IITD-INNOV8 Challenge
Predicting Troop Betrayal in the War Against the Phrygians

Vidit Aggarwal, Amritanshu Tiwari, Cherish Puniani, Ronit Chawla, Rishit Mathur

September 2024

Abstract

This study aims to identify soldiers at risk of betrayal in the conflict against the Phrygians by examining various stressors that can lead to disloyalty. Key factors include financial pressures, external influences, loyalty dynamics, psychological resilience, and sociocultural isolation. We quantify these influences using measurable metrics, such as debt-to-income ratios, spending habits, communication patterns, and disciplinary records. An autoencoder model will analyze these factors to predict potential betrayal. This approach seeks to enhance troop cohesion and address vulnerabilities, ultimately strengthening the military's operational integrity.

1 Financial Considerations

1.1 Debt and Bankruptcy

Hypothesis: Soldiers facing significant financial difficulties, such as debt or bankruptcy, may be more vulnerable to enemy offers of financial security. Financial desperation can lead to irrational decision-making, making soldiers more likely to prioritize personal survival over loyalty to their unit. This financial instability creates an environment where betrayal might seem like the easiest escape.

Quantification: Measuring the Debt to Income Ratio: a high ratio is indicative of a higher risk of default.

1.2 Recent Spending Habits

Hypothesis: Excessive or unusual spending patterns can indicate financial stress, greed, or even the reception of a large amount of money, which might be some form of bribe. Soldiers indulging in lavish or risky spending may be seeking fast monetary gain, which makes them prime targets for enemy incentives.

Careful monitoring of financial behavior can help identify those who might prioritize wealth over loyalty.

Quantification: The quantification of spending habits is a bit tricky and would require measuring outgoing transactions from the soldiers' accounts. Here, we measure this by calculating a monthly average spend of the soldiers over certain time periods, such as the previous 12 months and 24 months. A significant deviation of this number from the average is indicative of malicious income sources.

1.3 Gambling or Betting Habits

Hypothesis: Engaging in gambling reflects a willingness to take significant risks for personal gain, which could translate into a higher susceptibility to enemy offers. Soldiers who are involved in gambling may experience financial loss or addiction, pushing them towards betrayal if they believe it could restore their financial situation. This risk-taking mindset is often exploited by adversaries seeking to lure individuals with quick, substantial rewards.

Quantification: Here, the risk of betrayal is directly proportional to gambling and/or betting habits. It's a simple binary classification since more detailed information on this is bound to be inaccurate and very impractical to obtain.

2 External Influence

2.1 Frequency of Communication with External Entities

Hypothesis: Frequent contact with individuals or groups outside of the military, particularly those with potential connections to the enemy, can be a strong indicator of vulnerability to betrayal. Such communications could involve subtle persuasion, promises of wealth, or threats that sway soldiers towards defection. Tracking these interactions can provide insight into external pressures influencing a soldier's loyalty.

Quantification: This can be broken down into two kinds of contact - digital and face-to-face.

- **Digital:** For measuring suspicious digital contact, we can simply monitor the number of messages or calls received by the person that were from an unverified or unidentified source. (This type of data is generally maintained by all government agencies and already monitored, so no new endeavors need to be made to collect this data.)
- **Offline:** Offline communications can only be indirectly monitored by getting the number of hours a soldier spent outside of his/her base.

3 Loyalty and Discipline

3.1 Substance Abuse

Hypothesis: Substance abuse can impair judgment, increase stress, and lead to financial instability, all of which might push a soldier towards defection.

Quantification: This is a simple binary quantification with yes/no (1/0) values.

3.2 Failure to Follow Orders

(Disciplinary records, infractions, conflicts with superiors)

Hypothesis: Soldiers with a history of poor discipline, such as frequent infractions or disregard for authority, are more likely to betray their unit. Consistent disciplinary issues often reflect deeper dissatisfaction with leadership, which adversaries can exploit by offering a more favorable position. This lack of respect for military hierarchy can make these individuals more susceptible to the enemy's promises of power or autonomy.

This factor should also include suspicious digital activities, such as receiving unsolicited emails, transferring sensitive data to unauthorized accounts, tampering with records, or introducing malicious code, all signaling potential insider threats. These behaviors suggest attempts at espionage, sabotage, or preparation for defection. Soldiers involved in these activities may already be in contact with the enemy or planning betrayal. Tracking digital footprints is crucial to flagging potential defectors before they cause significant damage to the unit's operations or security.

Quantification: The weight of this factor is dependent on the number of offenses and the severity of each offense. Here, we assign weights to each class of infractions, let's say 1 for the least severe infractions and higher integers for more severe ones. We then calculate the weighted sum of the offenses and severity of offenses to get our factor:

$$f = s \cdot n$$

where f represents the factor, s represents the severity, and n represents the number of offenses.

3.3 Previous Incident Reports (Criminal Proceedings)

Hypothesis: Soldiers with criminal backgrounds or past legal issues may be more willing to defect in exchange for protection or rewards from the enemy.

Quantification: Since outside of army court, the severity of offense is not

generally classified categorically, we only count here the number of criminal proceedings against the soldier recorded outside of the army. A significant number of proceedings is a serious indicator of the aggression and unfaithfulness of the soldier.

4 Psychological Resilience (Mental Health and Stress)

4.1 Evaluation through Psychometric Test Scores

Hypothesis: Low psychological resilience, as measured by psychometric testing, can indicate a higher susceptibility to stress, which in turn makes soldiers more vulnerable to enemy manipulation. Soldiers with weaker mental health may struggle to cope with the pressures of military life, making them more likely to defect under intense psychological warfare from the enemy. Also, betrayal to one's kind is a very big event that is bound to take a toll on an individual's psychological health.

Quantification: This factor is quantified by considering a stress evaluation exam, e.g., the standard Perceived Stress Scale (PSS), to calculate the stress level of the soldier. We will then use this number to calculate the percentile rank of the soldier among other soldiers and convert it into a number between 0 and 1. For example, a soldier with stress higher than 95.58% of the soldiers would have a score of 0.9558 in this quantifier. We can have multiple variables indicating a similar score of various psychometric tests to soundly cover the psychoanalysis of the soldier.

4.2 Recent Traumatic Experiences (Combat Exposure, Personal Losses)

Hypothesis: Traumatic experiences, whether from combat or personal losses, can lead to psychological breakdowns, leaving soldiers feeling hopeless or disillusioned. In such states, soldiers may become more susceptible to enemy propaganda that promises relief from their emotional pain. This vulnerability can increase the likelihood of betrayal as they seek a way out of their distress.

Quantification: This can also be determined as part of the psychometric evaluation mentioned above.

5 Sociocultural Isolation

Hypothesis: Soldiers who experience social or cultural isolation may feel disconnected from their peers or the military institution. This isolation can stem from differences in ethnicity, religion, or political views that make the soldier

feel alienated or marginalized. When soldiers feel like outsiders, they are more vulnerable to enemy propaganda that plays on these divisions, increasing their likelihood of defection.

Quantification:

- **Cultural/Religious Affiliation:** Track the soldier's participation in cultural or religious groups and whether they feel marginalized. Surveys or interviews can be used to gauge their sense of belonging within the military community.
- **Social Network Size:** Measure the soldier's social network size and quality of interactions (number of meaningful connections within the unit). Social isolation can be quantified by the ratio of social interactions to average unit interactions.

6 Ideological or Political Disillusionment

Hypothesis: Soldiers who become disillusioned with the political or ideological stance of the military or governing body may experience internal conflicts that could lead to betrayal. This disillusionment can be sparked by policy changes, leadership corruption, or a clash of values, making soldiers susceptible to enemy narratives that promise alignment with their personal beliefs.

Quantification:

- **Surveys on Political Alignment:** Conduct surveys to determine the soldier's alignment with the political or ideological views of the military. Soldiers expressing strong dissent may be at higher risk for betrayal.

7 Opportunism (Risk vs. Reward Perception)

7.1 Security Clearance and Information Access

Hypothesis: Soldiers with access to sensitive or classified information may perceive the rewards of defection as outweighing the risks, especially if the enemy promises significant wealth or protection. Their unique position gives them leverage, and the temptation to exploit it for personal gain increases their susceptibility to betrayal. Additionally, such soldiers are at an elevated risk of outside persuasion due to the vast array of information they bring with them.

Quantification: This can also be measured on a percentile basis, which will then be converted to a variable from 0-1. For example, if a person has clearance to information that only the top 5 percent can view, he will receive a score of 0.95 on this scale.

8 Data Collection Model

We have put careful considerations into the data points we are collecting. Extensive effort has been made to ensure that most of the data points are already a subset of the standard data collected and maintained by any army. This makes the system effective, lightweight, and easier to implement, thereby making the overall project scalable.

The following types of data are being collected:

8.1 General Information

- **Enrollment ID:** The enrollment ID or identification number of the army personnel.
- **Psychometric Test Score:** We consider the score of a psychometric test such as PSS. This type of test is generally conducted during the recruitment of personnel and is also mandatorily performed at fixed intervals, making it an easily obtainable parameter.
- **Experience:** The experience of the personnel in years.
- **Security Allowance:** Security access of the personnel measured as a percentile score.

8.2 Financial Information

- **Income:** Income of the personnel
- **Debt:** Debt thats is currently pending for the personnel
- **Previous 1 year expenditure:** Last 12 months expenditure of the personnel
- **Previous 2 year expenditure:** Last 24 months expenditure of the personnel

8.3 Behavioural Information

- **Known Gambling Habits:** Yes/No column field
- **Known Substance Abuse:** Yes/No column field
- **Average amount of hours spent outside of army camp:** Average numbers of hours spent outside the army camp measured over a period of let's say a week'
- **Number of Unknown/Unverified Calls/Messages**

This is the part of information that is relatively harder to access or maintain with complete reliability. Assuming that vigilance and surveillance are strict due to the impending sense of war, we hypothesize that a fair estimate of these factors (e.g., inspection of call records, tracking movement of personnel, etc.) can be established.

8.4 Past Offences

- **Insubordination:** Records of any insubordination acts performed by the personnel.
- **Civilian Offences:** Records of any civil offences committed by the personnel.

9 Data Preprocessing

Assuming that each category of data is maintained in a separate database table, we first need to combine the tables to obtain the required data. We combine the data tables based on the unique enrollment ID of each personnel to obtain a data table containing all the necessary 14 factors considered.

Here, we build our dataframe by calculating the quantitative measures from the above-collected factors for each personnel. The following quantitative factors are calculated:

- Enrollment ID
- Debt to Income Ratio
- Spending Habit Change
- Number of Unverified Calls/Messages
- Hours Spent Outside Army Camp
- Insubordination Acts
- Committed Civil Offences
- Psychometric Test Score (PSS Score)
- Experience
- Security Allowances
- Gambling
- Substance Abuse

Data is checked for any duplicate rows and missing values. Once these are removed, our data becomes ready for passing onto the insider threat classifier model.

10 Model

For our threat identification model, we use an autoencoder to generate a top-k list of army personnel that are most likely to defect to the Phrygian forces.

10.1 Working of Autoencoder

Autoencoders are a type of neural network designed for unsupervised learning, primarily used for dimensionality reduction or feature extraction. They work by compressing input data into a lower-dimensional latent space (encoding) and then reconstructing the data from this compressed representation (decoding).

An autoencoder consists of three components: the encoder, the latent space, and the decoder. The encoder maps the input data $\mathbf{x}$ to a latent representation $\mathbf{z}$ using a function $f(\mathbf{x}; \theta)$, where θ represents the network parameters. For our purposes, we use two dense layers for encoding. Another two dense layers are used for decoding, which then reconstructs the original data $\hat{\mathbf{x}}$ from $\mathbf{z}$ using another function $g(\mathbf{z}; \phi)$, with ϕ as its parameters.

The network is trained by minimizing the difference between the original input and the reconstructed output, using mean squared error (MSE) loss:

$$\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|^2$$