BACHELOR COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE

AI: MACHINE LEARNING FOUNDATIONS

ARMA: PROACTIVE CYBERSECURITY MACHINE LEARNING MODEL

Baptiste Etroy, Jaskaran singh ghai, Fynn Frühling, Adrian de la Torre, Larbi Benamour, Juan Diego Fernandez

Executive Summary

In our **introduction** of the project, we will delve into Cybersecurity challenges as well as its limitations, and present our AI-driven Solution. This Machine learning model will oversee all network traffic, and interfere whenever a malicious connection, transaction or operation is detected. Not only does it identify the threat, but it also reacts accordingly and assigns a mitigation recommendation based on the detected danger.

Our **value proposition** focuses on keeping a reliable service without compromising any scalability or adaptability. We aim to meet the needs of small businesses which face cyber threats by implementing specific anomaly detection machine learning models (decision trees, neural networks…) in network traffic.

If we dive into a more **technical description** of ARMA, we use advanced machine learning models, including decision trees and Random forests for building predictive models and neural networks for complex anomaly detection for maximum security. Our data strategy uses multiple datasets to enhance training and model performance.

We displayed the **project's implementation** through a step-by-step development roadmap. We run through the different phases, from the initial development phase to the full deployment with a structured plan for model training and regular updates. We made sure to highlight any potential risks with their respective solutions to increase the security of our project.

Our **business model** is based on 3 main concepts: a subscription service, consultation and the sale of our ML model. Small companies and startups will be able to subscribe to our service, in which case we will be overseeing all network traffic, interfering when malicious, will also consult the business to improve their system. Otherwise, companies will have the choice to buy our model and implement it on their own.

For our **market strategy** we will aim to portray ARMA as the leader of network security solutions. To achieve this, we will focus on strong digital marketing and establishing strategic partnerships to hopefully create a presence in key cybersecurity expositions. In the case of future expansions we will explore new markets and incorporate advanced AI capabilities.

In **conclusion**, ARMA will serve as a strong cybersecurity protection model for businesses in need of it. Not only do we identify the threats but we give out recommendations to counter them and better your software.

Introduction

Evolving Cybersecurity Challenges

In a world where digital connectivity is so essential in the business scene, cyber attacks are becoming increasingly more common. With cyber-attacks improving on a day-to-day basis it is hard for traditional security measures to keep up, these factors leave various companies and individuals very vulnerable.

Limitations of Conventional CyberSecurity Approaches

Traditional cybersecurity solutions are mainly based on known past threats and, therefore have problems adapting to new attacks. In some cases, only when hacked entirely can the software identify the threat. For example, Firewalls and antivirus packages may only recognise new or complex attacks long after they have bypassed perimeter protection. This slow reaction leaves organisations vulnerable and exposed to data breaches and server network disruptions.

Introducing ARMA: Proactive, Intelligent Network Defense

Introducing ARMA, a change in cybersecurity. By integrating advanced machine learning algorithms, we built proactive and intelligent models capable of providing real-time monitoring and threat detection. It learns continuously from network traffic, making it able to detect new anomalies, thereby predicting attacks, and neutralising them. It also provides descriptive alerts for detected anomalies, to help improve the software.

Objective of ARMA

Our objective is to transform the role of network security systems to further fortify network defences. Our implementation of machine learning adds a factor of intelligence to our system which is capable of predicting potential threats and dealing with them, therefore preventing future problems. This approach is much safer than the typically used passive technique.

Value Proposition and Customer Insights

Customer Motivation

Small businesses and startups are often victims of cyber attacks due to vulnerabilities in their network systems and security. These companies are in the midst of growth, they're concentrated on building their idea rather than protecting it, moreover they lack the resources to implement cybersecurity software, making them a perfect target. That is why, the typical customers for ARMA are said startups, in need of a quick, robust, efficient network security solution that not only protects their data but indicates how to improve based on the attacks it detected.

Their primary motivations include:

- **Security Assurance:** To secure their business data, customer information and transactions from any exterior threats.
- **Cost-Effectiveness:** To manage these cyber attacks without the need for extensive IT infrastructure, which can be expensive.
- **Scalability:** To have a solution that grows with their business, adapting to new challenges as they expand.

Solution Overview

ARMA addresses all these needs by providing a machine-learning cybersecurity service that continuously monitors network traffic to identify and respond to potential threats and attacks. The service includes:

- **Real-Time Monitoring:** Thanks to our algorithms, ARMA analyses the network in real-time and passes every operation through its ML model that correctly identifies the danger of said operation.
- **Automated Threat Detection and Response:** Upon detecting a threat, ARMA neutralises it and stops further communication with this address.
- **User-Friendly:** The service also works in a user-friendly fashion, ensuring that every threat detected is followed by a descriptive alert of the attack and a mitigation recommendation.

Competitive Advantage

ARMA offers many advantages on top of what traditional cybersecurity services and VPNs already offer:

- **Speed:** The machine learning algorithms we implemented in ARMA detect anomalies much faster than any traditional system would since they usually work with signature-based detection which can lag behind the latest threats.
- **Accuracy:** With ARMA made to continuously learn from network traffic we can reduce the amount of false positives and increase the likelihood of pinpointing actual threats.
- **Adaptability:** ARMA’s constant adaptation to the digital environment allows for security measures to evolve in correlation with new business developments and threat vectors.
- **Cost Efficiency:** Thanks to the self-learning algorithms we minimise the need for a large cybersecurity team therefore making it much more viable for smaller businesses with smaller budgets

Technical Description

As mentioned a few times, ARMA uses machine learning to enhance network security by detecting malicious activity and providing alerts and mitigation. Our Decision Tree Classifier was trained with two individual datasets, initially using real network traffic, then transitioning to synthetic randomised data to improve adaptability. This synthetic model now approaches overfitting and it improves the classifier's ability to recognize suspicion.

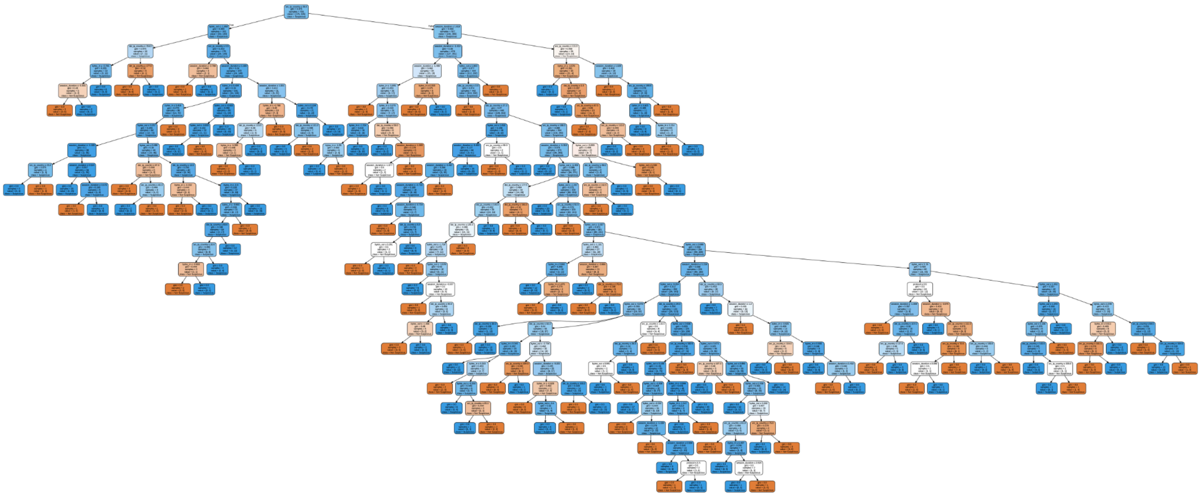
ML Model Selection

We employ Decision Tree Classifiers for their efficiency in handling both numerical and categorical data, they’re crucial for analysing network traffic which consists of diverse amounts and categories of data types such as IP addresses, byte sizes, geographical locations, and more.

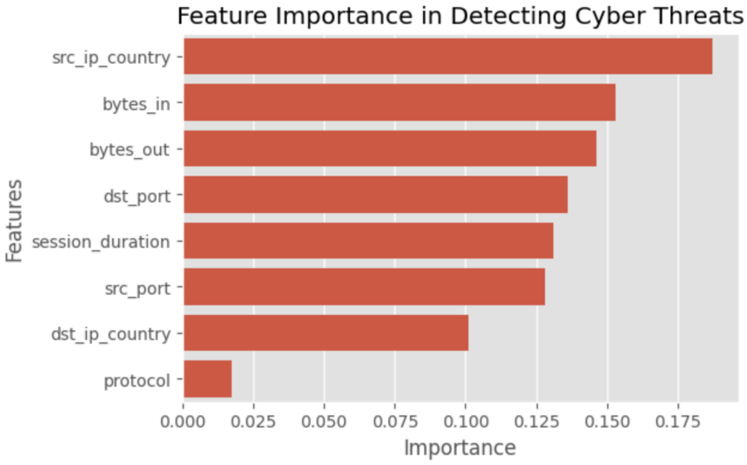
Decision Trees were chosen for their:

- **Interpretability:** The ability to visualise decision paths makes it easier to understand model decisions, which is very useful when applying cybersecurity. Also easy to verify decisions taken by the model.
- **Adaptability:** Easily handles changes in data distribution, a common scenario in network security.

Here is the decision tree we trained to identify malicious networks:



Feature Importance Analysis: In our notebook, we added a section on feature importance; it ranks every dataset feature based on its influence on the model's predictions, allowing us to focus on the most impactful features for anomaly detection. As seen below, the source of IP addresses is the most important feature with bytes_in in second place.



Data Strategy

Our data strategy involves:

- **Initial Data Training:** At first we used a cybersecurity dataset found on the internet(Kaggle), but it was too monotone, already cleaned, with the malicious IP addresses already identified in the content.

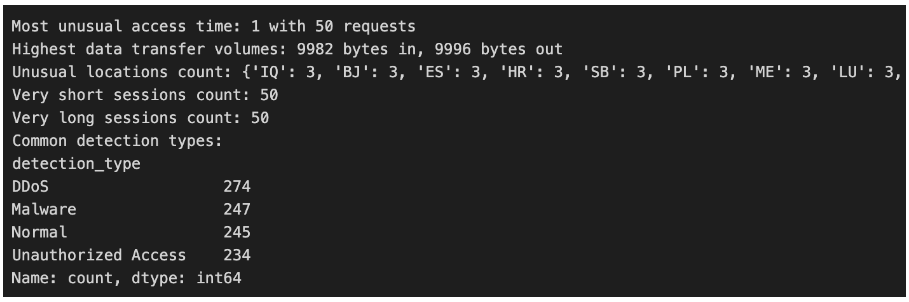
- **Synthetic Data Enhancement:** To counter this homogeneity in the initial dataset we made our own dataset using the Faker library, introducing more complex scenarios, and randomised values that the model might face in real-world applications, thereby enhancing its ability to predict threats.

Data Processing: To make sure our data was ready to be used we performed different steps of preprocessing such as normalisation, outlier detection, and encoding categorical variables to prepare the data for machine learning.

System Architecture

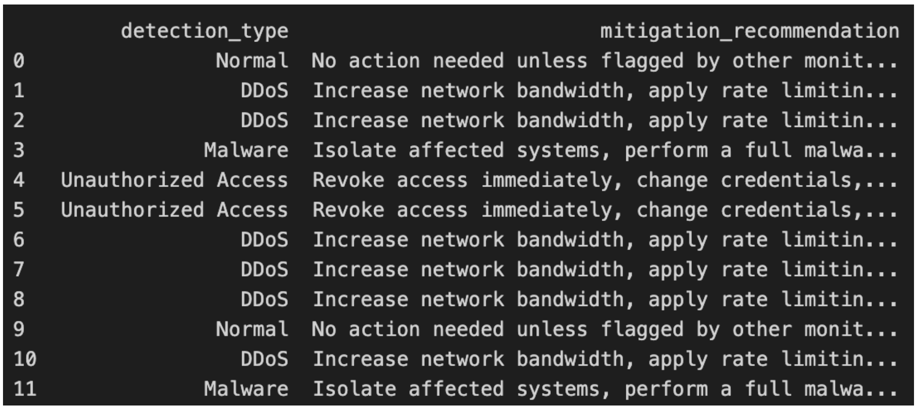
The system architecture contains loads of different sections, some we oversaw, like the data collection modules that capture and preprocess the network data, the analysis engine that uses the trained decision tree classifier to evaluate traffic and anomalies, and also the alert system that notifies users of potential threats. But we've also implemented the following algorithms:

- **Enhanced Security Analysis:** This is an additional monitor for extreme data transfers, unusual access times, and geolocation analysis to pinpoint and alert on sophisticated security threats.



```
Most unusual access time: 1 with 50 requests
Highest data transfer volumes: 9982 bytes in, 9996 bytes out
Unusual locations count: {'IQ': 3, 'BJ': 3, 'ES': 3, 'HR': 3, 'SB': 3, 'PL': 3, 'ME': 3, 'LU': 3,
Very short sessions count: 50
Very long sessions count: 50
Common detection types:
detection_type
DDoS                   274
Malware                247
Normal                 245
Unauthorized Access    234
Name: count, dtype: int64
```

- **Mitigation Recommendations:** An automated algorithm that gives out suggestions based on the type of anomaly detected, making sure that each threat is met with an appropriate response, like such:



```
         detection_type                          mitigation_recommendation
0                Normal  No action needed unless flagged by other monit...
1                  DDoS  Increase network bandwidth, apply rate limitin...
2                  DDoS  Increase network bandwidth, apply rate limitin...
3               Malware  Isolate affected systems, perform a full malwa...
4   Unauthorized Access  Revoke access immediately, change credentials,...
5   Unauthorized Access  Revoke access immediately, change credentials,...
6                  DDoS  Increase network bandwidth, apply rate limitin...
7                  DDoS  Increase network bandwidth, apply rate limitin...
8                  DDoS  Increase network bandwidth, apply rate limitin...
9                Normal  No action needed unless flagged by other monit...
10                 DDoS  Increase network bandwidth, apply rate limitin...
11              Malware  Isolate affected systems, perform a full malwa...
```

- **Practical Application - Transaction Monitoring:** We also did a practical application in a real-world scenario. The model was used to monitor financial transactions.

Project Plan and Implementation

Development Roadmap

Initial Development: In this phase we focused on setting up the foundational elements of the ARMA system. This includes integrating the development of the initial machine learning model architecture into an existing network infrastructure of a small business. We will also involve the following:

- **Setup of Data Collection Mechanisms:** Establishing the processes for continuous data collection from network traffic.
- **Preprocessing and Initial Model Training:** Using historical traffic data to train the initial models.

Beta Testing: We will go through a beta testing phase with the ARMA system after the initial development. This will be achieved by deploying it with a group of early adopter companies. With this phase, we aim to:

- **Gather Real-World Feedback:** Changing the model depending on how it performs in real-world settings.
- **Iterative Improvements:** Improving the machine learning algorithms and system interfaces based on user feedback.

Full-Scale Deployment: We will roll ARMA out to a broader market focusing on the following:

- **Scaling Up:** Adapting the system's capacity to handle higher amounts of data and customer numbers.
- **Customer Onboarding:** Simplifying the process of implementing ARMA into small businesses.

Model Training and Maintenance

Continuous Learning Approach: ARMA models will continuously learn from new data thereby adapting themselves to evolving network behaviours and emerging threats. This includes:

- **Ongoing Data Collection:** Automating the collection of new traffic data to optimise our models.
- **Regular Model Updates:** We will schedule updates for the machine learning models to allow the incorporation of new data and insights.

Maintenance Strategy: With regular maintenance, we will aim to maintain the system's reliability. Two main methods will be used:

- **Model Health Checks:** Regular checks of the model's health to help maintain its performance.
- **Feedback Loops:** User feedback will be added as a part of its model training to improve usability and efficiency

Risk Management

Technical and Security Risks: Various factors can create technical or security risks. Some of the main factors we identified are data breaches, model overfitting and performance degradation. We decided to deal with these through the following methods:

- **Robust Security Protocols:** The use of reliable and robust security protocols.
- **Regular Security Audits:** Identifying vulnerabilities through frequent audits of the system to allow for rectification
- **Performance Monitoring:** Continuous monitoring of the system performance to find possible degradation that could harm the system

Business Model and Outcomes

Our Business model follows two main parts, which can be stretched into three:

Subscription Service: Our foundational offering is a subscription-based service. Businesses subscribed to us will receive security over their network. We will oversee all communication and traffic, neutralising potential threats using our machine-learning model. The service is tiered depending on different security needs of small companies and startups:

- **Basic Tier:** Includes essential monitoring and protection.
- **Advanced Tier:** Adds more sophisticated monitoring tools, regular security audits, and incident response services.
- **Premium Tier:** Provides all features of the Advanced tier plus dedicated support and advanced threat detection and predictive analytics.

Custom Solutions & Consulting: Beyond our subscriptions, we also offer custom solutions and consulting services, thanks to our mitigation algorithm, tailored to the needs of our clients.

ML Model Sales: For companies that prefer to handle their own security, we provide the option to purchase our machine learning model. This allows businesses to use our technology but still maintain control over their internal security operations without having to pay us monthly.

Expected Outcomes:

- **Stable Revenue Growth:** As a result of the multiple offers, from a subscription service to the sale of our model, we're predicting a steady revenue stream.
- **Client Retention and Satisfaction:** Our range of services, and efficient security, coupled with the option for customisation, leads to high client satisfaction and retention.
- **Market Expansion:** The flexibility of our business model allows us to continually expand into new markets. Today a company in any sector needs cybersecurity.

Market Strategy and Expansion

Marketing and Promotion:

- **Digital marketing:** Our strategy involves searching for an adequate social media platform where the small business owners would be the most active and investing in pay-per-click advertising within it, furthermore we'd implement targeted digital marketing campaigns using SEO.
- **Partnership:** Partnering with industry associations, tech communities and other business service providers would enable us to expand our audience through co-marketing our services while also helping us gain credibility
- **Local Business Outreach:** We will engage directly with local business communities through workshops and seminars that help small business owners understand their cybersecurity risks and the solutions we offer.

Expansion opportunities:

- **Industry-specific Cybersecurity Packages:** We can expand our services to reach and specialise in specific industries, for example, have a model trained purely on healthcare or finance. We can also include mobile security solutions and IoT device protection.
- **Deep Learning & AI:** We are committed to staying up to level with emerging technologies, including deep learning, artificial intelligence as well as predictive analytics. These would help enhance our threat detection capabilities and provide better, more precise solutions.
- **Blockchain Implementation:** Blockchains are the epitome of digital security, so the implementation of that technology can provide a higher level of integrity.

Conclusion

In conclusion, ARMA is a robust cybersecurity solution designed to support businesses, especially small firms and startups, against the growing cyber threats in the digitised world. Leveraging top-of-the-line machine learning technology, such as decision trees and neural networks, ARMA offers businesses a proactive platform for monitoring, detection, and response against cyber threats in real-time and strategic mitigation measures. Our business model approach involves offering subscription services, bespoke solutions, and selling the ML model we develop, ensuring that we attain a wide market field and address the varied security needs of all customers.

Through calculated advertising and marketing as well as durable collaborations ARMA is made to end up being a leader in network safety and security services, constantly broadening right into brand-new markets as well as improving its offerings with the most recent technical improvements.