Title : Gambling Web Defacement Detection system (GWDD) in Chiang Mai University Domain

Author(s) :
1. Mr. Naruebordhin Pakwan   **Student ID :** 650510667
2.   **Student ID :**
3.   **Student ID :**

Major : Computer Science

Advisor(s) :
1. Dr.   Worawut Srisukkham
2.
3.

Type of presentation* *(choose 1)* :

☐ Oral Presentation   (เฉพาะ ตัวแทนนศ.ที่สาขาเลือกให้นำเสนอแบบบรรยาย)

☐ Poster   (กรณี นำเสนอผลงานปัญหาพิเศษ/การค้นคว้าอิสระ)

☑ Cooperative Education   (กรณี นำเสนอผลงานสหกิจศึกษา)

ABSTRACT

In recent years, website attacks in the form of embedded links or scripts from online gambling platforms have been steadily increasing, particularly within educational institution websites and the detection process of such embedded gambling content is often time-consuming. Therefore, this project aims to develop a detection system for gambling website embedding within the domain of Chiang Mai University to identify and resolve gambling-related issues more efficiently. The proposed approach employs a Large Language Model, XLM-RoBERTa, trained on two datasets: gambling-related content embedded within websites under educational domains in Thailand, and normal content collected from websites under the Chiang Mai University domain. For the prediction stage, the trained XLM-RoBERTa model was combined with a Mean-Pooling and Rule-Based method. The experimental results indicate that XLM-RoBERTa with a Mean-Pooling combined with Rule-Based prediction method achieves high performance, with an accuracy of 0.9937 on balanced data, a F1-score of 0.9592 on unbalanced data, and a F1-score of 0.9333 on context-based data. Finally, it can be concluded that XLM-RoBERTa combined with Mean-Pooling and Rule-Based prediction is highly effective for detecting gambling websites embedding within Chiang Mai University domain.

*Type of presentation must be matched with an option you choosing on student upload system.

**The abstract can be more than one page and <u>must be approved</u> by project advisor before upload.