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

Table of Contents

[Title 2](#_Toc70539788)

[Aim 2](#_Toc70539789)

[Problem Statement 2](#_Toc70539790)

[Background 2](#_Toc70539791)

[Scope 3](#_Toc70539792)

[SMART Objectives 3](#_Toc70539793)

[Methodology 3](#_Toc70539794)

[Project Plan 4](#_Toc70539795)

[Known Risks 5](#_Toc70539796)

[Source and Use of Knowledge 5](#_Toc70539797)

[Ethical, Legal, Social, Security and Professional Concerns 6](#_Toc70539798)

[References 7](#_Toc70539799)

# 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 2-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 aren’t 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 for 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 approach 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 isn't strange for those applications to ask 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 to appropriately tie-down admittance to their administration. A mainstream approach to get such access is by asking for passwords from clients during login measure (Irvan et al., 2021).

In any case, passwords and other information-based validation strategies like a 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 technique 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 known as 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 is taken, 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 will be implemented. This will be referred to as the 3FA or the 3-Factor Authentication.

# Background

In April 2019, Kaspersky researchers revealed enormous scope 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 to eventually take 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 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 3rd 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 will not be selected.
  + A minimum of 3 papers must be selected for the literature review before 30th January.
  + The Literature Review must complete before 3rd March.
* Dataset
  + The dataset must be finalized before 9th March.
  + The Data cleaning must be complete before 15th March.
  + The Dara pre-processing must be complete before 8th April.
  + The Statistical Analysis of the dataset must contain the Descriptive Analysis of all 5 Central tendency measures.
* Machine Learning
  + A minimum of 3 different Machine Learning algorithms must be compared 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 22nd May.

# Methodology

The complete implementation of the project will be done using Python. The playground will be designed using basic HTML, Bootstrap. The 2FA modules will be implemented using the Twilio and Google Authenticator libraries available for Python. The implementation for 2FA will use the Twilio API and Will be developed in Python. The application for the 2FA will be a very basic website that only uses a simple username and password-based authentication. Once the code for this is completed, the 2FA will be implemented. For the basic website, Flask, a framework for quick and easy website development will be used. The implementation of this website is essential to the research project since the layout and the working of the 2FA in this website will affect how the Machine Learning layer will be implemented.

The Machine Learning Implementation requires further research to finalize it properly. The most crucial part for the implementation of the Machine Learning part is the dataset. The dataset for training Algorithms on anomaly detection in 2FA isn’t available as of yet. Further research may uncover an existing dataset or the dataset will have to be created from scratch.

In case the dataset has to be created from scratch, dummy data from the playground developed can be recorded and anomalies can be created 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

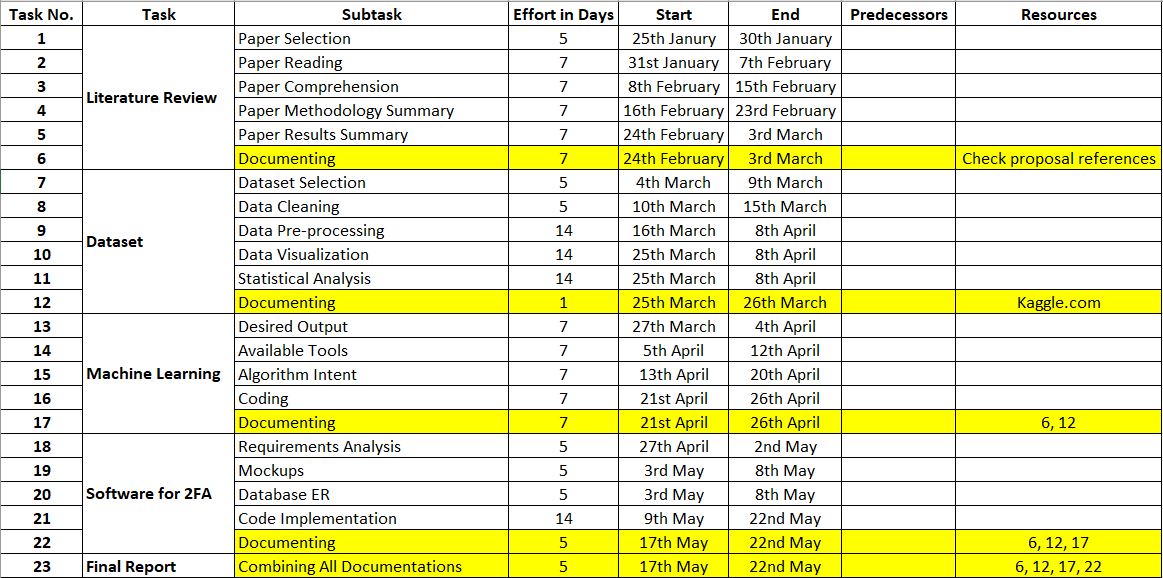


Figure - Project Plan

Figure - Gantt chart

# Known Risks

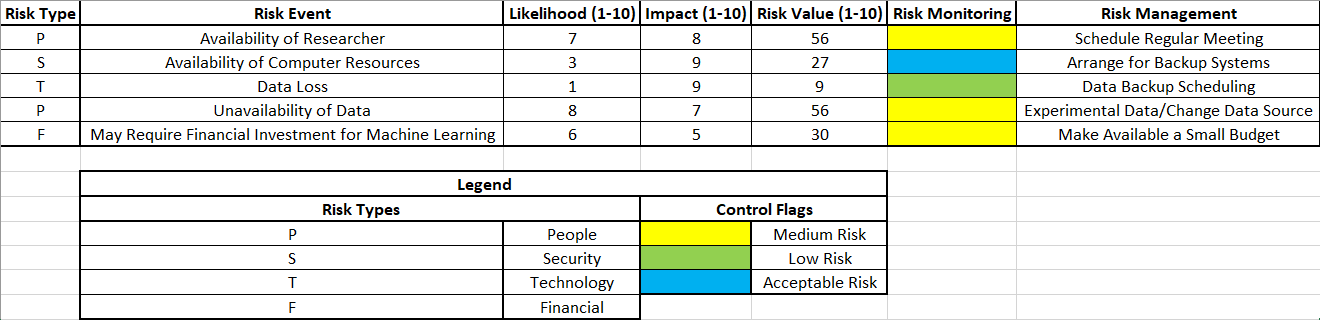


Figure - Known Risks and Risk Management

# 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 with 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.

There are three modules in this project that 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 is rarely done with such an approach. The reason behind this is the unavailability of the dataset. But the available research can be used to prepare an algorithm/ensemble that can perform well when needed.

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

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

It is basic to indisputably perceive the degree of the data and what it is to be used for to all individuals before the collection of such data. That data ought to simply be used for the inspirations driving this expected and devastated after a reasonable period has passed. To this end, all individuals ought to be occupant 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) will be seen as the legitimate Supervisory Authority.

As the assessment is focused on the examination of an individual mailing affinity over discrete period care ought to be taken to guarantee both that resource and the organization used. 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, all fundamental or arranged data will be obliterated. Any data conveyed of this environment will be anonymized and summarized as requirements be. As the assessment may be used by 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 be pleasantly settled. The errand recommendation is to pass on a strategy or computation for conveying adjusted phishing endeavours and doesn't struggle with the goal affiliation's standard exercises or business zones. The ethics of building sensible or regardless, imitating certifiable world phishing amusements ought to be assessed. Brand name and protected innovation law ought not to have infringed which curbs the legitimacy of impersonated attacks. Entertainment providers give articulations determining that ensured advancement rights ought to be respected.

Tries ought to be made, henceforth, to make non-infringing diversions. Thought ought to be made 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. It is suggested that perhaps gamification of phishing declaring may be a positive turn of events. 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. 7th 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.