Hubert MĄKA, III rok
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
Wydział IEiT
Koło Naukowe Telephoners

MODEL SZTUCZNEJ INTELIGENCJI ROZPOZNAJĄCY POTENCJALNY PHISHING NA PODSTAWIE ANALIZY SKŁADNI HIPERŁĄCZA

Celem projektu jest stworzenie i wytrenowanie modelu sztucznej inteligencji za pomocą uczenia nadzorowanego, którego przeznaczeniem będzie analiza hiperłącza w celu detekcji, czy jest ono potencjalnym wektorem ataku związanego z próbą nakłonienia ofiary do zaufania fałszywemu podmiotowi oraz przejście do spreparowanych zasobów (np. strony internetowej) stworzonych w celu zdobycia danych wrażliwych atakowanego podmiotu. Przygotowany model będzie rozwiązywał problem klasyfikacji binarnej w oparciu o zadane wejście będące hiperłączem w postaci napisu.

Do wytrenowania modelu zostaną użyte publicznie dostępne etykietowane zbiory danych, które zawierają hiperłącza w postaci tekstowej oraz informacje o nich. Dane zostaną przeanalizowane oraz odpowiednio przetworzone przed przesłaniem ich do stworzonego wcześniej algorytmu.

Do implementacji modelu w języku Python zostaną użyte:

- Biblioteki TensorFlow, Scikit-learn (Uczenie maszynowe)
- Biblioteki Pandas, Numpy, Matplotlib (Analiza i wizualizacja danych)

W celu rozwiązania problemu zostaną przetestowane różne architektury sieci neuronowych oraz algorytmy uczenia maszynowego. Za pomocą bibliotek Pandas, Matplotlib oraz Numpy dane będą poddane odpowiedniemu przygotowaniu i wyczyszczeniu w celu użycia ich do uczenia modeli. Następnie potencjalne modele zostaną zaimplementowane dzięki TensorFlow (sieć neuronowa) oraz Scikit-learn (klasyczne algorytmy klasyfikacji). Ponadto biblioteka Matplotlib umożliwi graficzną wizualizację otrzymanych wyników oraz procesu uczenia modeli. Jako końcowy produkt, który będzie polegał na modelu (funkcji) przyjmującej hiperłącze na wejściu i zwracające na wyjściu prawdopodobieństwo czy dany link jest potencjalnie niebezpieczny zostanie wybrany model z najlepszym wynikiem uzyskanym podczas testów.

Opiekun naukowy referatu:
Dr hab. inż. Marek Natkaniec