**. We then formatted each of the columns to have the correct data type so as to be consistent with the events from GDELT (i.e. specifying double/integer, string/date).

To prepare the GDELT data, the *first step was to filter the events based on location*. We used the state shapefiles for India and QGIS to filter the events down to the relevant states (Orissa, Tamil Nadu, and Uttarakhand) that are included in our DesInventar data. The *second step was to filter the GDELT events by event type*. Filtering by event type was done by studying the GDELT documentation on CAMEO codes to determine the events to keep. It was decided to keep events that included *humanitarian aid* or *social unrest events* (see Table 1). The decision to keep humanitarian aid events was made because research into disaster events and social unrest showed that government response to a disaster has a significant impact on potential unrest.

|  |  |
| --- | --- |
| CAMEO Code | Description |
| 073 | **Provide Humanitarian Aid-**  Extend, provide humanitarian aid, mainly in the form of emergency assistance. |
| 140 | **Engage in political dissent, not specified in other codes-**  All civilian demonstrations and other collective actions carried out as protests against the target actor not otherwise specified. |
| 141 | **Demonstrate or rally, not specified below-**  Dissent collectively, publicly show negative feelings or opinions; rally, gather to protest a policy, action, or actor(s). |
| 1411 | **Demonstrate or rally for leadership change-**  Dissent collectively, gather, or rally demanding leadership change. |
| 1412 | **Demonstrate or rally for policy change-**  Dissent collectively, gather, or rally demanding policy change. |
| 1413 | **Demonstrate for rights-**  Dissent collectively, gather, or rally demanding political, social, economic,  or other rights. |
| 1414 | **Demonstrate for change in institutions, regime-**  Dissent collectively, gather, or rally demanding major institutional, constitutional, or regime change. |
| 142 | **Conduct hunger strike, not specified below-**  Protest by refusing to eat until certain demands are met, not further specified. |
| 1421 | **Conduct hunger strike for leadership change-**  Refuse to eat until demands for leadership change are met. |
| 1422 | **Conduct hunger strike for policy change**-  Refuse to eat until demands for policy reform are met. |
| 1423 | **Conduct hunger strike for rights-**  Refuse to eat until demands for political, social, economic, or other rights are met. |
| 1424 | **Conduct hunger strike for change in institutions, regime-**  Description Refuse to eat until demands for major institutional, constitutional, or regime change. |
| 143 | **Conduct strike or boycott, not specified below-**  Protest by refusing to work or cooperate until certain demands are met, not specified further. |
| 1431 | **Conduct strike or boycott for leadership change-**  Refuse to work or cooperate until demands for leadership change are met. |
| 1432 | **Conduct strike or boycott for policy change-**  Refuse to work or cooperate until demands for policy reform are met. |
| 1433 | **Conduct strike or boycott for rights-**  Description Refuse to work or cooperate until demands for political, social, economic, or other rights are met. |
| 1434 | **Conduct strike or boycott for change in institutions, regime-**  Description Refuse to work or cooperate until demands for major institutional, constitutional, or regime change. |
| 144 | **Obstruct passage, block, not specified below-**  Description Protest by blocking entry and/or exit into building or area, not otherwise specified. |
| 1441 | **Obstruct passage to demand leadership change-**  Obstruct passage, block entry/exit to demand leadership change. |
| 1442 | **Obstruct passage to demand policy change-**  Obstruct passage, block entry/exit to demand policy reform. |
| 1443 | **Obstruct passage to demand rights-**  Obstruct passage, block entry/exit to demand political, social, economic, or other rights. |
| 1444 | **Obstruct passage to demand change in institutions, regime-**  Description Obstruct passage, block entry/exit to demand major institutional, constitutional, or regime change. |
| 145 | **Protest violently, riot, not specified below-**  Protest forcefully, in a potentially destructive manner, not further specified. |
| 1451 | **Engage in violent protest for leadership change-**  Protest forcefully, in a potentially destructive manner, to demand leadership change. |
| 1452 | **Engage in violent protest to demand policy change-**  Protest forcefully, in a potentially destructive manner, to demand policy reform. |
| 1453 | **Engage in violent protest to demand rights-**  Protest forcefully, in a potentially destructive manner, to demand political, social, economic, or other rights. |
| 1454 | **Engage in violent protest to demand change in institutions, regime-**  Protest forcefully, in a potentially destructive manner, to demand major institutional, constitutional, or regime change. |
| 1323 | **Threaten to impose curfew-**  Threaten to enforce a deadline beyond which inhabitants of an area are not  permitted to be on the streets or in public places. |
| 1324 | **Threaten to impose state of emergency or martial law-**  Threaten with suspending certain given rights or the whole constitution by imposing state of emergency or military rule. |
| 153 | **Mobilize or increase police power-**  Increase the number of military personnel and/or weapons |
| 1663 | **Expel or withdraw aid agencies-**  Terminate the presence of aid agencies or other non-governmental organizations helping civilians. |
| 1723 | **Impose Curfew-**  Set a deadline beyond which inhabitants of an area are not permitted to be on the streets or in public places. |
| 1724 | **Impose state of emergency or martial law-**  Suspend normal constitutional rights and provisions by installing state of emergency or military rule. |
| 173 | **Arrest, Detain-**  Legal or extrajudicial arrests, detentions, or imprisonments. |
| 175 | **Use repression-**  Actively repress collective actions of dissent by forcing subjugation through crowd control tactics, arrests, etc. |

**Table 1. CAMEO Codes and Descriptions**

Once the data from our two sources was prepared, the data was combined into a single dataset by appending DesInventar events onto the set of GDELT events. A new column called *timeline\_event* was created with three different values to show whether an event was a disaster event, humanitarian aid event, or an unrest event.

**3.3. Data Analysis**

For our data analysis, here are the general steps. First, we derive a timeline for each disaster event by adding GDELT events that occurred within a specific number of days after the date of the disaster event (e.g., 180 days, 365 days) and within a geographic distance radius (e.g., 40 km, 80 km, 120 km). The rationale behind building each timeline is to allow us to study the impact of a disaster on subsequent GDELT events. Furthermore, we focus only on two general categories of GDELT events, as alluded to earlier in Section 3.2 above: negative events that are protests, and positive events that are government/humanitarian aids. Second, we study the timelines to discern possible patterns, trends, differences and similarities among the disaster events to explore viable techniques (e.g., clustering, recurrent neural networks) to build a classifier to predict social unrest events. Within this step, we also peruse additional attributes associated with each disaster such as the number of injuries or deaths incurred and economic impact. Third, we aim to build a classifier to predict social unrest events, possibly in two designs. The first design involves a classifier that predicts *whether* the amount of unrest activities will increase or decrease in the next time step given the timeline activities up to current time step. The second design involves a classifier that predicts how much the unrest activities in the next time step will change from that at the current step.

**3.3.1 Step 1: Timelines**

**3.3.1.1 Configuration Parameters of Timelines**

The different configuration parameters used to create a set of timelines were: events within 40, 80, or 120 kilometers of the timeline’s origin event, and events that occurred within 180 or 365 days following the origin event. To provide an idea of the timelines we created, simple statistics about them are provided in **Table 2**. To explain what these numbers mean: if we look at the first entry of **Table 2** (40 km & 180 days), this means that each timeline has one disaster event (‘origin’ event) and all of our GDELT events that were within 40 km of the origin event or occurred equal to or less than 180 days after the disaster. Across all of the timelines created with these parameters, there was a mean of 4.120 events, then we also provide the median, max, min, and standard deviation.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Km range** | **Day Range** | **Mean** | **Median** | **Max** | **Min** | **Standard Deviation** |
| 40 | 180 | 4.120 | 1 | 220 | 1 | 11.795 |
| 40 | 365 | 7.691 | 2 | 234 | 1 | 24.489 |
| 80 | 180 | 11.175 | 4 | 301 | 1 | 20.310 |
| 80 | 365 | 22.593 | 6 | 302 | 1 | 41.861 |
| 120 | 180 | 17.015 | 8 | 334 | 1 | 24.357 |
| 120 | 365 | 35.510 | 14 | 360 | 1 | 51.821 |

**Table 2.** **Timeline Statistics**

**3.3.1.2 Timeline Visualization (Heatmaps)**

Visualizing these timelines can be confusing because a set of timelines has 3 dimensions (each timeline, each event in each timeline, and each attribute for each event) one way of visualizing these timelines is by creating Heatmap tables (**Figure 1** and **Figure 2**) to show how many timelines (Z) have Y events, X days from the origin event at the beginning of each timeline. The X and Y values will have bins of 5 days or events accordingly. In order to make these heatmaps more insightful and simpler, rather than making one heatmap for 1-180 days and one for 1-365, only one was made for 365. Also, we split the events into two sets of heatmaps: one where the timelines only included the origin event and the *negative* (*protest*) GDELT events that occurred after the disaster, and the other had only the origin event and *positive* (*aid*) GDELT events that occurred after the disaster.

|  |  |  |  |
| --- | --- | --- | --- |
| Heat-map |  |  |  |

**Figure 1: *Negative*-Event Timeline Heatmaps for Varying Ranges**

|  |  |  |  |
| --- | --- | --- | --- |
| Km Range | 40 | 80 | 120 |
| Heat-map |  |  |  |

**Figure 2: *Positive*-Event Timeline Heatmaps for Varying Ranges**

When looking at the heatmaps in figures 1 & 2, the following things stood out to us:

* **Observation 1:** *The positive-event heatmaps had very large spikes of aid events in the 1-20 days following a disaster event*. This makes sense as aid is regularly provided to areas hit hard by disasters.
* **Observation 2:** *The negative-event heatmaps had a much smaller increase of events in the 1-20 days period than the positive-event heatmaps*, though it still has one. This could be indicative of a small number of disaster events causing an increase in protests within about a month of a disaster occurring. This will be investigated further.
* **Observation 3:** *There is a spike of events for both the positive-event and negative-event heatmaps in the 260-365 day range.* This could be explained by seasonality of disasters and protests and having a lag between a disaster and the protests it causes. This will be investigated further.

**3.3.2 Step 2: Patterns and Trends**

This section aims to address the observations reported in Section 3.3.1, two of which require further investigation. The questions we can pull from these observations are:

1. Do larger disasters have more events—larger spikes—in the 1-20 or 260-365 day ranges?
2. As part of Question 1 above, what parameters can we use to determine the ‘size’ of a disaster?
3. Are disasters and protest events seasonal?

Question 1 focuses on whether the size of a disaster effects protest events. Follow-up questions to this would include whether the type of impact a disaster has will change the effect on protest events and whether larger events has a larger geospatial or temporal effect. For Question 2, we examine economic impact, number of deaths, and number of injuries as potential metrics for the size of a disaster. When combined with Question 1, we start asking about whether a disaster with a large economic impact but small death impact has a different than a disaster of the opposite variety. Question 3 is asked as a simple explanation for the spike in the 260-365 day range. If disaster events were primarily seasonal and protest events were also primarily seasonal and were focused around 260-365 days after a disaster, then it would easily explain the spike observed in our heatmaps (**Figures 1 and 2**).

**3.3.2.1 Economic Impact**

This section studies Questions 1 and 2 by looking at the economic impact of a disaster as a measure of disaster size. Are disasters with economic impacts of varying size more or less likely to have protest events in the ranges we are investigating? If this were the case, then we would see a higher proportion of timelines having events in the bins of interest if we filtered to only timelines with at least a specific impact size. The economic impact data included in the DesInventar dataset has the following attributes: (1) Houses Destroyed, (2) Houses Damaged, (3) Economic Loss (infrastructure), and (4) Economic Loss (with Agriculture).

The first step to test this would be to generate another set of heatmaps using the same approach as that for **Figure 1** and **Figure 2**, but using *only* disaster events that have a measured economic impact. By only using disaster events with a measured economic impact (*provides a value and value is greater than $0*), our number of timelines is reduced from ?? to 1,304.

Placeholder

**3.3.2.2 Deaths**

Similar to investigating economic impact, we look into the number of deaths caused by a disaster as a measure of disaster size. Are disasters with larger death tolls more likely to cause protests in the day ranges we are investigating? If this were the case, then we would see a higher proportion of timelines having events in the bins of interest if we filtered to only timelines with at least a specific death toll.

The DesInventar dataset provides us with the total death toll of a disaster, as well as a binary value of whether or not there were deaths. We can use these attributes to investigate how the death toll of a disaster correlates with the amount of events in the day ranges we are investigating.

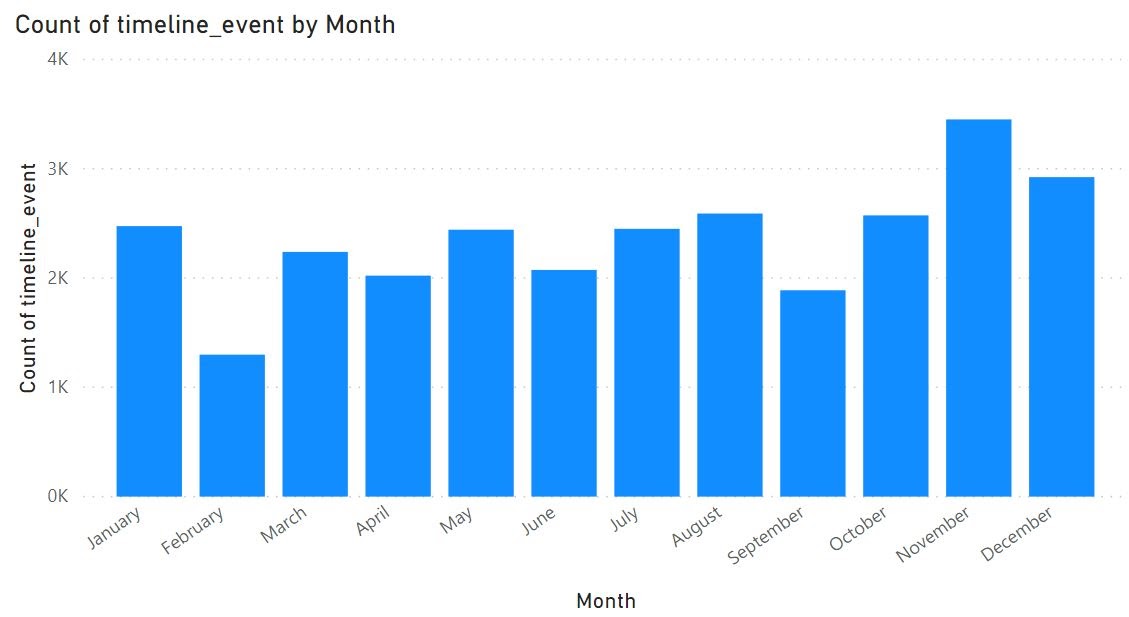
**3.3.2.3 Injuries**

Here, we look at the number of injuries caused by a disaster as a measure of disaster size. Are disasters with larger amounts of injured people more likely to cause protests in the day ranges we are investigating? If this were the case, then we would see a higher proportion of timelines having events in the bins of interest if we filtered to only timelines with at least a specific amount of injured victims.

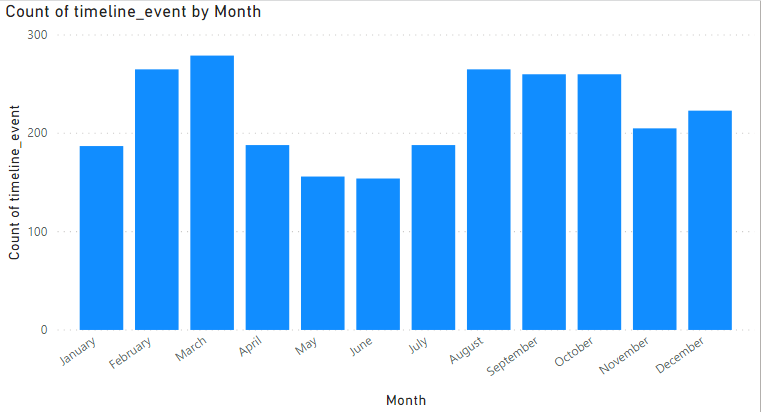
The DesInventar dataset provides us with the total number of people injured by a disaster, as well as a binary value of whether or not there were injuries. We can use these attributes to investigate how the injuries caused by a disaster correlate with the amount of events in the day ranges we are investigating.

**3.3.2.4 Seasons**

This section studies Question 3 by looking at the potential seasonality of disasters and protests. Are the disaster events and protest events seasonal? If the disaster events and protest events were seasonal and the seasons were about 9-12 months apart, it would explain the 270-day separation being so common. The set of disaster events seemed to be fairly evenly spread throughout the year (this can be seen in **Figure 3**). This observation reduced the possibility that the spike was due to protests being seasonal, occurring at a certain time (or certain times) of each year. Likewise, we can further reduce the possibility that the 260-365 day spike is caused by seasonality by looking at **Figure 4**. We see that protest events do seem to occur more around February/March and August-October, with the earlier months having a higher amount of protests. If the 260-365 day spike were cause by seasonality, we would have two similarly-sized spikes in the heatmaps, about 6 months apart. Since both of these spikes do not appear in the heatmaps in **Figure 1**, we can be confident that the 260-365 day spike was not caused by seasonality of disaster/protest events.



**Figure 3: Yearly Distribution of Disaster Events**



**Figure 4: Yearly Distribution of Protest Events**

**3.3.2.5 Temporal Relation of Events**

This section takes a more general approach and adds a question to those at the beginning of this section; “does the order or temporal distance of certain events from the origin event correlate with protest events?”. This question brings up the idea of the events that happen after a disaster determining or signaling whether there will be an increase or a decrease in protest events. If this were true, we could look at various metrics of when events happen after the disaster and whether we see a trend between those and protests. Some of these metrics could include:

1. The number of positive aid events before the first protest event
   1. Looks at whether the amount of aid received after a disaster could impact protests. This would combine well with impact metrics and look at the ratio of disaster size to aid received.
2. Ratio of the number of positive events to the number of negative events in the first N days after a disaster
   1. This looks at whether there were more negative events than positive events close to the disaster occurring, which could potentially lead to a chain reaction and allow for more negative events in the future.
3. The number of days elapsed before the first aid events following a disaster
   1. This is similar to the first metric discussed aboveand focuses on what might impact people’s perception of how the government is handling a disaster.
4. The number of days elapsed before the first protest event following a disaster
   1. This is similar to the second metric discussed above and focuses on what might affect people being more likely to continue protesting.

These metrics will be produced and tested out in clustering algorithms along with the other metrics mentioned in this section to investigate the relationship they might have and how it correlates with changes in protest events. Note also that for the particular dataset that we have, it appears that there is always an aid event on the first day after a disaster, and so is true for a protest event one day after a disaster. That means that instead of looking for the first day of the occurrence of such an event, perhaps look for the day of the occurrence of the *k*th event of the same type. For example, when is the day of the occurrence of the 10th protest event? How to identify *k* could be based on the daily number of events that typically occur regardless of a disaster.

**3.3.3 Step 3: Clustering Analysis**

One method that will be used to analyze the data is clustering. By testing out different parameters and combinations of parameters, we can investigate which clustering methods are most effective, and what clusters are discovered when we allow for combination of different parameters. We will then take these clusters and investigate what parameters led to them and pursue that combination on how the parameters could be working together. This section will detail various types of clustering methods, the results of using them, any interesting clusters they produced, and how we pursued investigating them.

**3.3.3.1 Background on Clustering Methods**

Before running different clustering algorithms on the dataset, discussing and identifying the best clustering methods for our dataset will improve the quality of our analysis. This section discusses various clustering methods and what makes them good or bad candidates for our dataset.

* **DBSCAN (ref) –** Density-based clustering is powerful for clusters of odd shapes and weak with clusters of differing densities. In addition, DBSCAN does not require a number of clusters as an input and handles noise better than many other clustering methods. With the complexities of our dataset, this method could be useful in identifying whether we have oddly-shaped clusters hidden in our data. The best approach for this dataset would be to use our set of parameters mentioned in section 3.3.2 to see if we can find clusters that other algorithms did not pick up on. If these clusters are found, the disasters in that cluster will be further investigated to learn about the cluster and whether this can be used for simple predictions.
* **Hierarchical Clustering (ref) –** Hierarchical clustering can be very useful when someone does not have a number of clusters in mind for their dataset and wants to learn about the groupings that could exist in their dataset. Experimenting with hierarchical clustering would be useful for identifying different numbers of clusters to use for methods that require a number of clusters as an input.
* **K-means Clustering (ref) –** K-means clustering divides the data into *k* clusters and takes in *k* as an input. K-means clustering is a useful and fairly simple clustering method, though it is highly vulnerable to outliers, so the outliers in the dataset will need to be taken into account and handled for this method to be reliable.
* **Fuzzy Clustering (ref) –** Fuzzy clustering allow for a data point to have varying degrees of membership to a cluster, which makes it more effective when it comes to datasets that have overlapping clusters. This method is similar to k-means, but will allow for data points to be included in multiple clusters rather than limiting them to only one.
* **Model-Based Clustering (ref) –** Model-based clustering works off of the assumption that the data were created by a model and attempts to recreate that model. This is a broad group of different clustering methods, and there is even a library in *R* called *Mclust* that exists to help identify the best model for a dataset. This method would be useful in essentially having the computer try to create a model that could potentially allow for simple prediction such as whether a given disaster *A* increases or decreases the likelihood of a protest within *x* kilometers and between dates *y* and *z*. This group of methods would be very useful as the ability for a program to perform predictions on our data would support the possibility of a relationship between disasters and protest events.

**3.3.3.2 Applying Clustering Methods**

This section will focus only on the inputs and outputs for various clustering methods, the analysis and further investigations will be detailed in **Section 3.3.3.3**. The attributes we will be using for clustering are as follows:

1. economic\_loss – the personal economic loss caused by the disaster
2. total\_affected – the total number of people affected by the disaster
3. duration – duration of the disaster in days
4. deaths – the number of deaths caused by the disaster
5. injured – the number of injuries caused by the disaster
6. houses\_destroyed – the number of houses destroyed by the disaster
7. economic\_loss\_infrastructure – economic loss in infrastructure caused by the disaster
8. houses\_damaged – the number of houses damaged by the disaster
9. directly\_affected – the number of people directly affected by the disaster
10. indirectly\_affected – the number of people indirectly affected by the disaster
11. events – the number of events in the disaster’s timeline
12. significantEvents – the number of events in the disaster’s timeline that occurred 260-365 days after the disaster
13. aidBeforeProtest – the number of aid events before the first protest event in the timeline
14. posNegRatio20Days – the ratio of positive to negative events in the first 20 days of the disaster’s timeline
15. daysBeforePositive – the number of days before the first positive event in the timeline
16. daysBeforeNegative – the number of days before the first negative event in the timeline

|  |  |  |  |
| --- | --- | --- | --- |
| **Attribute Name** | **Data Type** | **Grouping** | **Justification** |
| Economic\_loss | Double | Economic | The total economic loss caused by a disaster can help us gauge the effect a disaster might have on the future/hope of the affected area. |
| Duration | Integer | Impact | The duration (days) of a disaster can point to the extended stress on an area, which would have a different impact than a disaster that was relatively instantaneous. |
| Deaths | Integer | Personal | The amount of deaths caused by a disaster can heavily affect people’s ability or motivation to protest. |
| Injured | Integer | Personal | The amount of injuries caused by a disaster can heavily affect people’s ability or motivation to protest. |
| Houses\_destroyed | Integer | Economic | The amount of houses destroyed by a disaster can point to the type of effect a disaster had on a specific community. |
| Economic\_loss\_infrastructure | Double | Economic | The economic loss caused by a disaster in terms of infrastructure can point to the general economic effect caused by a disaster, which could have a long-term effect. |
| Houses\_damaged | Integer | Economic | The amount of houses that were damaged by a disaster points to an economic impact. |
| Directly\_affected | Integer | Personal | The amount of people who reported that they were affected directly by a disaster shows how widespread the initial impact of a disaster was. |
| Indirectly\_affected | Integer | Personal | The amount of people who were indirectly affected by a disaster show how widespread the overall impact of a disaster was. |
| Events | Integer | Timeline | The number of events in a disaster’s timeline (number of event within [40, 80, 120] km and 365 days of a disaster). |
| Significant\_Events | Integer | Timeline | The number of events in a disaster’s timeline that are within the timeline area of interest (occurred over 259 days after a disaster). |
| AidBeforeProtest | Integer | Timeline | The number of aid events that occurred in a disaster’s timeline before a protest event occurred. |
| PosNegRatio | Double | Timeline | The ratio of positive (aid) events to negative (protest) events in the first 20 days of a disasters timeline. |
| DaysBeforePositive | Integer | Timeline | The number of days before a positive event occurred in a disasters timeline. |
| DaysBeforeNegative | Integer | Timeline | The number of days before a negative event occurred in a disasters timeline. |
| NegativeGovActions | Integer | Gov. | The number of negative government actions that occurred during a disaster’s timeline. |
| GovProtestRatio | Double | Gov. | The ratio of negative government actions to protest events in a timeline. |
| GovAidRatio | Double | Gov. | The ratio of negative government actions to positive aid events in a disasters timeline. |
|  |  |  |  |
|  |  |  |  |

**3.3.3.2.1 K-Means Clustering**

The first focus in applying K-Means clustering to a dataset is determining the *k* value, or the number of clusters we want our data divided into. In order to do this, we will be computing the ‘goodness’ of various *k* values and comparing them to find the best *k* for our dataset.

**3.3.3.2.2 Computing Goodness of a *k* value**

In order to determine an optimal *k* value to use, we wanted to choose the value that produced the optimal Good value where:

To compute the centroid of a cluster , , we use the following steps:

1. Cluster the data into *k* groups.
2. Select *k* points at random as cluster centers (as temporary cluster centroids).
3. Assign every object to its closest cluster centroid according to the *Euclidean distance* function.
4. Update the centroid location (mean of all objects) for each cluster.
5. Repeat steps 2, 3 and 4 until the centroid location of each cluster does not change in consecutive rounds.

**3.3.3.2.3 Evaluating Goodness Values**

**4. Results**

**5. Discussion**

**6. Conclusions and Future Work**

References

Olson, R. S. and A. C. Drury (1998) ‘Disasters and Political Unrest: An Empirical Investigation’, *Journal of Contingencies and Crisis Management*, Volume 6, Number 3, Pages 153-161

Becker, S.L. and Reusser, D.E. (2016) ‘Disasters as opportunities for social change: Using the multi-level perspective to consider the barriers to disaster-related transitions.’, *International Journal of Disaster Risk Reduction*, Volume 18, Pages 75-88

BBC News. “Lebanon's PM-Designate Adib Fails to Form New Government after Blast.” *BBC News*, BBC, 26 Sept. 2020, www.bbc.com/news/world-middle-east-54307896.