EÖTVÖS LORÁND UNIVERSITY

FACULTY OF INFORMATICS

Thesis Topic Registration Form

Student's Data:

Student's Name: Ishaq Muhammad

Student's Neptun code: PXPRGK

Educational Information:

Training programme: Computer Science BSc

I have an internal supervisor

Internal Supervisor's Name: *Md. Easin Arafat*

Supervisor's Home Institution:	***Department of Data Science and Engineering***
Address of Supervisor's Home Institution:	***1117, Budapest, Pázmány Péter sétány 1/C.***
Supervisor's Position and Degree:	*PhD Candidate*

Thesis Title: Phishing Detection Using OSINT-Enhanced Features

Topic of the Thesis:

(Upon consulting with your supervisor, give a 150-300-word-long synopsis os your planned thesis.)

Problem to Be Solved:
Phishing remains one of the most persistent cybersecurity threats, exploiting deception and social engineering to extract sensitive user information. Traditional detection methods rely mainly on textual or structural email features, lacking external intelligence about domain authenticity. This thesis aims to design a phishing detection system using machine learning (ML) and natural language processing (NLP), enriched with open-source intelligence (OSINT) features such as WHOIS data, domain age, DNS records, and reputation sources.

Motivation:
Enhancing phishing detection with OSINT data provides richer context and greater interpretability, allowing the model to detect suspicious patterns missed by text-only classifiers. This approach supports improved detection accuracy and transparency, addressing the growing need for explainable and adaptive cybersecurity systems.

Where It Is Applied:
The system can be applied in email security gateways, corporate cybersecurity tools, and academic research platforms. It may serve as a prototype for integration into real-world phishing defense systems or awareness tools.

How It Will Be Implemented:
The project will be developed using Python and FastAPI. Free and open-source tools will be used, including scikit-learn, spaCy, python-whois, dnspython, Google Safe Browsing API, and PhishTank dataset. The web-based tool will allow users to submit suspicious messages or URLs, process them through the ML model, and display both classification results and OSINT-based explanations. The entire project will run locally and be published as an open-source repository.

Expected Outcomes:
Deliverables include a fully functional prototype web tool, evaluation on public phishing datasets, and a detailed performance comparison between baseline and OSINT-enhanced models. The final thesis will also discuss limitations, dataset biases, and future improvements for large-scale or real-time phishing detection.

Budapest, 2025. 10. 09.