# 7CS997 Independent Studies Project Plan

## Student/supervisor details

**Submission deadline for signed and completed project plan: 19/05/2023 12:00 Midday**

Project Title: Detection of Spam Emails Using Machine Learning/Deep Learning

Name: **Ikenna Obumnaeme Oranyiridike**

Student Id: **100621591**

MSc: Big Data Analytics

Supervisor: Dr Asad Abdi

## Brief description

*Spam email detection employing machine learning or deep learning approaches entails utilising algorithms for recognising unsolicited or unwelcome emails sent to a substantial number of recipients. These messages frequently include fraudulent and harmful material that can be detrimental to both persons and organisations. Machine learning algorithms are developed on massive datasets of spam as well as non-spam emails to detect spam emails. These representations learn to recognise patterns and attributes that separate real emails from spam. While training the model, features that include the sender's email handle, the topic line of the email, or the body of the email itself are all considered in account.*

*Deep learning models make this technique a step further by employing neural networks to digest or analyse the email's content in more depth. These models are capable of detecting complicated patterns and language clues that suggest an email is spam. Once trained, the model is able to be used to categorise fresh incoming emails as spam or not. This is accomplished by analysing the email's features while contrasting them to the trends established throughout training. If the model concludes that the email is spam, then may be sorted into another folder or removed automatically.*

*In the end, employing machine learning / deep learning models for identifying spam emails provides an effective technique to safeguard individuals and organisations from the detrimental impacts of spam emails. Everyone should expect even more accuracy in recognising as well as filtering out emails that are spam in the future as these models get more sophisticated.*

## Project aims and objective

**Aim:** *The primary Aim of this project is to detect the spam emails by utilizing machine learning/ deep learning models*

**Objectives:**

* *To perform a complete analysis of present machine learning / deep learning algorithms and methodologies for email spam detection.*
* *To investigate the impact of different hyperparameters on the effectiveness of deep learning models, which include training rate, the size of the batch, the number of layers, as well as activation functions.*
* *To determine the suggested approach's drawbacks and constraints, which include the generalizability of the models trained to fresh data, possible bias as well as fairness problems and the compromise between precision and computing cost.*
* *To assess the efficacy of supervised and unsupervised learning algorithms in identifying spam emails, particularly to see if using labelled data increases model accuracy.*
* *To examine the viability of leveraging pre-trained models on huge datasets of natural language processing duties to increase the accuracy of spam email detection models utilising transfer learning approaches.*

## Plan of work

***4.1 Literature Review***

*In today's world, e-mail has become one of the most widely utilised ways of communication, making it an ideal target for spam or phishing assaults. Because of the advanced strategies employed by attackers to avoid detection, identifying spam emails is a difficult undertaking. Algorithms based on deep learning and machine learning have demonstrated promising results in accurately identifying spam emails. Multiple studies on email spam identification employing machine learning / deep learning models are being undertakenDada et al. (2019) introduced a deep learning-based algorithm that used a multi-layer perceptron (MLP) to categorise emails as spam or real. The suggested model attained an accuracy of 98.25%, according to the findings.*

*Kumarand Sonowal, (2020), suggested a machine learning-based solution which utilised many criteria, which includes email header, content, including attachments, to identify emails as spam as well as legitimate. The analysis found that the proposed method was 97.35% accurate. FurthermoreAhmed et al. (2022) suggested a hybrid model that combines machine learning as well as rule-based techniques for identifying spam emails. The model that is suggested has a 98.15% accuracy. In general, the findings indicate that machine learning as well as deep learning models are quite successful at detecting spam emails. On the other hand, further study must be conducted to increase the models' effectiveness and adaptability for real-time applications.*

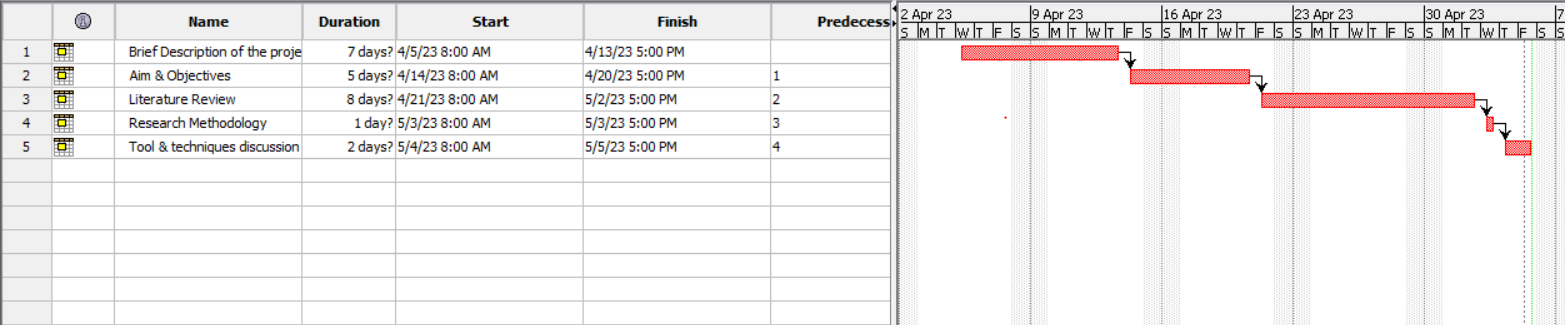
***4.2 Methodology***

*In order to conduct the task, the waterfall methodology will be considered as Utilising the waterfall methodology for creating a spam email identification system may offer an organised strategy that can be helpful in regards to controlling the design and development process, maintaining that the needs are correctly determined, along with ensuring that the entire system has undergone extensive testing before deployment.*

***4.3 Tools & techniques***

*For conducting the spam email detection different types of python libraries will be executed that helps the load or manage the spam dataset and make predictions based on the dataset. Further, CNN or other deep learning model will be utilized to detect the message is spam or not. Jupyter notebook tool will be consider for this task.*

## Gantt Chart

**

**Figure 1: Time Plan**

(Source: Self-created in ProjectLibre)

## Meeting plan

*In the table below, the project week number and dates correspond to the week numbers that were presented in the introductory lecture that are designated meeting weeks. You should fill in the date, time, and venue for the meeting schedule that you have agreed with your supervisor as part of planning your project.*

|  |  |  |  |
| --- | --- | --- | --- |
| Project week | Meeting date | Meeting time | Venue |
| 1: 05/06/2023 |  |  |  |
| 3: 19/06/2023 |  |  |  |
| 5: 03/07/2023 |  |  |  |
| 7: 17/07/2023 |  |  |  |
| 9: 31/07/2023 |  |  |  |
| 11: 14/08/2023 |  |  |  |

## Ethical approval

I confirm that I have discussed and agreed the project plan with my supervisor, and have completed the ethical approval process. The PDF confirming ethical approval has been completed and uploaded to the relevant submission point on course resources.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Name Supervisor Name

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Signature Supervisor Signature

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date Date