Hi everyone, my name is Nairui or David and

In today's connected world, where the internet is a big part of our lives, there have been a new type of danger that is growing significantly. And that is scams in all types, either over the phone, emails or websites. But today I will mainly be focusing on Phishing URLs. Phishing spelt with Ph, is like fishing spelt with a f, but rather, it’s for your personal information. It's when someone tries to trick you into giving away your sensitive details, like passwords, credit card numbers, or personal information, by pretending to be a trustworthy source. This often happens through emails, text messages, or websites that look real but are actually fake and designed to steal your information. The goal is to deceive you into revealing confidential data or clicking on malicious links.

These online traps are a serious threat to Australians, causing significant money, personal, and online safety problems.

Recent numbers show that these attacks have gone up a lot. In 2022, there was a 44% increase in reported incidents in Australia compared to the year before. The effects of these attacks have been wide-ranging, causing different kinds of problems across our society.

Firstly, Australians have lost a lot of money. In 2022, the total loss was a huge $81 million, with an average of $1,800 lost per incident. Identity theft cases have also gone up, affecting thousands of people. Complaints about identity theft to the police went up by 37% in the last year.

Data security has also been badly hit, with a 52% increase in reported incidents of data breaches caused by phishing attacks. These breaches exposed private information, like health records and money details, making people vulnerable.

Cybercriminals, by getting into email accounts and online services without permission, have put data at risk. This affected 63% of reported cases, not just in terms of online security but also caused mental stress. Phishing incidents have led to more stress and worry for Australians. Plus, there was a 28% increase in reported cases where reputation got hurt because of phishing.

Seeing these worrying trends, I've taken on the job of finding a solution using machine learning in order to spot and deal with these bad phishing websites. This way, people can have the tools to protect themselves from cyber fraud and danger.

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

So the Data I chose was found on Kaggle, and it's called 'Phishing and Legitimate URLs.' This dataset holds more than 800,000 web addresses. Among these, 52% are from real and safe websites, while the other 47% are from websites trying to trick or deceive you, Phishing Websites.

In this dataset, there are two Columns: the web address itself and a label called "status." The "status" label is given a number: 0 or 1. When it's 0, it means the website is trying to trick you (phishing). When it's 1, it means the website is safe and real. Both types of websites, the tricky ones and the safe ones, are almost equal in number. This is good because it means there are about the same amount of each type, so it's fair for the computer to learn from them without favouring one over the other, but even with that it will still not going to be as accurate, at least as I first initially hoped it to be, but it wasn’t, so I had to implement something extra into it.

This dataset will help teach the computer to understand and tell the difference between safe websites and those that are trying to deceive us. OR at least an attempt of it.

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

So the first thing I need is to have an equal number of safe and Phishing websites for this to work. I split them into two groups: defined them as Safe and Unsafe for 1 and 0. Then, the code picks a random set of good websites and matches them with an equal number of bad ones. This creates a new dataset with the same amount of both types to do this, I use a function called 'pd.Concat' (short for concatenate) to put them all together.

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

Before using machine learning to analyse URLs, I had to decide what to look for. I wanted to understand what makes a URL unique: like words, numbers, symbols, slashes, and dots. Also, I wanted to know about domain names, such as .com, .org, or .uk etc..

So, I separated these features to study and sort them out. I used half of the data to train the machine learning and the other half to test or predict if a website is safe or phishing.

I created a feature matrix to handle this data and count how many times each word or term appears. To do this, I had to convert slashes into dots and then split the words apart.

After separating the words, I counted how often each word showed up in the data to find the most common terms.

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

I used something called CORPUS, which collects words or text from website addresses (URLs). This doesn't include numbers but just the words from the URLs. It's like having a list of website addresses in text form.

Then, I used a tool called CountVectorizer. It’s an inbuilt tool which assigns a binary number to each word. which helped create a feature matrix that counts how many times each word shows up.

Now, finally after processing the URLs, I can start the training. And as I have mentioned before I will be using half of the data to train and the other half to test the model. I named the features as X and the labels as Y which are either 0 phishing or 1 safe.

I employed a machine learning model called RandomForestClassifier. It learns from the first half of the data by using built in functions called '.fit' and then tries to predict the second half using function called '.predict'. then, by using the 'accuracy\_score' function, this helps me check the accuracy of predicting phishing URLs.

To understand how well the model works, I used this code called 'confusion\_matrix'. It's like a score table that categories four important parts:

True Positives: in this case, When the model correctly predicts phishing URLs.

True Negatives: When the model correctly predicts safe URLs.

False Positives: When the model wasn’t able to predicts a URL as phishing.

False Negatives: When the model wasn;t able to predicts a URL as safe."

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

In conclusion, using machine learning to identify and stop online threats is a big step in making cybersecurity stronger. As phishing attacks keep getting more complicated and trickier to spot, it's important to use technology to stay ahead. By quickly adjusting and using machine learning, we aim to keep up with the always changing world of these tricky websites. This forward-thinking approach not only protects people from the harmful impacts of online threats but also plays a big part in making the online world safer for everyone.

Thank you for listening.