National College of Ireland

Configuration Manual

MSc Research Project

MSc in Data Analytics

Dixita Yadav

Student ID: x23175796

School of Computing

National College of Ireland

Supervisor: Aaloka Anant

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Dixita Yadav		
Student ID:	x23175796		
Programme:	MSc in Data Analytics	**Year:**	2024
Module:	MSc Research Project		
Lecturer:	Aaloka Anant		
Submission Due Date:	12/12/2024		
Project Title:	Configuration Manual		
Word Count:	863	**Page Count:**	9

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Dixita Yadav

Date: 11/12/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	✓
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	✓
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	✓

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Dixita Yadav
Student ID: x23175796

1 Introduction

This configuration manual provides detailed information of the system configuration, hardware specifications, software used as well as the steps that were taken to carry out the Research project, Evaluating the Impact of Drone Strikes on Terrorism Dynamics.
The information about hardware and software specification is discussed in Section 2 of this manual. Section 3 holds the information of environment setup configuration, Data Gathering and preparation, Importing libraries. Section 4 describes the data preparation and section 5 explains the design and execution of the models.

2 System Configuration

This Section provides detailed information of hardware and software configuration used to implement this project.

2.1 Hardware Requirements

Table 1: Hardware Configuration

Operating System	Window 11
System Type	64-bit operating system, x64-based processor
Installed RAM	8.00 GB
Runtime Model Name	1th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz 1.69 GHz
OS Storage	459 GB

2.2 Software Requirements

Table 2: Software Configuration

Programming Language	Python 3.12.1
IDE	Jupyter Notebook
Web Browser	Google Chrome

Email Account	Gmail account to get verified and link to download GTD dataset
Other Softwares	Microsoft office

3 Environment Setup

This section describes the steps for setting up environment(Jupyter Notebook) and data collection procedure.

3.1 Jupyter Notebook Environment Setup

Due to it's flexible nature of being accessed on multiple devices and platforms, Juypter Notebook (Jupyter Server 2.12.5)[1] 1 is used for this project. It is an open source software and can be downloaded directly via google chrome. Moreover, Jupyter Lab is installed and used for running the code and visualizing the data. Once Installed, open the "jupyter-lab" and run it as an administrator as shown in Figure 1.

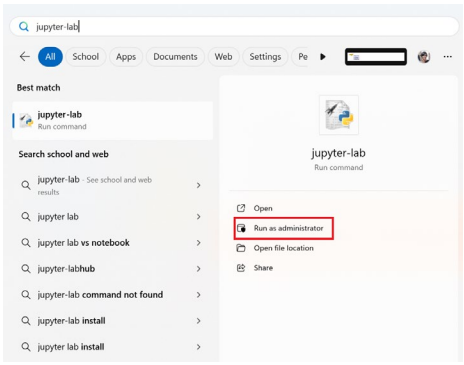


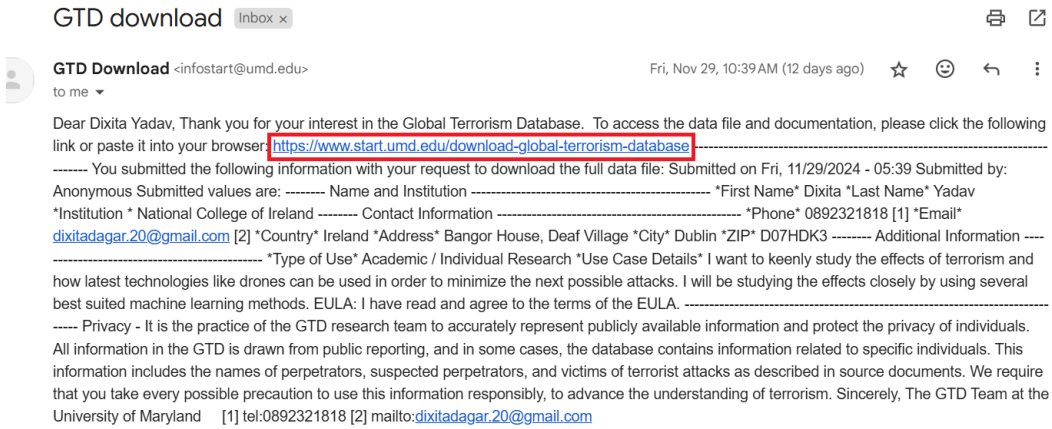
Figure 1: Jupyter-Lab

3.2 Data Collection

Two datasets (GTD and DroneWars Dataset) from different and open source websites are extracted in xlsx formats.
Global Terrorism Database(GTD)[2] , maintained by the University of Maryland. In order to download this data, fill your details with reason to download this dataset. Later, an email with link to download the dataset is sent via gmail, as shown in Figure 2. Click on the link, it will redirect to the START website as shown in Figure 3.

[1]https://jupyter.org/
[2]https://www.start.umd.edu/download-global-terrorism-database

Drone Wars Data set[3], maintained by Bureau of Investigative Journalism(BIJ) on United States drone strikes in Afghanistan, Yemen, Pakistan and Somalia. It can be downloaded directly as shown in Figure 4.



GTD download Inbox ×

GTD Download <infostart@umd.edu> Fri, Nov 29, 10:39 AM (12 days ago)

to me

Dear Dixita Yadav, Thank you for your interest in the Global Terrorism Database. To access the data file and documentation, please click the following link or paste it into your browser: https://www.start.umd.edu/download-global-terrorism-database ---------- You submitted the following information with your request to download the full data file: Submitted on Fri, 11/29/2024 - 05:39 Submitted by: Anonymous Submitted values are: -------- Name and Institution ---------- *First Name* Dixita *Last Name* Yadav *Institution * National College of Ireland -------- Contact Information ---------- *Phone* 0892321818 [1] *Email* dixitadagar.20@gmail.com [2] *Country* Ireland *Address* Bangor House, Deaf Village *City* Dublin *ZIP* D07HDK3 -------- Additional Information ---------- *Type of Use* Academic / Individual Research *Use Case Details* I want to keenly study the effects of terrorism and how latest technologies like drones can be used in order to minimize the next possible attacks. I will be studying the effects closely by using several best suited machine learning methods. EULA: I have read and agree to the terms of the EULA. ---------- Privacy - It is the practice of the GTD research team to accurately represent publicly available information and protect the privacy of individuals. All information in the GTD is drawn from public reporting, and in some cases, the database contains information related to specific individuals. This information includes the names of perpetrators, suspected perpetrators, and victims of terrorist attacks as described in source documents. We require that you take every possible precaution to use this information responsibly, to advance the understanding of terrorism. Sincerely, The GTD Team at the University of Maryland [1] tel:0892321818 [2] mailto:dixitadagar.20@gmail.com

Figure 2: GTD Download link sent to mail

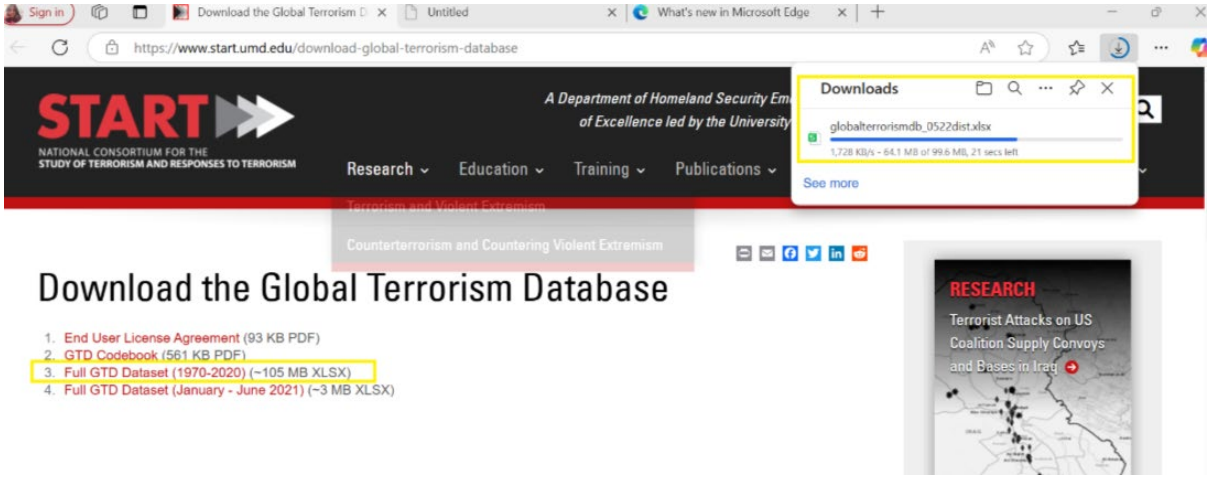


Figure 3: Link to download GTD

[3]https://dronewars.github.io/data/

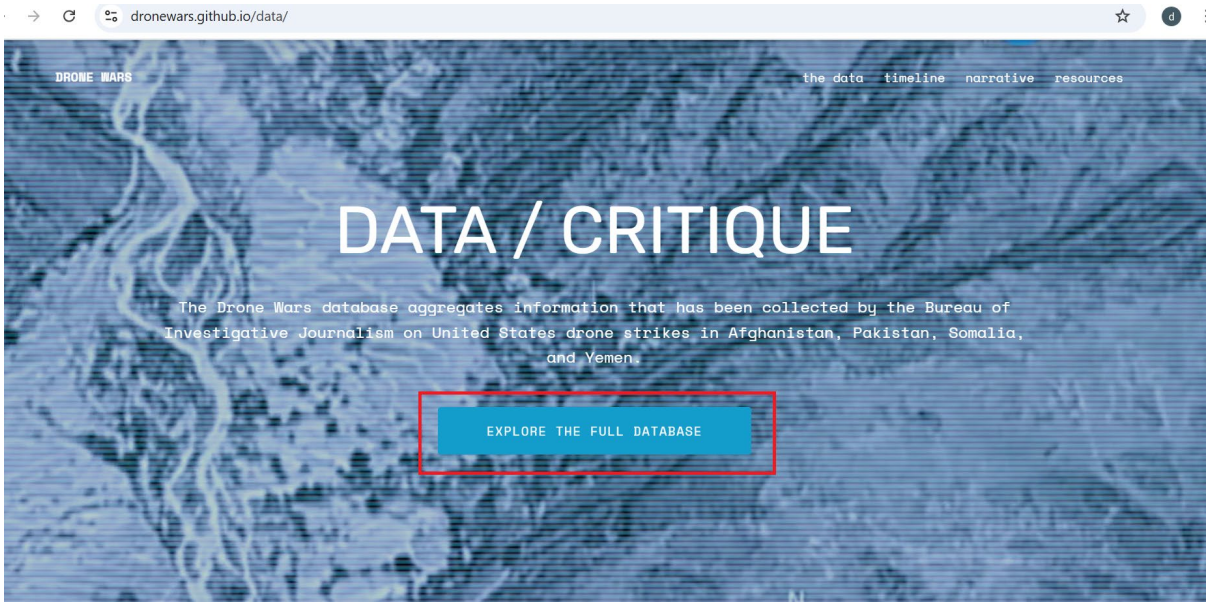


Figure 4: Link to download Drone Wars Dataset

NOTE: Drone Wars Dataset have multiple sheets, the sheet "All" is used for this research as shown in Figure 5.

	A	B	C	D	E	F	G	H
1	Strike ID	Country	Date (MM-DD-YY	President	Most Specific Location	Most Specific Lat/Long	Latitude	Longitude
2	B50	Pakistan	01-01-2009	Bush	Kari Kot, South Waziristan, Pal	32.270651,69.554179	32.270651	69.554179
3	Ob53	Pakistan	01-01-2010	Obama	Ghundikala, North Waziristan, I	32.959854,70.156371	32.959854	70.156371
4	Ob181	Pakistan	01-01-2011	Obama	Mandi Khel, North Waziristan, I	33.001031,70.364339	33.001031	70.364339
5	Ob182	Pakistan	01-01-2011	Obama	Mandi Khel, North Waziristan, I	33.001031,70.364339	33.001031	70.364339
6	Ob183	Pakistan	01-01-2011	Obama	Datta Khel, North Waziristan, F	33.150049,70.433361	33.150049	70.433361
7	Ob184	Pakistan	01-01-2011	Obama	Boya, North Waziristan, Pakist	32.943336,69.899944	32.943336	69.899944
8	YEM115	Yemen	01-01-2013	Obama	Shabwa, Yemen	14.754630,46.516261	14.754630	46.516261
9	AFG001	Afghanistan	01-01-2015	Obama	Spera, Khost, Afghanistan	33.19392, 69.515022	33.19392	69.515022
10	B51	Pakistan	01-02-2009	Bush	Madin Village, South Waziristar	32.320237,69.859740	32.320237	69.859740
11	Ob306	Pakistan	01-02-2013	Obama	Lajhmarai area, Tehsil Birmal,	32.378567,69.430782	32.378567	69.430782
12	AFG136	Afghanistan	01-02-2016	Obama	Shaltan Darra, Shegal, Kunar,	34.846589,71.097316	34.846589	71.097316
13	SOM081	Somalia	01-02-2018	Trump	Mogadishu	2.046934,45.318162	2.046934	45.318162
14	Ob54	Pakistan	01-03-2010	Obama	Mosaki, North Waziristan, Paki	32.943065,69.955033	32.943065	69.955033
15	Ob307	Pakistan	01-03-2013	Obama	Ghundi Killi, Tappi area, Tehsil	32.947289,70.147813	32.947289	70.147813
16	AFG002	Afghanistan	01-03-2015	Obama	Gayan, Paktika, Afghanistan	32.984041,69.430782	32.984041	69.430782
17	AFG003	Afghanistan	01-03-2015	Obama	Spera, Khost, Afghanistan	32.8078448,68.6456967	32.8078448	,68.6456967
18	YEM116	Yemen	01-04-2013	Obama	Radaa, Bayda, Yemen	14.411870,44.836513	14.411870	44.836513
19	Ob358	Pakistan	01-04-2015	Obama	Alwara Mandi, Datta Khel or Sl	33.150049,70.433361	33.150049	70.433361
20	Ob56	Pakistan	01-06-2010	Obama	[illegible] North Waziristan, Paki	32.943065,69.955033	32.943065	69.955033

Afghanistan | Somalia | Yemen | Pakistan | Variables | All | Unknown_Locations | All_With

Figure 5: "All" sheet in Drone Wars Dataset

3.3 Importing Python Libraries

Once the Jupyter Notebook and Jupyter Lab is setup, required python libraries are imported. The code for importing the libraries is shown in Figure 6. The libraries required for the implementation of this project are pandas, geopandas, shapely, sklearn, matplotlib, seaborn.

```
import pandas as pd
import matplotlib.pyplot as plt
import geopandas as gpd
from shapely.geometry import Point
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier, VotingClassifier
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import accuracy_score, classification_report
```

Figure 6: Importing library

3.4 Loading the Dataset

Once the necessary libraries are imported, both the datasets are loaded. The code to load the dataset is shown in Figure 7 and Figure 8.

```
# File path
file_path = 'globalterrorism.xlsx'

# Read the Excel file into a DataFrame
df = pd.read_excel(file_path)
```

Figure 7: Loading GTD Dataset

```
# File path
file_path = 'DroneWarsData.xlsx'

# Read the Excel file into a DataFrame
df_drone = pd.read_excel(file_path)
```

Figure 8: Loading DronesWars Data

4 Data Preparation

Both the datasets are first analyzed properly and descriptive columns with repetitive values that doesn't add any significance to the records are ignored and only relevant columns are retained as shown in Figure 9. As this project demands to merge both the datasets, which is only possible if both the datasets have geospatial coordinated i.e. latitude and longitudes. Therefore, records which lacked geospatial values were removed. Now, after overlaying both the datasets, duplicate columns and records are generated and once again, only relevant columns required for modeling are retained as shown in Figure 10.

```
# Select only relevant columns
df = df[['Day','Month','Year', 'Country', 'Region', 'city', 'latitude', 'longitude', 'AttackType',
         'Killed', 'Wounded', 'Target', 'extended', 'Group', 'Target_type', 'Weapon_type', 'Motive']]
```

Figure 9: Selecting relevant columns from GTD

```
# Remove duplicate rows
filtered_df = filtered_df.drop_duplicates()

# Check the shape of the DataFrame after removing duplicates
print(f"Shape of the DataFrame after removing duplicates: {filtered_df.shape}")
```

```
Shape of the DataFrame after removing duplicates: (9957, 44)
```

```
# List of columns to retain
columns_to_retain = [
    'Day_left',
    'Month_left',
    'Year_left',
    'Region',
    'AttackType',
    'Killed',
    'Wounded',
    'Target',
    'Group',
    'Target_type',
    'Weapon_type',
    'Motive',
    'Strike ID',
    'President',
    'Most Specific Location',
    'Most Specific Lat/Long',
    'Casualties',
    'buffer',
    'Drone_Strike_Occurred'
]
```

Figure 10: Duplicate columns are removed and necessary columns retained from overlaid data

5 Models Implementation

The codes for implementing the predictive model is described below.

5.1 Classification using Voting Classifier

For classifying the severity, the original dataset is split into 80% training and 20% testing sets as shown in Figure 11.

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier, VotingClassifier
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import accuracy_score, classification_report

X = filtered_df.drop(columns=['Severity', 'Killed By Terror Attack','buffer'])  # Remove target and unnecessary col
y = filtered_df['Severity']  # 'Severity' could be Low, Medium, or High
```

```
label_encoder = LabelEncoder()
for column in X.select_dtypes(include=['object']).columns:
    X[column] = label_encoder.fit_transform(X[column])
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Figure 11: Train-Test Split

Voting Classifier model is employed by using three base models i.e. Random forest, SVC and Logistic Regression as shown in figure 12.

```
# Initialize base classifiers for ensemble
rf = RandomForestClassifier(n_estimators=100, random_state=42)
svc = SVC(kernel='linear', random_state=42, probability=True)
logreg = LogisticRegression(max_iter=1000, random_state=42)
```

```
voting_clf = VotingClassifier(estimators=[('rf', rf), ('svc', svc), ('logreg', logreg)], voting='soft')
```

```
voting_clf.fit(X_train, y_train)
```

```
C:\Program Files\Python312\Lib\site-packages\sklearn\linear_model\_logistic.py:469: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
```

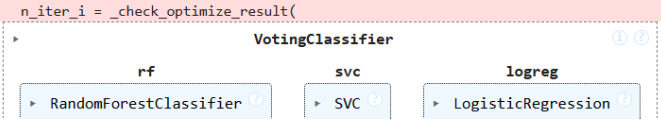


Figure 12: Voting Classifier Model

To check the results and other accuracy scores, a classification report is ran as shown in Figure 13.

```python
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix, roc_curve, roc_auc_score
import seaborn as sns
from sklearn.preprocessing import label_binarize

# Make predictions on the test set
y_pred = voting_clf.predict(X_test)

# Accuracy
print(f"Accuracy: {accuracy_score(y_test, y_pred)}")

# Classification report
print("Classification Report:")
print(classification_report(y_test, y_pred))
```

Figure 13: Classification Report

5.2 Plotting Confusion Matrix

To check the severity of attacks and get a better understanding of the results a confusion matrix is plotted as shown in figure 14.

```python
# Confusion Matrix
cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=['Low', 'Medium', 'Severe'], yticklabels=['Low', 'Medium', 'Severe'])
plt.title('Confusion Matrix')
plt.xlabel('Predicted Labels')
plt.ylabel('True Labels')
plt.show()
```

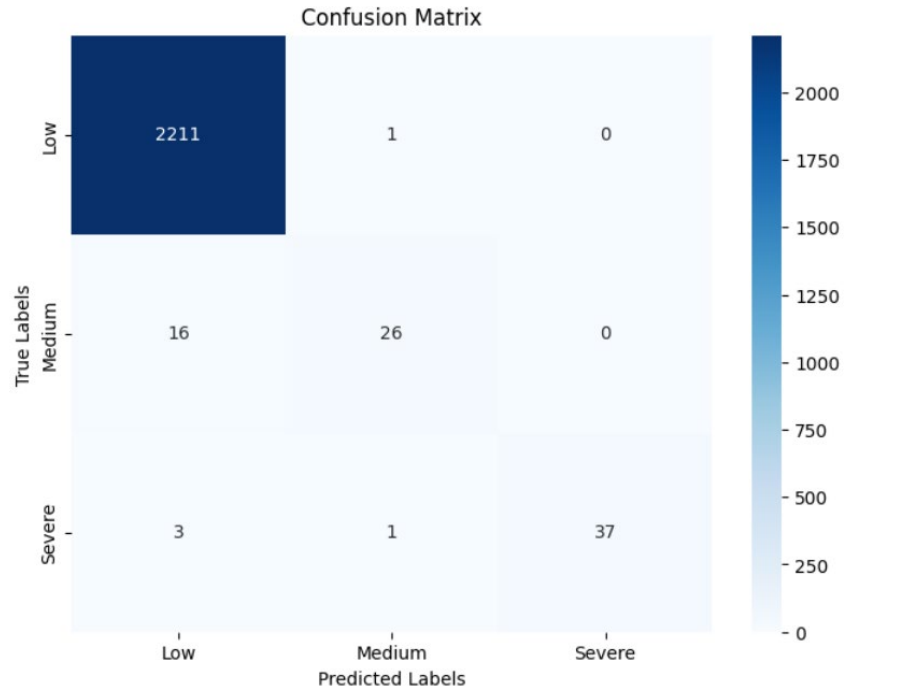


Figure 14: Confusion Matrix

5.3 Plotting Map to check the severity in different regions

To check the severity of the drone strikes, a map is generated across the globe by categorizing green dots - low severity, yellow dots - medium severity and red dots - severe severity for different regions as shown in Figure 15 - Drone strikes severity in Pakistan, Figure 16 - Drone strikes severity in Yemen and Figure 17 - Drone strikes severity in Somalia.

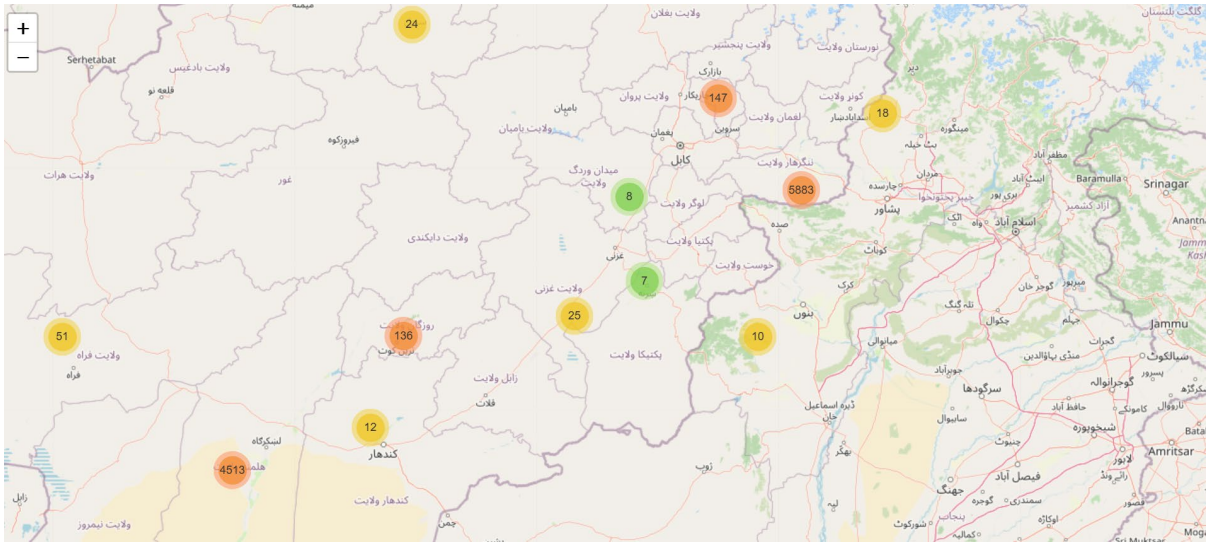


Figure 15: Drone attacks severity in Pakistan

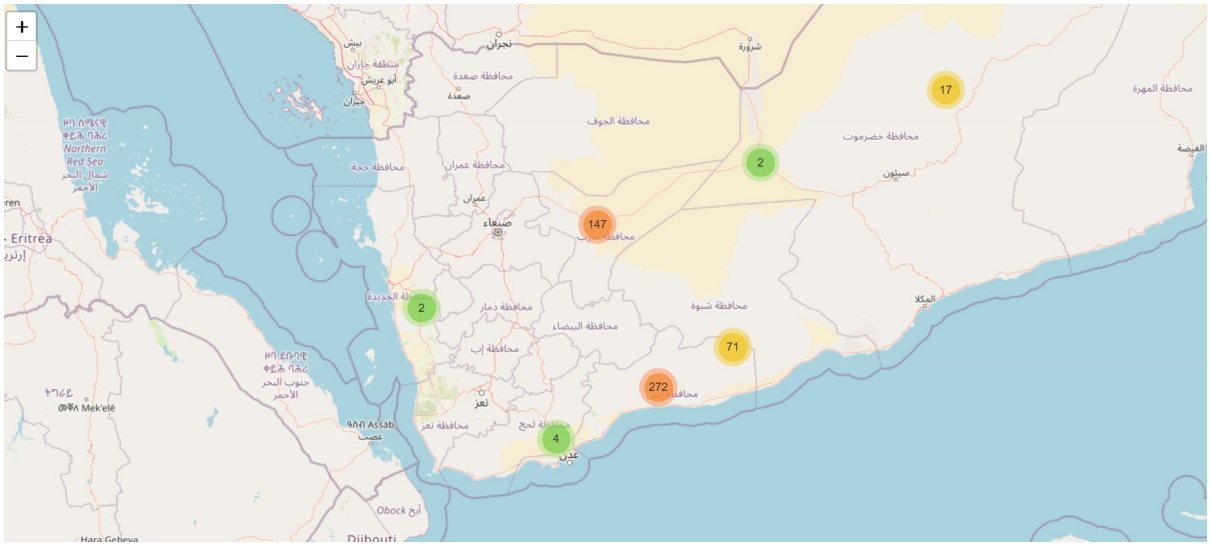


Figure 16: Drone Strikes severity in Yemen

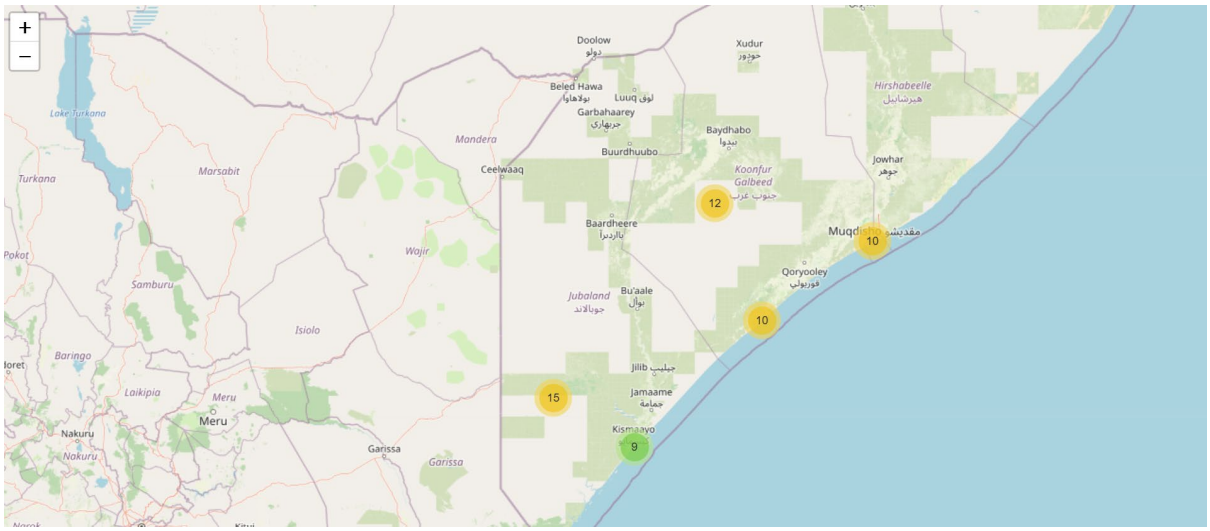


Figure 17: Drone strikes severity in Somalia