|  |  |
| --- | --- |
| Module: | LD7028 – Research and Project Management |
| Module Tutor: | Hamid Jahankhani / Shelagh Keogh |
| Assignment Title: | MSc Research Proposal |
| Student Name: |  |
| Student University Identifier: |  |
| The course of Study: | MSc Cyber Security |
| Supervisor: |  |
| Second Marker: |  |
| Review Date: |  |
| Word Count (Only Content): | 1696 |
| Total Word Count: | 1867 |

Table of Contents

[Table of Figures 1](#_Toc99966777)

[Title 2](#_Toc99966778)

[Aim 2](#_Toc99966779)

[Problem Statement 2](#_Toc99966780)

[Background 2](#_Toc99966781)

[Scope 3](#_Toc99966782)

[SMART Objectives 3](#_Toc99966783)

[Methodology 3](#_Toc99966784)

[The 2FA Playground 3](#_Toc99966785)

[Machine Learning Model 4](#_Toc99966786)

[Project Plan 4](#_Toc99966787)

[Known Risks 5](#_Toc99966788)

[Source and Use of Knowledge 5](#_Toc99966789)

[Ethical, Legal, Social, Security and Professional Concerns 6](#_Toc99966790)

[References 6](#_Toc99966791)

# Table of Figures

[Figure 1 - Project Plan 4](#_Toc70539800)

[Figure 2 - Gantt Chart 5](#_Toc70539801)

[Figure 3 - Known Risks and Risk Management 5](#_Toc70539802)

# Title

“Identifying Anomalies in Mutli-Factor Authentication with AI”.

# Aim

To develop a third approval layer including arranged Machine Learning models prepared to perceive peculiarities after the main affirmation stage.

# Problem Statement

With the advent of technology, the security and safety of our electronic lives are not far-fetched. The only problem is that the development in security technology is a double-edged sword due to the freedom of access people have to technology. Thus, hackers also get their hands on this technology easily and find ways to get past it. Two-factor authentication (2FA) is one of the most innovative security measures to prevent hackers and malice from approaching our social accounts. 2FA requires a third-party app or device to verify the account details during login.

Cell phones assume significant parts in numerous individuals' day-by-day life. Individuals usually use cell phone applications to take photographs, send messages, book rides, or shop on the web. It is not strange for those applications to ask for private data (like names, sex, or Visa data) from their clients to improve the nature of their administration. The delicate idea of those private data requires application engineers appropriately tie-down admittance to their administration. A mainstream approach to getting such access is by asking for passwords from clients during login measures (Irvan et al., 2021).

In any case, passwords and other information-based validation strategies like PIN (individual ID number) codes convey extraordinary danger as clients will in general utilize similar passwords across different administrations. Accordingly, numerous administrations right now require extra belonging based verification techniques before allowing access. An ordinary method of this execution is by sending an interesting code through SMS (short message administration) to clients' telephone numbers. This additional progression is 2-factor validation (2FA) or multifaceted confirmation (MFA). Tragically, ownership-based validation techniques carry possible bothers to clients since e they may need to convey extra gadgets, which can be effectively lost. Numerous clients additionally utilize the same cell phone to enter passwords and get 2FA codes. In this manner, if their cell phone goes missing, assailants can sidestep 2FA checks (Irvan et al., 2021).

To prevent this, a novel solution of creating a third layer in the 2FA that will detect anomalous attempts using machine learning has taken its bearings for implementation. This also claims the title 3FA or the 3-Factor Authentication for ease of discussion in this paper.

# Background

In April 2019, Kaspersky researchers revealed an enormous scope of SIM trade misrepresentation activities focusing on clients in both the Portuguese-talking countries of Brazil and Mozambique had the option to utilize social designing, pay off, and straightforward phishing assaults eventually taking cash from casualties. Danger entertainers did these assaults by assuming responsibility for a casualty's telephone number by capturing accounts and catching two figure confirmation strategies in which the subsequent verification factor is an SMS message or a call put to the portable number.

Two-factor validation, the additional security step that requires individuals to enter a code shipped off their telephone or email, has generally attempted to protect usernames and passwords from phishing assaults. In any case, security specialists have shown a mechanized phishing assault that can slice through that additional layer of safety—likewise called 2FA—conceivably fooling clueless clients into sharing their private qualifications.

Cases like these are popping up all over the country. Various hackers have compromised the integrity of the 2FA security. Researchers are trying to keep up with their security measures but the requisite of something intelligent and robust still exists.

# Scope

This project covers the following aspect of the domain:

* Development of the 2FA playground for testing the novel proposed methodology.
* Development of the Machine Learning algorithm.

# SMART Objectives

The problem this project needs to tackle is the existence of hackers that break the 2FA authentication using phishing attacks or any other type of attacks like sim-swapping. This problem belongs in the cyber-security domain since its concerns lie with 2FA. The solution is to introduce the third tier in the 2FA architecture that includes Machine Learning algorithms that will help detect any anomalistic behaviour indicative of a hacker. The following objectives will help achieve this.

* Literature Review
  + Papers older than the year 2014 are not eligible.
  + Finalizing a minimum of three papers for the literature review before 30 January.
  + The Literature Review must complete before 3 March.
* Dataset
  + Finalizing the dataset before 9 March.
  + The Data cleaning must be complete before 15 March.
  + The Dara pre-processing must be complete before 8 April.
  + The Statistical Analysis of the dataset must contain the Descriptive Analysis of all five Central tendency measures.
* Machine Learning
  + Comparing a minimum of three different Machine Learning algorithms together for the validation of performance and selection.
  + The performance metrics should provide an accurate result of 85% or greater.
* The software development process must be complete before 22 May.

# Methodology

The project has two basic modules. The modules are as follows:

## The 2FA Playground

The 2FA playground will be, as the name suggests, a playground for testing and experimenting in this project. The playground will be enabled with two main features i.e., the 2FA auth. and request tracking.

The programming language used for this module will be JavaScript and the tools employed from JS will be the **Node.js** and **express** server libraries to create the backend server that the app will run from and the **speakeasy** library to implement the 2FA. The **speakeasy** library functions combined with the **Google Authenticator browser extension** to generate tokens that verify the user Login process. The main functions as mentioned previously for this module will be to provide an API that allows 2FA user registration, user verification and user validation. In addition, this module will be responsible for collecting client request data from the requests on this module. User data like the **IPv4 address, IPv6 address, email and password** are targets for the collection so that they can train the Machine Learning Model to identify anomalies.

This module will use the free Heroku server for hosting. This will make the module online and accessible by different devices having different IP addresses. This will allow us to collect data from different devices under different scenarios also providing a much-needed variation in the data. The information of the user freshly registered will be stored as JSON in an onboard JavaScript object whose access will be to all the project files to avoid any access issues down the road.

## Machine Learning Model

The Machine Learning Implementation requires further research to finalize it properly. The most crucial part of the implementation of Machine Learning part is the dataset. The dataset for training Algorithms on anomaly detection in 2FA is not available as of yet. Further research may uncover an existing dataset or the data collection will be the choice.

In case the dataset requires creating it from scratch, dummy data from the playground developed will provide the necessary input and anomalies doped and induced as a proof of concept for training the algorithm and providing results. Depending upon the features in the dataset (created or found online), the Machine Learning approach can vary from simple binary classification to Deep Learning or even unsupervised learning using clustering. Further research will help finalize these facts.

# Project Plan

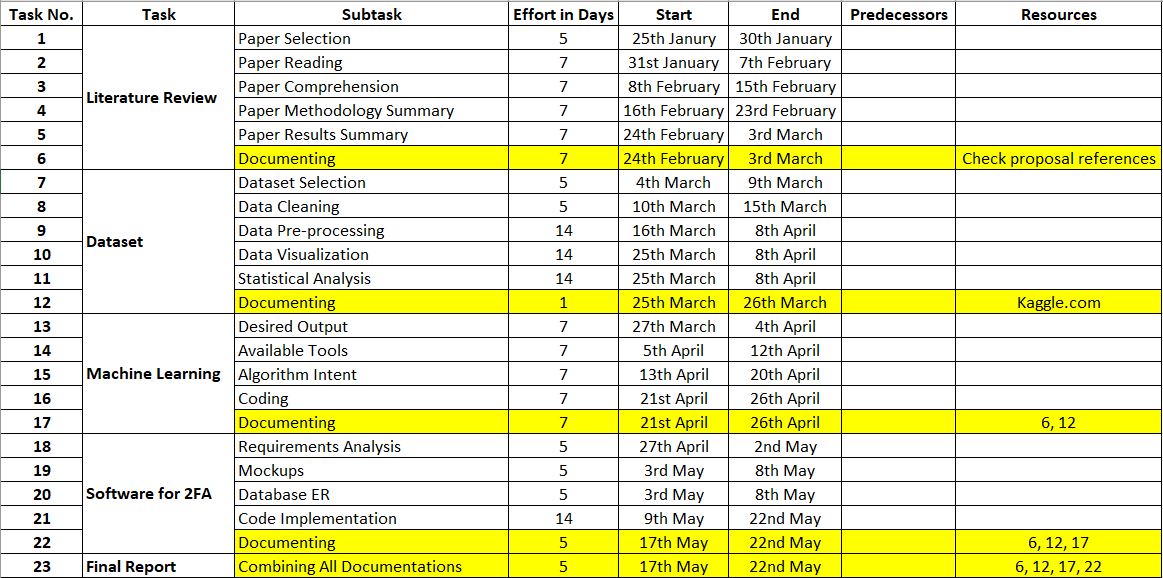


Figure 1 - Project Plan

Figure 2 - Gantt chart

# Known Risks

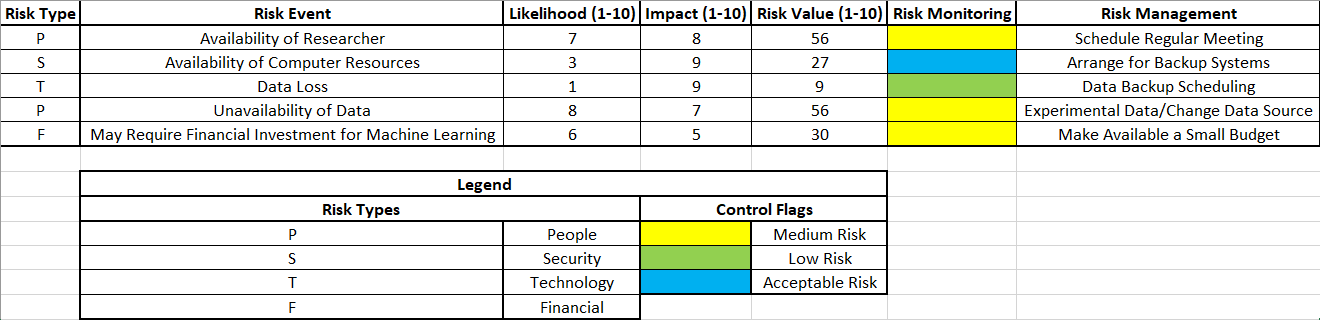


Figure 3 - Known Risks and Risk Management

# Source and Use of Knowledge

Since this is a novel approach, to the extent of the literature review done for this project, there is no availability for research references on the same topic. This makes it difficult to find material to study and prepare. However, the approach for this research project has been adapted to find a way through it.

Three modules in this project require extensive research reviews:

Data: The availability of any dataset for training machine-learning algorithms to identify anomalies in the 2FA authentication system is uncertain. Further research may uncover this to be false and provide the dataset requisite.

Machine Learning: Although numerous articles talk about the implementation of Anomaly Detection using Machine Learning, Anomaly detection concerning this application with such an approach is a rarity. The reason behind this is the unavailability of the dataset. However, the available research makes itself useful to prepare an algorithm/ensemble that can perform well when needed.

2FA: Two-factor authentication implementations are available in huge amounts on the internet. However, a proper, customizable 2FA implementation is a little difficult to find since 2FA is a secure protocol and not meant to fiddle with. However, implementation of a modular 2FA for the research in which the ML attribute of this project can be a part is in the works.

These three modules of this project require extensive research. The source of information is the internet. Particularly websites like scholar.google.com and arxiv.org. The only restriction is that any implementation/research material that is considered for this research must not be older than 2014.

# Ethical, Legal, Social, Security and Professional Concerns

Indisputably perceiving the degree of the data, what it does for all individuals before the collection of such data is basic knowledge. That the use of data ought to be for the inspirations driving this expected devastation after a reasonable period has passed. To this end, all individuals ought to be occupants inside the United Kingdom and British nationals. Under both the British Data Protection Act and the European General Data Protection Regulation, the Information Commissioner's Office (ICO) claim the legitimate Supervisory Authority spot.

The focus of this assessment is the examination of an individual mailing affinity over a discrete period; care encapsulated guarantee for both that resource and the organization used exists. It very well may be legitimate to hold both the data and any system utilized inside a guaranteed network, for instance, that given by the goal association. It is sensible and appropriate to use encryption still and in transit for the rough data got from individuals. Toward the completion of the endeavour, obliteration ensues of all fundamental or arranged data. Masking the true origins of data conveyed in this research is of utmost importance along with anonymization and summarization as a pre-requisite. As the assessment might claim employment from the target relationship, there is a potential hopeless circumstance between the necessities of the organization, the school and keeping the investigation self-governing and objective.

To this end, the assessment will adhere to the ethical limit legitimate to an MSc project. Comparative substances also have clashing cases on Intellectual Property that may arise out of the endeavour that ought to dissolve easily. The errand recommendation is to pass on a strategy or computation for conveying adjusted phishing endeavours and does not struggle with the goal affiliation's standard exercises or business zones. Assessing the ethics of building sensible or regardless, imitating certifiable world phishing amusements is mandatory. Brand name and protected innovation law ought not to have infringed which curbs the legitimacy of impersonated attacks.

Efforts to make non-infringing diversions will take place. Thought about how to help the mission to thwart it ending up being, or individuals feeling like it has become a denouncing, shaming or fault managing measure. Anonymization midway aids anyway a distinction in culture will not occur without such great requests being eased. Perhaps gamification of phishing declaring may be a positive turn of events will be a possible suggestion. A couple of affiliations have "Star Awards" and this may be a phase towards making interest a positive association.

# References

* Irvan, M., Thao, T., Kobayashi, R., Nakata, T. and Yamaguchi, R., 2021. Learning from Smartphone Location Data as Anomaly Detection for Behavioral Authentication through Deep Neuroevolution. Seventh International Conference on Information Systems Security and Privacy.
* Ali, G., Ally Dida, M. and Elikana Sam, A., 2020. Two-Factor Authentication Scheme for Mobile Money: A Review of Threat Models and Countermeasures. Future Internet, 12(10), p.160.