# Security and social media:

# In modern day, communications play an essential role in military command and control. Effective exchange of clear and concise information and orders have always been key and vital to the success of any mission. With the rapid evolvement of all aspects of war, communications are advancing in an exponential ration. Nowadays, the success of “combined operations” has effective, reliable and fast communication as its corner stone. This evolvement is a 2-edged sword which present huge advantages on one side, but on the other, as it is available for free to all, may present a real threat and in the same time could be a weapon in its own.

# The concept of social media was to find easy ways that can connect people together in a cheap and fast way. Organizations did not wait for long to participate in the networking applications as they find in them great business opportunities. Social media changed the way of exchanging information, and that change came is fast way that made infiltration behind the lines easier for the enemy.

# Security concerns started to build quickly with the advent of social networking. In the modern world of technology, whether local of foreign, attackers find it easier every day to implement their plans and commit actual attacks on persons, companies, national and governmental institutions. These attacks include but not limited to: identity theft, viruses, malwares, phishing, privacy of data, tracking of persons, exposing confidential sensitive information ….

# More importantly, the use of social media platform to spread propaganda, bombarding the public opinion with lies and fake news, represent real challenge for homeland security, posing major threat to the national and economic security. Cyberspace is the new playfield for attackers/terrorists, and cybercrimes are culminating.

# That invasion of internet and social media made the world smaller and smaller but making at the same time its security threats wider and wider. It is really a big challenge for intelligence and security staff to overcome or at least cover that growing pool of threats.

# The technology that made internet and social media platforms in every house or even with each human, is on the other hand, giving us as intelligence members, a bunch of tools, that if properly used, could give us the lead….

# So, it is vital for us to plunge in the realm of big data, use the same tools available for terrorists and suspected people, analyze what they are spreading within their network, predicting their intentions, and most of all taking actions at the correct time to eliminate the threats.

# Methodology:

# In order to respond to those security threats and act against them, I decided to tackle the problem on two levels: Data analysis and GIS.

# For data analysis mining and demining, targeted databases will be Twitter and Facebook, due to their huge increase in data day by day, and the option available to access their databases.

# The approach is as follows:

# Prepare a list or category of predefined words related to security events (explosives, bombing, robbery, gun fire, manifestations, steeling, arrest, …).

# Collect Data (api, web scrapping)

# Prepare the data: Cleaning and Filtering.

# Feature extraction using deep learning techniques.

# Extract the geographic locations of incidents to plot on GIS platform.

# Proceed with Named Entity Recognition to cluster the tweets related to the predefined category.

# Do a Sentiment Analysis to determine the polarity relative to certain topics which are considered as indices in intelligence?

# Draw network gathering positive classes.

# Compare with local databases in a concurrent way to link suspected people.

# Locate suspected targets in different ways (cell phones, internet access points)

# Every security action must be executed at a certain location, because after all in military and security location-based actions must be taken, so once GIS data is collected, it will be molded using the powerful GIS tools to complete the scene for a security action. Here comes the role of integrating multiple geodatabases together in a unique and global GIS platform.

# The first task is to prepare a list of predefined words directly related to security events, and create a category of these words, in both English and Arabic, and if possible, a custom language consisting of Arabic words written in English alphabet and Hindi numbers. This is due to the fact that, in Lebanon, besides of Arabic and English, many users use the Arab English vocab made by writing Arabic words in English characters as they are pronounced, in conjunction with special characters from the Hindi numbers (2, 3 and 7).

# The second and most hard task is to the data collection. It can be done using api or web scrapping. For Twitter data it must be simplest than for Facebook. Several third-party services are available for this task, for instance: Zapier, Dexi, ScrapeStorm, Content Grabber, Pattern, Graph API, meltwater.com(datasift.com), nodexl.com, topsy.thisisthebrigade.com

# For Twitter database, they provide an academic developer account with elevated privileges (content of tweets and number of tweets you can download per day), but unfortunately in my case as Officer at the Lebanese Army, I was not able to subscribe in that account. And there is also a difficulty to extract the incidents’ locations due to the fact that the stored location is that where the account was created and not where the incident took place.

# Downloaded data, in its raw format, cannot be used directly in natural language processing, machine learning and deep learning models. It must be cleaned and filtered. This step is also known as pre-processing the data, and can be done using the python NLTK library which contains several tools to accomplish many related processes: Letter casing, tokenizing, noise removal, stop word removal, normalization, stemming and lemmatization.

# This phase is the core phase in which pre-processed words are converted to vectors. Many methods could be used here, from lexicon methods (dictionary and context), syntactic and semantic, machine learning and deep learning methods.

Deep learning models are pre-trained models that have built-in feature extraction and word embedding techniques, and can be fine-tuned with already classified data to give much better results.

The most widely used models are for instance: CNN, LSTM, GPT-3 and BERT.