Steven Szachara
Graduate Co-Op Application
9/21/2024

stevenszachara@gmail.com
ss9270@g.rit.edu
3034895962

Machine Learning Model for Asset Identification

Theorizing a Machine Learning Model to Classify New Foreign Assets

This document will focus on a hypothetical machine learning model solution to identify and classify new foreign assets. Given the limited knowledge known in my current position about the levels of access to data that can be used for the model, I can only theorize, and in many ways, this can be interpreted as "wishful thinking". Even so, this is how I would approach.

Why Use a Model

The common use-case for machine learning models is a reduction of time and an improvement in efficiency. The amount of time it takes to identify patterns in data manually is cut drastically. A real-world example is the difference between manually identifying, and a model identifying, key proteins in biological studies, being reduced from months to days comparatively. A model's predictive capability is also something to note. Models can be designed to predict things, for example, cancer clusters, before they are formed enough to be seen by the human eye.

> *"[with machine learning models]… identifying, key proteins in biological studies, being reduced from months to days comparatively."*

Machine learning models have become a recent phenomena, not only because of their power, but because of their ease of implementation through the work of countless individuals over the last 10 years. This would be the latest and greatest form of technology to implement, given the United States goal to be ahead of it's competitors.

The Goal of the Model

An asset, for the purposes of this document, is a foreign citizen, group, or government party who has information, property, access, and any other item of potential value, that can be used to gather intelligence on a current geopolitical situation, problem, or in general, responding to a specific request. The goal for this model, would be to take

stevenszachara@gmail.com
ss9270@g.rit.edu
3034895962

travel patterns, purchasing patterns, and the digital footprint patterns of an asset, and determine whether or not it is of value to track/pursue as a should-be current asset, a future asset, a hostile asset or a non-asset. This model will not be limited to a one-time set of data, and will be able to continually update the predictions. Ultimately, and theoretically, this model will be able to predict a valuable asset in advance, based on previous learning and updated training.

Where to Start: Data

With any set of data, one is presented with many problems, but the first step in building a machine learning model to decipher this data, is to effectively "clean" it. Without proper data, there can be no solution. Garbage in, garbage out, as they say. When it comes to data, just what are we referring to? Given the plethora of data that has likely already been mined and is available for use within, as well as through partnerships with other foreign government agencies, the goal here would be to filter. The data theorized here could be travel patterns, purchasing patterns, digital footprint patterns, connectivity within the government, friends, family, wealth, as a couple of examples.

Theoretical Use Cases

As a general use case, the model would be given a request on a specific topic, key words, emotional patterns, jobs held, travel capabilities/past travel, current and past digital footprint analysis to indicate said topic. A data prep for a specific region will have to be implemented, a distance/radius, and theoretically, the model will be able to limit the necessary "start points" needed to be given on a topic, to generate a report. The report will be a list of should-be current assets, a future assets, hostile assets, or non-assets.

A hypothetical use case: "Need more information on the political situation in Lebanon. 100 mile radius from the capital. High profile targets only. All asset classifications." From the hypothetical data cleaning process of the constantly updating data that is being informed and stored, the model will have been trained for the specific region, and specific country, we have given it. Based on the parameters above, the model will be able to give a list of the four asset classifiers. In theory, the model will identify should-be current assets as those likely to give information, assets that have the potential

stevenszachara@gmail.com
ss9270@g.rit.edu
3034895962

to give future information (think younger political figures, business figures, sons and daughters of these figures), hostile to being an asset (with the notion that, the enemy of my enemy is my friend type scenario, which can be valuable to know) and non-assets, which will likely be the biggest group (not enough information known or not valuable enough). With the non-asset category, this can be further modeled for specificity, however, given the large amount of data that will be looked at, specifics are key.

Multiple Models

I mentioned before that the model can be adjusted to allow more or less precision, which would require using different models that are trained to do exactly this. However, once the structure of a model is in place, this process of making different types of models becomes much easier.

Once the initial model has been built on a general set of parameters, the adjustments for region, tensions, history, and specifics needed to build other models on top of the original can be utilized. For example, a model that has a high precision will eliminate a lot of “potential” assets, and only present the absolutely close to 100% should-be assets in the report. However, a model that has high recall will present more of these should-be assets, but not all will be as “good”, theoretically. But, having the option for utilizing both versions of these models is key, depending on what the scenario requires.

How to Measure Success

How do we know if the model is working correctly? In any machine learning model, it is important to know what an accurate result looks like. In this case, one approach is to base the model on previous examples of assets being discovered and utilized to learn what exactly an asset “is”. That information would be critical to identifying the list it generates. Once the model is aware of what the “accurate results” are in classifying should-be, potential, hostile, and non-assets, numerically of course, we can then test it against past assets not used in the training phase, that have been identified in this manner, and see the level of accuracy the model displays.

stevenszachara@gmail.com
ss9270@g.rit.edu
3034895962

"... base the model on previous examples of assets being discovered and utilized to learn what exactly an asset "is"."

Another way to measure success? Manual verification. Just because a model predicts a should-be asset, does not always mean that they as a human, with many complexities and unknowns, will be an actual asset. A model won't be able to make something an asset. However, cutting through the manual labor of sifting though data, and having a helpful tool as a starting point, will, in theory, give the benefits of time efficiency, eliminating some human error, and redirecting focus.

Final Thoughts

The amount of effort and initial man-power needed for this effort is debatable, depending on the hiring and leadership of the people involved, as well as the access to data, but still, it will take time. I am also aware that some version of what this document entails is already likely being worked on, and I also realize I have no business in the level of access to this type of mass data and implementation requires, however, I wanted to express how I think my skillset can be utilized, and my hypothetical idea.