**CHAPTER I**

**INTRODUCTION AND ITS BACKGROUND**

Spam messages refer to irrelevant messages sent through various communication channels, like email, text messages, and social media platforms. These messages contain promotional ads, phishing attempts, fraudulent texts, and other forms of unsolicited content. Not only do they waste our time and attention, but they can also pose serious risks, leading to identity theft and financial losses.

Communication is a vital part of our lives. We communicate every time in different ways. One of which is by sending a text message or SMS. Most of people nowadays owns a mobile phone, but due to some scammers, receiving of spam SMS was an issue identified. These SMS contains fake links that can trick users to input classified information like passwords, bank accounts and more. A lot of people have fallen to the scam and lose large amount of money just by clicking a link.

SMS became one of the best forms to communicate because of many reasons, the first being that SMS is efficient and reliable. SMS can be delivered instantly, even if you are far from your target receiver, depending on the signal strength of your mobile phone. SMS is also cost-effective, in the Philippines, there are promos that you can register to have unlimited text even to all networks for such a low price.

Despite those advantages of SMS, hackers and scammers use it to trick people and get their information and money. This type of trick is called Smishing, Smishing targets victims through mobile text messaging or SMS. Individuals usually receive unsolicited messages that include links that redirect to fraudulent websites which steal personal data, introduce mobile malware, and even facilitate the commission of fraud (Philippines: NPC Warns the Public Against Smishing, 2021).

According to Dela Cruz, in 2022 Globe Telecom Inc. (Globe) have blocked over 138 million spam and scam text messages from January to June 15, preventing Filipinos from engaging with malicious entities through SMS.

Spam messages may sometimes contain the users’ name which is a bit bothering because it is sensitive information. Leandro Angelo Aguirre, deputy privacy commissioner of the National Privacy Commission, said Tuesday that our names may have been manually or automatically scraped by scammers from certain apps.

Aguirre told ABS-CBN’s TeleRadyo that their complaints and investigation division has seen a pattern in these targeted smishing attempts as they bear names of users of “popular payment applications, mobile wallet or messaging applications.” (Gregorio, 2022).

In the present day, there are various spam filtering technologies that exist, however, most of them are designed for emails only. There are only a few ways to filter out and determine spam SMS.

To address this problem, the researchers aim to design and develop a mobile app and utilize Machine Learning, Deep Learning and Natural Language Processing in the identification of a spam SMS. The proposed app will filter out incoming/old SMS on a users’ phone, then block the number if it is sending spams frequently. Protecting the user from the possibility of clicking a fake link that will lead to identity theft and loss of money.

**Project Context**

According to Rosales (2022) the recent deluge of text scams has claimed millions of dollars from network subscribers in the Philippines who ended up clicking on malicious links and giving up sensitive information.

Around 1.1 billion scam and spam messages were intercepted by Globe in the first three months of 2023 in its ongoing campaign against online fraud, coupled with the implementation of the SIM Registration Act.

The number of blacklisted SIMs also increased with the help of Globe's Stop Spam portal. It reached 22,455 from January to March this year against 1,812 recorded in the same period in 2022. Globe also deactivated 647 SIMs, of which 610 were involved in sending scam or fraudulent messages while the remaining 37 were used to send spam messages (Amojelar, 2023).

To lessen the damage and protect individuals from smishing, the researchers will propose a mobile spam SMS detector app. The said application will detect and archive old and incoming messages on a users’ mobile phone to protect them from clicking the link on those messages and prevent being scammed. The application will also automatically block the sender of spam messages so the scammer can’t send spam messages frequently by using a single mobile number.

The application will use Natural Language Processing for the text preprocessing of the SMS, then use Machine Learning and Deep learning techniques to analyze and classify SMS based on their contents. If the application classifies an SMS as spam, it will automatically add the SMS to the archive so the user won’t have to see it and accidentally click the link on the SMS.

In a particular project setting, this research focuses on analyzing the effectiveness of an AI spam detector on filtering out spam SMS. The following research problems must be handled by researchers: (1) How to design and develop a mobile app that will automatically detect and block spam messages on a users’ phone. (2) How will Machine Learning and Deep Learning models will perform in identifying a spam SMS? (3) How to examine the performance of the models developed integrated into the system through actual testing?

**Research Problem**

Most people nowadays own a mobile phone, but due to some scammers, receiving malicious spam SMS was an issue identified. This study aims to design and develop a mobile app that will handle the automatic filtering of spam SMS in a users’ phone using Natural Language Processing for the text preprocessing of the SMS and Machine Learning or Deep Learning Algorithms for the analysis and classification of the SMS. With the help of modern technology this study will address the following problems:

1. How to apply Natural Language Processing and Deep Learning algorithm to a mobile app that will filter out spam SMS?

2. How can we determine the best performing deep learning algorithm that can be used in the classification of spam SMS?

3. How can the researchers collect messages for the dataset that will be used in this study?

4. How effective will the combination of Natural Language Processing and a Deep Learning Model be in identifying spam SMS messages?

5. How can we test the performance of the mobile app in a mobile device as it works on filtering spams?

**Research Objectives**

The main objective of this study is to design and develop a mobile app and utilize Natural Language Processing for the pre-processing of the message and Deep Learning for the identification of a spam SMS to alert and protect mobile phone users from identity theft and loss of money.

Specifically, it aims:

1. To develop a mobile application that utilizes Natural Language Processing and Deep Learning algorithms that detect and filter out spam SMS messages.
2. To determine the best performing Deep Learning algorithm that can be used for the identification of spam SMS
3. To collect messages (spam and ham) for the dataset that will be used in this study.
4. To determine the effectiveness of Natural Language Processing and Deep Learning Model in identifying a spam SMS.
5. To test the performance of the mobile app in a mobile device while it filters out spams.

**Theoretical Framework**

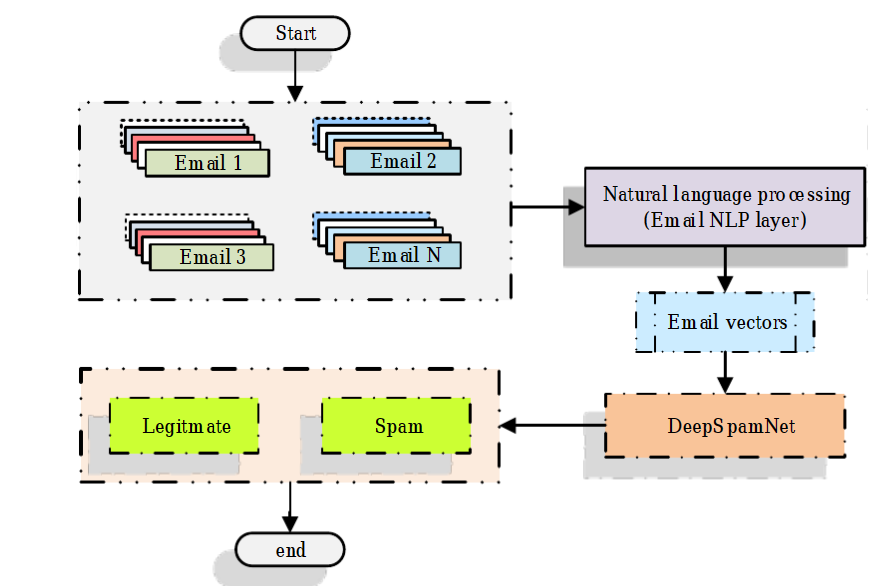


Figure 1. DeepSpamNet Email Classification using NLP and Deep Learning.

In the study of Srinivasan et al. (2020), they proposed a strategy that is called asDeepSpamNet, it includes three phases: preprocessing, feature extraction, and classification. During the preprocessing stage, the emails are converted into a feature vector through text representation techniques. The strategy includes different methodologies to determine the optimal text representation approach, considering the unique attributes of each technique, including TDM, TF-IDF, TDM with SVD, TDM with NMF, TF-IDF with SVD, TF-IDF with NMF, Keras embedding, FastText, NBOW, and word embedding.

In contrast to other feature engineering methods, the proposed strategy provides explicit feature engineering. Instead, it turns towards autonomous feature acquisition. However, the extraction of keywords is accomplished through email processing, categorizing emails based on their content. Other information, including source IP and email addresses, are collected from the emails. This involves the retrieval of both email addresses and the corresponding source IP addresses. After processing of each email set, statistical metrics are generated. This compilation of information goes in a dedicated database, adaptable to updates with the introduction of new datasets. Notably, consistent updates to these databases yield an elevated rate of spam detection coupled with a proportional reduction in false positives.

Assigning an email correctly as spam corresponds to a true positive, while misclassifying a legitimate email as spam corresponds to a false positive. The potential risk of losing communications and misunderstanding due to wrong classification is a matter of concern. Addressing this, the model requires training on contemporary datasets to improve its effectiveness.

**Conceptual Framework**

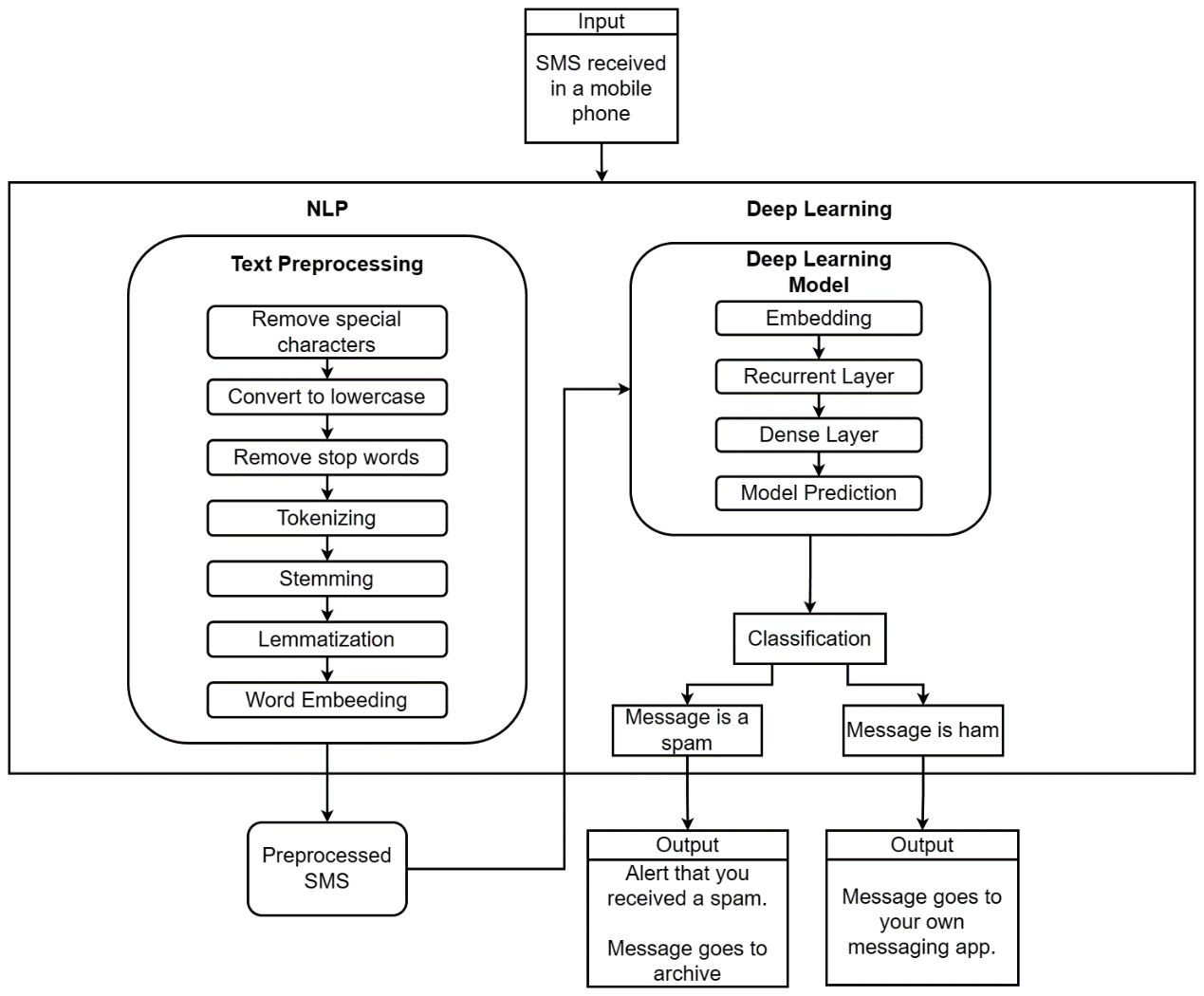
****

Figure 2. Conceptual Framework of SPAMTZY: Mobile AI Spam Detector using Natural Language Processing and Deep Learning

In the figure provided above, the conceptual framework of the spam detector application is illustrated, showing the processes happening inside the app. To elaborate, the initial inputs consist of SMS messages received on a mobile phone. These SMS messages undergo a series of text preprocessing steps utilizing NLP techniques, resulting in the generation of preprocessed SMS content.

The preprocessed text is embedded, meaning that the words are being represented in multi-dimensional vectors. This technique is used to capture semantic relationships between words. Words that are contextually or semantically similar in a given dataset tend to have similar vector representations. This property allows models to understand and generalize relationships between words, making them capable of recognizing synonyms, antonyms, and more.

Subsequently, the preprocessed SMS content is inputted into a meticulously trained Deep Learning Model. This model then takes the task of predicting whether the SMS content is categorized as spam or ham. If the model identifies the SMS as spam, an automated action takes place: the spam SMS is directed to an archived list within the app, and simultaneously, a notification is displayed to the user, informing them of that they have received a spam message. Conversely, if the SMS content is classified as ham (non-spam), an alternative course of action happens: the message is directly routed to the messaging application on the user's mobile phone, immediately notifying them of the arrival of a new message.

**Scope and Limitations of the Study**

The scope of this study involves;

* The development of a mobile app that will filter out and block incoming/old spam SMS on a users’ phone.
* This study will use Natural Language Processing and Deep Learning algorithm in order to identify if an SMS is a spam or ham.
* This study will also evaluate the effectiveness of the developed mobile app through user acceptance testing and gathering feedback for further improvement.

The limitations of this study are;

* the proposed mobile app is only designed for Android Operating Systems only; it may not be compatible with other mobile Operating Systems.
* The mobile app will also only filter out and block SMS from the phone’s messaging app and not messages from other messaging apps like Messenger, Viber, or Telegram.
* The app also will only predict base on the content of the message received.

**Significance of the Research**

This study will be beneficial to phone users that receives numerous amounts of spam messages. The proposed app will filter out spam SMS, making them safer from being scammed by giving their information without them knowing it. The significance of this study lies in the potential of the application to identify and filter out spam messages from a users’ phone. Specifically, the proposed application will be a great help to the following individuals or groups:

* **General Mobile Phone Users.** They are the ones that spend most of their time communication through their phone, thus making them prone to Smishing. The app can reduce the amount of spam they receive, which can save them a lot of time and money. The app can also help to protect them from scams and phishing attacks.
* **Job Seekers.** The app can protect them from spam messages that offers fake jobs. This can save a lot of time and money.
* **Business Professionals.** Professionals can guarantee that important work-related SMS are not buried by spams, improving their productivity and preventing missed opportunities.
* **Entrepreneurs and Small Business Owners.** Individuals running businesses can ensure efficient communication by preventing essential client or customer messages from being buried in spam folders.