**Document Overview:**

To develop a solution for organizing intelligence data available from multiple agencies in structured/ semi-structured/ unstructured forms and correlate to event matrix under big data framework alongwith tools to analyse patterns for predictive anaylsis. The application should be capable of identifying triggers in terms of identified activities which may lead to an impending event. This document states one method of doing so by putting forward an automated machine learning solution to overcome the challenge faced by the militaries in terms of overflow of data and no sufficient manpower or processing power to do the same. Thus machine learning can be implemented to automate the categorising of military data and prediction of certain events and alert system for the same. This can also be used to predict the outcomes of strategies and thus act as a simulator for certain events and help in making counteracting strategies. Thus this trained model will be able to provide strategic, predictive alarming support and further increase the relevance of the large amounts of military data by parsing it through an automated product which will not ignore a single detail.

With this paper, we intend to further the data analytics and collection departments of military systems and make advances in the field of information and cyber security. This document will lay some light on the approach and methodology to look at battles as a whole according to microinstances and thus summing up many algorithmic decision.

**Background:**

Terrorism prevention and military strategy prediction is something which has been done in conference halls and meetings for a long time. From simple recon scout soldiers in the Roman Empire to collect information about positions and numbers of the enemy, we have come a long way to United States' unmanned aerial drone, the AAI RQ-7A/b "Shadow", which has the capability to root out enemies and launch live missiles which can be directed to hit the optimum location for maximum destruction/disability. The Shadow has an additional functionality with 12 cameras which are used for data collection. In an age of big data we can use exabytes of data to enhance our knowledge with machines which can process this data and give us useful results from all these estranged facts.

Besides having so much data, we now possess processing capabilities which surpass our strongest machines. Thus we have our own supercomputers to process this humongous amount of data in a way we see fit and fitting all this data to model certain results is the next step. While these methods are already in employ, there are only a few which actually embrace the military strategy prediction stage. This is because it requires a lot of privileges into data and thus security is a major issue. It is also imperative that the model predicts with high accuracy since these scenarios are crisis situations and thus will require solutions which are accurate. While faster a committee deciding a situation, the question arises over accuracy of models. Earlier algorithmic models consisted of linear regression models which use a brute force mechanism on all microinstances which may occur over a battlefield. However, with the oncoming of Artificial Neural Networks and further, Deep Neural Networks, it is possible to now see the bigger picture and thus have more control and accuracy over crisis situations.

**Project Contacts:**

Our Team :

1. **Akshay Ramdev**

Phone : +91-8698618646

Email : akshay.ramdev@yahoo.com

2. **Akash Ramdev**

Phone : +91-8698618642

Email : akashramdev@gmail.com

3. **Rohan Datta**

Phone : +91-8698574158

Email : rohandatt@gmail.com

4. **Saloni Verma**

Phone : +91-7768933872

Email : sal.verma48@gmail.com

**Time Frame:**

The project work shall be divided between four people and will have a time period of 8-12 weeks. The work shall be ditributed over workforce and time as follows:

Week 1(Tentatively from May,28th,Monday)

1. Collection of data using scrapy spider starts. This is basic data gathering stage and will take 2-3 days. All data gathered will be further analysed.

2. Data will be now analysed and filtered superficially to ignore all external misguiding references. Meanwhile a python/octave script to be written to start Data Preprocessing.

3. According to data abundancy and credibility, algorithm to be run to be decided(simple regression,SVM,etc.).

Deliverables of the week: Rough data metrics to be given out along with tentative algorithm selection

Week 2(4th June)

1. The collected data is now organised and will be preprocessed to rectify missing values, invalid data and nonsense data.

2. Further, the entire dataset is assembled and preprocessing starts to rectify all data.

Deliverables of the week: Preview of dataset and tentative features selection.

Week 3(11th June)

1. Previous renderings of the idea are searched once again to further enhance own model

2. Nodes of the DNN/RNN or weight training of the regression model are now found out by training on the cloud.

Deliverables of the week: Any improvements made are reported and results of training are shared.

Week 4(18th June)

1. More training of models on the cloud is carried out. DNN model is trained against itself with certain features. Overfitting possibilities are taken into account and revision is carried out for the features.

Deliverables of the week: Any change in features or any overfitting cases are reported. Node structure and flowchart is drawn out.

After week 4, all weeks are tentative and thus deliverables are TBD.

**Approach and Methodology:**

There are significant potential benefits to developing predictive methods of defending against adaptive adversaries, in which opponents evolving strategies are anticipated and these in-sights are employed to counter novel attacks. This section considers the following concrete instantiation of the predictive defense problem, given some history of attacker actions we will design a defense system which performs well against both current and future attacks.

All security tasks can be put into five categories: Prediction, Prevention, Detection, Response and Monitoring.

Our Project will help to include :

1.Regression:

Regression, or in other words prediction, is a simple task.The proposed approach to designing a predictive defense system which works well against both current and future attacks. We want to utilize our knowledge about existing data to make opinions on new data.In Defence, it can be implemented to predict the stratergies of the Enemy force and is capable of identifying triggers in terms of identified activities which may lead to an impending event.The difficulty of deriving realistic representations for attacker behavior is overcome by recognizing that the actions of attackers that can be modeled as attempts to transformdata in such a way that malicious and innocent activities are indistinguishable. It is possible to model attacker actions as transformations of data because, within the problem formulation, historical attack data are available in the form of training instances. Therefore, we are predicting the attacker strategy through computation of attack and then learning an appropriate countermeasure by applying ML to the transformed data.

2.Classification:

Here, based on the activity reported to the model, the model has to predict the classification of the model. Therefore it will classify actions and cases in different classes. For example: Invasive tactics, Aggressive motion, Neutral activity, Diplomatic stance change, etc. These classes will further be analysed by the model to give us an idea of what motives could be hidden between these actions. These motives per each mocroinstance will be recursively carried on to each next recursion to determine an overall motive using pushback on a neural network

3.Clusterization:

In this stage, to prevent independencies of events and combining all events to make sense in a broader perspective, a solution will be brought about to find the overall class of a series of events to be passed on to the next recursive call as a background motive. Thus, clusterization will be very necessary to maintain an overall perspecive and not overfitting for microinstances. As military operations are a combination of more than one microinstance, these microinstances need to be grouped together and thus clusterization comes into effect.

4.Recommendation:

It can be used primarily in incident response. If a defence force faces a wave of incidents and offers various types of responses, our system can learn what type of response it should recommend for a particular incident. Risk management solutions can also benefit in that they automatically assign risk values for new vulnerabilities or misconfigurations built on their description.

We can solve all of them with different levels of efficiency for various needs.

Besides these methods, we will be implementing Deep Neural Networks to gain an overall perspective and further allow pushback between the nodes of the network. This will enable us to enhance our capabilities as we will not be checking all possibilities but will already havea list of likely possibilities from the first neural layer which will further be shortened to another likely layer in the second neural layer and so on. This DNN will be further implemented using Deep Learning libraries such as Tensorflow to refine our layers. We will have various filters to give us our final prediction and each will be defined with pushback and convolutionary layers to give us a broader perspective. In the end, clusterisation will take place to average out the findings and push it further into another iteration with new variables in concern.

In our case, we will give the model a few initial sets in a hypothetical battle which may have taken place long ago. The factors(features) to be included will be: Media, food, siege status, arsenal, political status, form of government, etc. These will all be taken into effect to give a good broad range of coverage for the model. These initial few training runs will be done manually with given data. Next, after finalising on the test data, we will now make an algorithm to randomise values and give them to the model for processing along with an assigned label of win/lose. These will act as microinstances to make the model realise the classification of actions and further dwell on an action to perform in retaliation. For each action there will be an assigned label. Thus, the model will learn on itself by practicing against itself and will therefore create a good training ground for itself like the IBM Watson which had trained against itself and then the Alpha Go model which had against reduced the billions of possibilities to a few likely possibilities. Thus the model will train and after putting it through a test phase, we will have a model ready to be deployed and tested in dummy runs for the military.

**Work Breakdown:**

**Akash Ramdev** - Preprocessing of initial dataset and deeming inappropriate entries, missing values, etc. and refinement of NLP output model.

**Akshay Ramdev(Co-ordinator)** - NLP filtering to find ways to output data in a paragraph form to make it understandable to the reader. This will be after model has been trained

**Saloni Verma** - Spider deployment using python (scrapy) and data collection and basic analysis to deem a sample worth of use. Alongside, working on the neural network for refinement of nodes.

**Rohan Datta** - Creation of a deep recursive neural network or support vector machine depending on requirements and availability of data.

Front end and presentation to be equally divided among the members.

**Project Risk Management**

Risks:

1. Outdated or incomplete information may persist due to the cost and/or effort of obtaining up-to-date information

2. Malicious parties may corrupt data (for example, cybercrime activity that alters data and documents)

3.An organization that uses multiple data sources may incorrectly interweave data sets and/or be unaware of causal relationships

between data points and lack proper data governance mechanisms to identify these inconsistencies.

4. There may be different versions of the same instance in time. As the victor writes the history. This may alter results. Therefore standardised data shall be used.

5. Due to high number of features, many iterations will be required thus a larger time frame has been alloted to make up for the same.

6. Data preprocessing to make actual usable data will take time as data available is majorly in the form of text and thus pre existing NLP models will be sought out to make sense of the text into data form.

**Conclusion:**

Machine learning has been gaining popularity immensely over the years. Combining huge amount of data base with the machine is helping in great technological advancements. Likewise, creating a program that challenges the designers in creating an adaptive AI opponent that provides a war strategy and focuses on tactical and strategic aspects of war play, can be majorly applied for strategic predictions. There are substantial potential benefits to considering predictability when designing defenses against adaptive adversaries, including increasing the ability of defense systems to predict new attacker behavior and reducing the capacity of adversaries to anticipate defensive actions. This paper adopts such a perspective, leveraging the coevolutionary relationship between attackers and defenders to derive methods for predicting and countering attacks and for limiting the extent to which adversaries can learn about defense strategies. These results basically indicate that the algorithms have foresight of the opponent’s strategy. Given this ,the machine learning algorithms outperforms the opponent.