Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
May 4, 2023
A Regional Conference of the Society for Industrial and Systems Engineering

Risk Evaluation for Command Operating Networks (RECON) for US Army Special Operations

Natalya Milian[1], Jonathan Parent[1], Joshua Thombleson[1], Ashley Ulricson[1], Tyler Vratny[2], Adah Spain, and Julia Coxen[1]

[1]Department of Systems of Engineering
United States Military Academy
West Point, New York 10996

[2]Department of Mathematical Sciences
United States Military Academy
West Point, New York 10996

Corresponding Author's Email: ashley.ulricson@gmail.com

Author Note: The authors of this paper conducted a year-long research project through the Department of Systems Engineering at the United States Military Academy under the advisement of Colonel Julian Coxen. COL Coxen currently serves as the Department Head in Systems Engineering. The views expressed in this paper are those of the authors and do not reflect those of the United States Army, the United States Military Academy, or the Department of Systems Engineering.

Abstract: The mission set of United States Special Operations Command (USASOC) is to coordinate and support special operations forces domestically and internationally. One point of friction that the organization faces is assembling a concise collection of data that informs traveling soldiers of relevant and important factors pertaining to their destination. This project aims to aggregate disparate data to model the risk for declared or undeclared personnel traveling to any country. The Risk Evaluation for Command Operating Networks or RECON System collects open-source data, aggregates it, and then evaluates it to produce a risk assessment for commanders to interpret. The compiled information will then be presented to the commander of the unit traveling overseas and allow them to make a more risk-informed decision.

1. Introduction

The United States Special Operations Command (USASOC) sends numerous individuals abroad to a variety of countries across the globe. When deciding whether to send personnel abroad, commanders must weigh the numerous risks associated with approving an individual X's travel to country Y. As described by our stakeholder within USASOC, Lieutenant Colonel Kevin Larrabee, when making these decisions, decision-makers task intelligence professionals to find relevant information regarding country Y and manually compile it to present inherent risks associated with this travel event. LTC Larrabee emphasized there is a need to streamline this current ad hoc process as it is time-consuming, labor intensive, and differs for every deployment and individual. A commander could potentially be left without relevant, timely, and detailed information pertinent to the mission of sending individuals abroad which could impact unit operations, mission readiness, and personnel safety. Through an interview with LTC Larrabee, he informed us the military examines risk as it impacts the mission and the force. The sensitive nature of special operations means that some individuals will be entering countries with ubiquitous surveillance. Risk to the mission includes potential data exposure, information on unit associations, and freedom of maneuver, while risk to the force includes individual safety like bodily harm, detention, or even death.

The world is becoming increasingly data-driven every day, which means the greatest threat to mission and force readiness is how other countries collect data on US personnel abroad, whether this be through biometrics, social media, or other forms that could link personnel to USASOC. Understanding risk as it pertains to the capabilities a country possesses to collect data, and their interest in how that data relates to the US, will help commanders consider how an individual's travel may threaten the overall mission and force. USASOC has an abundant amount of data on different countries, but they lack standardized methods of cleaning, aggregating, and presenting this information to individual travelers and commanders. This is why USASOC needs a model that measures the will and ability of a country to detect abnormalities in an individual's data signature, then display this information to commanders through a dashboard so they can make a risk informed decision before sending personnel abroad. We define ***will*** as a country's desire to investigate an individual traveler and ***ability*** is defined as a country's capacity to recognize data abnormalities amongst travelers. Risk is defined as the negative impact of foreign travel to mission and force as it relates to will and ability.

2. Background

To gain a better understanding of the problem to be solved, it is necessary to understand the effect of smart cities and artificial intelligence ethics on data collection procedures, as well as the important concepts of linear regression and black box models for the modeling process used. The phrase "smart cities" is the term coined for the mass collection of data and its interpretation through surveillance on a city-wide scale to create a more efficient system (Axhausen et. Al., 2012). This intense process underlies innovations such as the expansion of digital ecosystems, the Internet of Things (IoT) (Cheema et. Al., 2021), and 5G capabilities (Cheein & Guevara, 2020). One drawback that is especially relevant to our study is the impact of data collection in smart cities and how this can possibly violate the privacy and security of travelers. Several existing literature sources already address how smart cities may lead to security violations of personal privacy, whether from the government (Yau, 2019) or data leaks (Dodge & Kitchin, 2017). Solutions provided to this problem range from better security frameworks to encryption techniques, but these solutions are not easily implemented by travelers. Each country's own values and culture shape how AI ethics are implemented and enforced around the world representing a complex challenge in analyzing the safety and security of personnel abroad. While the exact policies and regulations on ethical AI remain varied, there are many key principles that remain the same amongst numerous countries. In a study conducted analyzing AI principles in various countries, it was found that transparency, privacy, fairness, security, and safety were the top five principles across all nations (Tidjon & Khomh, 2022). In summary, most nations around the world can agree on similar principles that shape AI guidelines and regulations, however, each continent differs in its personal enthusiasm to implement such policies to protect transnational data.

A random forest regression model was chosen for this project because in recent years there has been a push back against black box models which are difficult to interpret. According to Cynthia Rudin, black box models have been causing problems throughout healthcare, criminal justice, and in other domains (2019). Black box models can become problematic due to a lack of interpretability which many have dire consequences (Rudin, 2019). The military is an organization built on people, and people are USASOC's greatest assets therefore using a black box model poses a greater risk to the safety and lives of individuals if we do not know how it is producing its results. Providing an easy to interpret model for commanders will deliver them with the greatest confidence in the model and a greater understanding of how to interpret this risk.

3. Methodology

To create the dashboard for USASOC commanders to assist them in determining the risk of sending Soldiers abroad, the problem was separated into two components: a quantitative and qualitative analysis. The first component involved developing a model that measures the will and ability of a country to detect abnormalities in an individual's data signature by collecting quantitative data. This is then implemented into a random forest regression model to predict a risk score. The second component is providing commanders with qualitative data to contextualize the predicted risk score. This was collected through webscraping the State Department Travel Advisory, Twitter, Reddit, and other websites to display current events within each country as it pertains to potential threats. This was then compiled into a user-friendly dashboard on Streamlit to present to commanders. See Figure 3 on page 4 for a process chart which outlines the process utilized to achieve this final output of a user-friendly Streamlit dashboard from the initial input of the situation to send an individual traveler to a specific country.

3.1 Quantitative Data Analysis

Our risk assessment is an ensemble model consisting of three individual models, two random forest models (one for will and one for ability) that produce two risk scores and a third model that combines them into a classification matrix. To provide a risk assessment with real time accuracy, our scores for will and ability need to reflect current data that is pertinent to USASOC's mission. There are no open book sources that provide an hourly, daily, or monthly measure for will and ability with relevant factors. Using a random forest regression model allows us to use proxy measures for will and ability while taking into consideration elements with a greater interest to USASOC. Our proxy measures acted as dependent variables predicted by approximately ten different independent variables. In terms of ability, we used the Global Innovation Index (GII) that provides a "rigorous statistical benchmark that attempts to capture national innovation ecosystems" as a proxy measure and/or a_{tq} under the assumption countries with greater resources, infrastructure, and technology will have a greater ability to detect abnormalities in a traveler's data signature (*Indicator Rankings & Analysis*). Equation 1 is how we formulated our ability model, where a_{tq} is the predicted ability score in year "t" for country "q", y_n is the independent variable, and b_n is the associated coefficient. y_n represents data pulled from the World Bank, International Monetary Fund, and US Energy Administration as seen in Table 1, for country "q" in year "t."

$$a_{tq} = b_1 y_1 + b_2 y_2 + \ldots + b_n y_n \qquad (1)$$

The Our World in Data Human Rights Distribution Index (HRDI) scored countries from 0 to 100 on the "extent to which people are free from government torture, political killings, and forced labor, they have property rights, and enjoy the freedoms of movement, religion, expression, and association" where 0 is a low protection of human rights and 1 is a high protection (Herre 2016). We used the HRDI as a proxy measure for will and/or w_{tq} under the assumption the level at which countries protect data privacy rights will be equivalent to how they respect individual liberties. Equation 2 is how we formulated our will model, where w_{tq} is the predicted will score in year "t" for country "q", x_n is the independent variable, and b_n is the associated coefficient. x_n represents data pulled from the Our World in Data, Freedom House, World in Data, and World Bank as seen in Table 1, for country "q" in year "t."

$$w_{tq} = c_1 x_1 + \ c_2 x_2 + \ldots + c_n x_n \tag{2}$$

Table 1. Data Sources Utilized for the Will and Ability Risk Score Factors

Ability – Global Innovation Index	Will – Human Rights Distribution Index
International Monetary Fund - National Debt World Bank - Armed Forces Personnel - Exports of Goods and Services (current US$) - Gross Domestic Profit (current US$) - Gross Domestic Profit Growth (annual %) - Imports of Goods and Services (% of GDP) - Individuals Using the Internet (% of the population) - Industry Value Added (annual growth %) - Inflation (annual %) - Military Expenditure (% of GDP) - Unemployment (% of total labor force) US Energy Information Administration - Energy Consumption - Energy Production	Our World In Data - Civil Liberties - Freedom of Press - National Rights Institutions Freedom House - Freedom House Score - Freedom on the Net Score World Data - UN Member - NATO Member World Bank - Control of Corruption - Government Effectiveness - Political Stability and Absence of Violence/Terrorism - Regulatory Quality - Rule of Law - Voice of Accountability

Our third model as seen in Figure 1b, is derived from the Army's Risk Assessment matrix from ATP 5-19 to categorize the risk as it pertains to the four risk levels: Extremely High (EH), High (H), Medium (M) and Low (L). During our stakeholder analysis, experts expressed a preference for visualization of risk in a familiar format that would be easy for commanders to interpret. We projected the risk model outcomes synthesized from the two sub-models onto the Army's Risk Assessment matrix as seen in Figure 1b. This not only incorporates the stakeholder's wishes, but also provides a simple and recognizable format that requires no additional training.

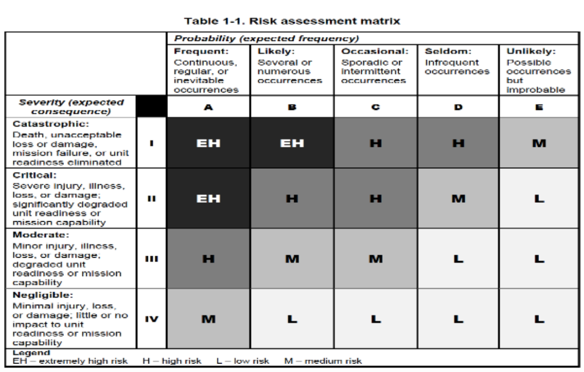

Table 1-1. Risk assessment matrix

		Probability (expected frequency)				
		Frequent: Continuous, regular, or inevitable occurrences	Likely: Several or numerous occurrences	Occasional: Sporadic or intermittent occurrences	Seldom: Infrequent occurrences	Unlikely: Possible occurrences but improbable
Severity (expected consequence)		A	B	C	D	E
Catastrophic: Death, unacceptable loss or damage, mission failure, or unit readiness eliminated	I	EH	EH	H	H	M
Critical: Severe injury, illness, loss, or damage; significantly degraded unit readiness or mission capability	II	EH	H	H	M	L
Moderate: Minor injury, illness, loss, or damage; degraded unit readiness or mission capability	III	H	M	M	L	L
Negligible: Minimal injury, loss, or damage; little or no impact to unit readiness or mission capability	IV	M	L	L	L	L

Legend
EH – extremely high risk   H – high risk   L – low risk   M – medium risk

Figure 1a. Army Risk Assessment Matrix ATP 5-19

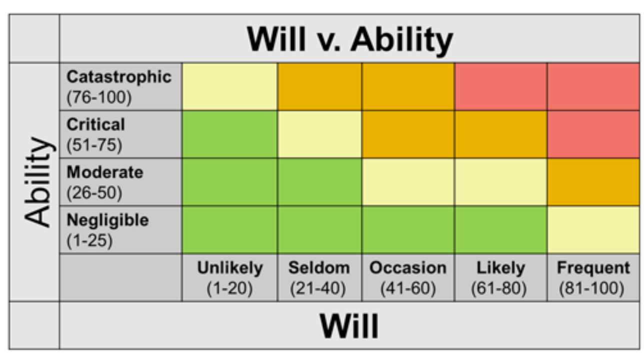

Figure 1b. RECON Risk Assessment Matrix

3.2 Qualitative Data Analysis

This research relies heavily on webscraping as the primary method for data collection. This has largely been done in Python utilizing the Beautiful Soup Python library. The purpose of the Beautiful Soup library is to extract data from any HTML or XML file which gives users the ability to extract different types of data from any website online ("Beautiful Soup Documentation" 2015). Webscraping was used to obtain the State Department scores, comments and ratings pertaining to the busiest airport in a country from skytraxratings.com, Reddit and Twitter comments on current events both in the airport and in the country, and some of the indicators for the risk model discussed in Section 3.1. For this portion of data collection to work, assumptions were made that comments and ratings were provided by rational actors and that conditions in the busiest airport of each country could be generalized to every other airport within that country assuming the busiest airport is the first to receive any technological advancements.

Webscraping was utilized to populate the "State Department Score," "Sentiment Analysis," and "Current Events" portions of the homepage of the qualitative analysis section. To populate the "Crime" portion, the World Values Survey was used which is a survey issued to each country that displays individual survey results on a variety of values-based questions ranging from the years of 1981 to 2022 (World Value Survey Contributors, 2022). The survey results regarding the different behaviors people exhibited for security, like not carrying money, avoiding going out at night, and even carrying a weapon, were used to assess the level of crime in a country and how that has changed over time, while using the United States as a baseline. This assumes that the higher the proportion of people in a country that exhibit these behaviors, the higher the level of crime within the country.

As previously discussed, our webscraping process can compile large amounts of data in string form that can then be used to infer the general opinion of a population. This data, drawn from Reddit, Twitter, and skytraxratings.com, allows us to populate recent firsthand user experiences to create a clear picture. The collected data can then be analyzed using sentiment analysis. Sentiment analysis characterizes a string by comparing the words in a sentence with a python library assigning each word a positive or negative value. Although an imperfect characterization of large bodies of text, we employ sentiment analysis across our data to illicit a generalized sentiment from user-generated content. Our analysis produces a score from -1 to 1 (negative to positive) that informs the decision maker on the public's feeling towards an individual airport or country. The decision maker can then use this score to create a more in-depth operating picture of the destination.

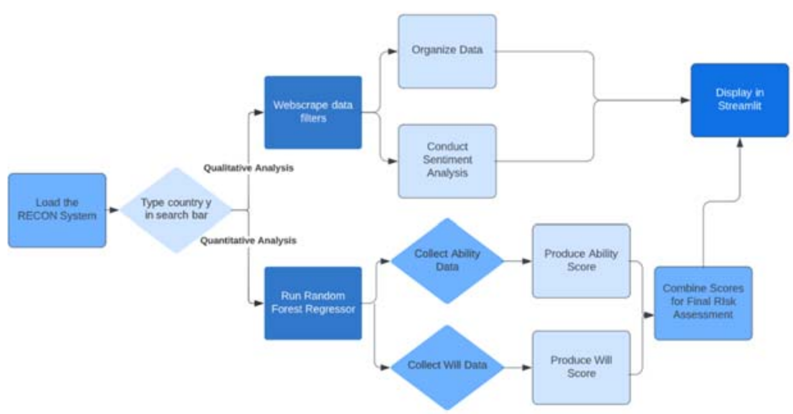

Figure 2. RECON Model Process Flow Chart

4. Results

4.1 Quantitative Results

The first portion of the quantitative analysis was training a predictive model for a proxy measure for ability. The sample size of this dataset, after cleaning and removing certain predictors, was 1036 (n = 1036). To select a model, the root mean squared errors (10-fold cross validation) of a decision tree regressor, random forest regression, and a multiple linear regression were compared. The highest performer on the training data was the random forest regression, so that model was optimized using a grid search cross validation technique to find the best hyperparameters of the model. The resulting model was exported for further use. The second set of results of the quantitative analysis was a predictive model for a proxy measure

ISBN: 97819384964-4-8

for will. The proxy measure that was chosen was the Our World in Data Human Rights Distribution score from 0-1. The sample size of this data set is 2975 (n = 2975). The same model selection techniques were used, and the highest performer on the will data was the random forest regression, in which the parameters were also optimized. The resulting model was exported for further use.

Table 2. Root Mean Square Error for All Tested Models

	RMSE for Ability Model	RMSE for Will Model
Multiple Linear Regression	6.940	0.109
Decision Tree Regressor	5.440	0.044
Random Forest Regression	1.463	0.043
Improved Random Forest Regression	1.417	0.043

Both models were then evaluated on 20% of the dataset. There were no major issues in either model with overfitting, but there is a minor limitation in the will model. This model performs very well as the output ranges from 0.8 to 1.0, and the model tends to overpredict in values below 0.6. This is most likely attributed to the null values within the dataset. To preserve the data, the null values were imputed with a 0. After there were two fully trained models for will and ability that were ready for implementation, the models were exported. Algorithms were created to search the will and ability dataset for the most recent data point so that we have the most accurate risk picture as possible. That data is then passed through the trained will and ability models to pass out a raw score for both models. The raw scores for will (0-1) and ability (0-70) are mapped from 0 to 100 so it is compatible with our risk model. In the case of Belarus (Figure 3), the most up to date risk and ability scores were scaled from 0 to 100 and then plotted onto our risk model. Future work may include writing algorithms for imputing null values with a value from a similar row (k-nearest neighbors), writing web scraping algorithms to gather the most up to date information on a country, or finding other methods for countries that have no data.

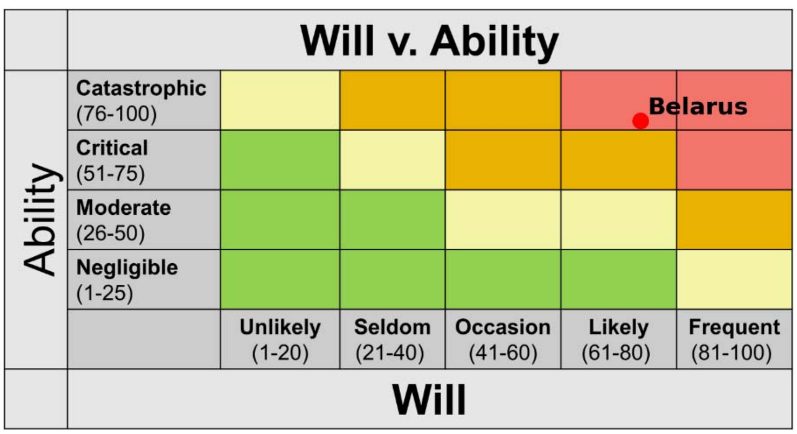

Figure 3. Risk Matrix for Belarus

4.2 Qualitative Results

The front-end dashboard used to visualize our results employs Streamlit to integrate the different components of the above methodology section into a single place for commanders to view to make risk-informed decisions regarding sending personnel abroad. Streamlit is an open-source framework to create easily navigable website applications using Python script (Streamlit Contributors 2023). A key factor in selecting this program to use for the dashboard is due to its ability to automate data collection by collecting user input which allows the user to type in the desired country to view. Streamlit provided our research with the capability to collect user input and change the display of data on the local web application dependent on the user's country of interest. On the home page, the composite risk assessment matrix will be displayed with the corresponding risk score given to the country based on the analysis of the country's will and ability. The State Department score, sentiment analysis, current events, and crime rating will also be displayed. Additionally, there is the option for users of the web application to download the datasets displaying the specific information if they choose to gain more insight. Finally, individual travelers can fill out an individual traveler section on the dashboard to answer basic questions regarding their traveler to include its

ISBN: 97819384964-4-8

purpose, whether the Soldier is going alone or with a group, and the duration. This questionnaire allows commanders to understand the travel situation for increased awareness. If you would like to see the application, contact the author personally.

5. Conclusion

Improved technological capabilities worldwide has made it necessary for U.S. special operations to understand their digital signature and its impact on the mission and the force. The assessment of a country's will and ability to seek out, distinguish, and utilize an individual's data signature against them helps inform a commander's ability to understand the risk to the mission and the force by that individual traveler. The final Streamlit dashboard developed gives commanders a holistic understanding of this potential risk. This has a greater utility than risk scores already developed, like that of the State Department Travel Advisory score, because this dashboard gives commander's a more in-depth understanding of the current state of a country in various areas to include airport security, crime levels, and even current events, while providing a separate risk score dependent on the country's will and ability. Further refinement of this dashboard involves including country relationships with one another and airport interrogation procedures should a Soldier be questioned in the qualitative analysis portion of the homepage. Additionally, it would be helpful to collect historical data on how the quantitative risk model performs to evaluate its performance and include additional indicators for model improvement. The creation of this dashboard is a clear step towards achieving an Army that possesses a greater understanding of risk to the mission and force as influenced by individual Soldiers traveling to a country with a unique will and ability.

6. References

Axhausen, K., Batty, M., Giannotti, F., Pozdnoukhov, A., Bazzani, A., & Wachowicz, M. et al. (2012). Smart cities of the future. *The European Physical Journal Special Topics*, *214*(1), 481-518. doi: 10.1140/epjst/e2012-01703-3

Cheein, A., & Guevara, L., F. (2020). The Role of 5G Technologies: Challenges in Smart Cities and Intelligent Transportation Systems. *Sustainability*, *12*(16), 6469. doi: 10.3390/su12166469

Cheema, M., Cheema, O., Li, H., Lu, H., & Umair, M., (2021). Impact of COVID-19 on IoT Adoption in Healthcare, Smart Homes, Smart Buildings, Smart Cities, Transportation and Industrial IoT. *Sensors*, *21*(11), 3838. doi: 10.3390/s21113838

Diplomat, T. (2019). Smart Cities or Surveillance? Huawei in Central Asia. Retrieved 27 September 2022, from https://thediplomat.com/2019/08/smart-cities-or-surveillance-huawei-in-central-asia/

Department of the Army. (2014). Risk Management (ATP 5-19).

Herre, B., & Roser, M. (2016). Human Rights. *Our World in Data*. https://ourworldindata.org/human-rights

Indicator Rankings & Analysis. Global Innovation Index. Retrieved March 2, 2023, from https://www.globalinnovationindex.org/analysis-indicator

Kelber, N. & Lawless, T. (2023). Sentiment Analysis with Vader. *Creative Commons*.

Kitchin, R., & Dodge, M. (2017). The (in)security of smart cities: Vulnerabilities, risks, mitigation, and prevention. *Journal of Urban Technology*, *26*(2), 47–65. https://doi.org/10.1080/10630732.2017.1408002

Moran, M. (2023). Social Media Statistics (How Many People Use Social Media). *Colorlib*. https://colorlib.com/wp/social-media-statistics/

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), Article 5. https://doi.org/10.1038/s42256-019-0048-x

Streamlit Contributors. (2023). Streamlit. *Snowflake Inc.* https://streamlit.io/

Tidjon, L. N., & Khomh, F. (2022). *The Different Faces of AI Ethics Across the World: A Principle Implementation Gap Analysis*.

Unknown (2015). Beautiful Soup Documentation. *Leonard Richardson Revision*.

World Value Survey Contributors. (2022). World Values Survey. *World Values Survey Association*. https://www.worldvaluessurvey.org/wvs.jsp