MACHINE LEARNING ENGINEER NANODEGREE

CAPSTONE PROPOSAL

CONTEXT OF TERRORISM IN SOMALIA

AHMED GAAL,   June 1, 2021

DOMAIN BACKGROUND

Somalia is a coastal country that lies at the Horn of Africa from the Gulf of Aden to the Indian Ocean. It has a coastline of more than 3, 300 km making Somalia the country with the longest coastline in Mainland Africa.

Somalia's population had **75%** literacy rate and a booming tourism industry hence earning the name **'The White Pearl of the Indian Ocean.'**

Somalia has immense natural resources and mineral reserves such as oil and natural gas, uranium, iron ore, tin, gypsum, bauxite, copper and many more.

Somalia once possessed the most lethal Air Force in Africa and as such it trained **Burundi's**[1] air force, protected borders of **Tanzania** & **Uganda** and flew fighter jets for **Zambia**.[2]

It defended **Mozambique** from the Portuguese, trained **South Africans** fighting apartheid and supplied **Eritrea's**[3] war of independence.

Somalia sent troops to **Angola**, supported **Egypt** with naval logistics, supported **Djibouti's** independence movement and also supported **Zimbabwe** and **Namibia's** forces in the war against apartheid.[4]

1. Burundi Air Force

2. Somali Air Force aids Zambia

3. Somalia is a key ally during the Eritrean War of Independence

4. Somalia role in Africa's Modern Warfare

SOMALI CIVIL WAR

In 1991, Somalia was ravaged with war which went on for almost the past 3 decades after the toppling of the military regime governed by Major Gen. Mohamed Siad Barre. This gave birth to Somalia's Civil War and clan factions began fighting for power. Many civilians were annihilated and many more fled the country as refugees in neighbouring Kenya and all over the world.

In 1993, the US Rangers conducted Operation Gothic Serpent which resulted in a failed mission after it lost 2 Sirkosky UH-60 Black Hawk helicopters, 19 US servicemen and at least 1,000 Somali civilians were killed and 4,000 were wounded.

In 2006, **the Islamic Courts Union (ICU)**, an Islamic organization, assumed control of much of the southern part of the country and promptly imposed Sharia' law.

On the dawn of 2007, TFG President and founder Abdullahi Yusuf Ahmed, a former colonel in the Somali Army, entered Mogadishu with the Ethiopian military support and relocated the government to **Villa Somalia** in the capital from its interim location in Baidoa.

This marked the first time since the fall of the Siad Barre regime in 1991 that the federal government controlled most of the country.

Following this defeat, the Islamic Courts Union splintered into several different factions. Some of the more radical elements, **Al-Shabaab**, regrouped to continue their insurgency against the TFG and oppose the Ethiopian military's presence in Somalia.

Throughout 2007 and 2008, Al-Shabaab scored military victories, seizing control of key towns and ports in both central and southern Somalia.

At the end of 2008, the group had captured Baidoa but not Mogadishu.

By January 2009, Al-Shabaab and other militias had managed to force the Ethiopian troops to retreat, leaving behind an ill-equipped African Union peacekeeping force to assist the Transitional Federal Government's troops.

PROBLEM STATEMENT

Terrorism is defined as the threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious or social goal through fear, coercion, or intimidation.[1]

Somalia has faced quite a strife for almost 30 years and most of those was between clans. Up to date, clans engage in armed conflict attributed to diverse motives that are not terrorism.

The objective of this project is to carefully examine, analyse, map and classify the events where the type of attack was an act of terrorism or not.

Also classify events where the targets of attacks were non-combatants (collateral damage) or not.[2]

This project will reveal insights on the types of attacks prevalent in Somalia, Where these attacks are mostly targeted and what types of weapons are employed. This will provide a clear understanding of context of terrorism in Somalia.

DATASET AND INPUTS

The data that will be used throughout the course of this project will be obtained from the **Global Terrorism Database**.

This database was granted by the **National Consortium for the Study of Terrorism and Responses to Terrorism** at the **University of Maryland**.

This data is the most appropriate and reliable source of information that would ensure that the objective of this project is met.

The dataset contains 135 dimensions and 191,463 data points. Most of the classes in the dataset are imbalanced.

1. Terrorism According to National Consortium for the Study of Terrorism and Responses to Terrorism.

2. USAF Intelligence Targeting Guide

SOLUTION STATEMENT

The solution I intend to deliver to the problems clearly defined in the problem statement will be:

- Create a classifier to classify types of attacks that constitute as an act of terrorism and those that are not.
- Create a classifier to classify events where the targets were non-combatants.
- Create a dashboard to visualize the types of attacks, target locations and assortment of weapons employed during these events.

BENCHMARK MODEL

I believe Decision Tree is the most suitable benchmark model for this project given that Decision Tree rarely tends to overfit and we will thereafter explore other models such as Random Forest, Gradient Boosting, Extreme Gradient Boosting, Logistic Regression and Support Vector Machines.

EVALUATION METRICS

The most suitable evaluation metrics we will implement given this is a binary classification problem is:

- **Accuracy Score.**
- **F1 Score.**

PROJECT DESIGN

- **Data Extraction**
  - We first need to extract our starting data from cloud storage.
  - We then need to filter our data and extract data that is related to Somalia so that we can initiate data pre-processing step.
- **Data Pre-processing**
  - We need to drop null values column-wise because it will not be appropriate to interpolate or fill in missing values to make the study reliable.
  - Some of the days and months have '**0**' observation which in real-time setting has no meaning and according to the data, it is not known. For this reason, I find it suitable to drop these values.
  - There are some features which are irrelevant to the objective we are aiming for so it is only convenient to drop it.
  - Using **year**, **month** and **day** features, we can introduce a new feature i.e., **Date** which will be vital in creating interesting visualizations.
  - Finally, we will rename the columns to more suitable names.
- **Data Manipulation and Visualization**
  - Create a scatter-map to visualize the types of attack and the actual location where it happened.
  - We will also reveal interesting insights about which locations are prone to which types of attacks.
  - We will investigate which types of weapons are of choice for which groups.
- **Exploratory Data Analysis**
  - We will compute the descriptive statistics to identify meaningful information and also help in feature extraction and selection.

- We will also compute pairwise correlation to find linear relationships.
- We will employ predictive power score to explore non-linear relationships and other relationships to filter the information leakage that may be presented by the pairwise correlation.[1]

- **Feature Extraction**
  - We will extract features using insights divulged at the Exploratory data analysis step to provide train and test data.
  - We will the split train and test data to a ratio of 80 – 20 respectively.

- **Model Training & Selection**
  - We will make use of scikit-learn's **pipeline** module to counter test data to leak on to train data when training our model. [2]
  - We will apply the **StandardScaler** method to normalize our data before fitting to a model. [3]
  - We will scroll through linear models, ensemble and support vector machines to select the most robust model.
  - We will employ **hyper-parameter tuning** whenever necessary to find the most potent parameters.

- **Model Evaluation**
  - We will use accuracy score, precision and recall score to evaluate our model performance.
  - We will implement a confusion matrix to detect outliers and detect whether the model is overfitting.

1. Predictive Power Score by 8080labs

2. Chaining estimators using Pipeline method

3. Standardize features by removing the mean and scaling to unit variance