



การพัฒนาระบบตรวจจับเว็บไซต์อันตรายแบบเรียลไทม์ ด้วยเทคนิคแมชชีนเลิร์นนิงผ่านเว็บซ็อกเก็ต (Development of Real-time Malicious URL Detection System using Machine Learning via WebSocket)

หลักการและเหตุผล

ปัจจุบันเว็บไซต์อันตรายเป็นภัยคุกคามทางไซเบอร์ที่สำคัญ ผู้ไม่หวังดีมักสร้างเว็บไซต์ปลอมเลียนแบบเว็บไซต์จริงเพื่อหลอกให้ผู้ใช้กรอกข้อมูลสำคัญ เช่น รหัสผ่านหรือข้อมูลทางการเงิน ซึ่งอาจก่อให้เกิดความเสียหายได้ ขณะที่ระบบป้องกันแบบเดิมที่อาศัยรายชื่อเว็บไซต์อันตราย ไม่สามารถรับมือกับโดเมนใหม่ที่ถูกสร้างขึ้นได้อย่างรวดเร็ว โครงการนี้พัฒนาระบบตรวจจับเว็บไซต์อันตราย โดยใช้ Machine Learning วิเคราะห์รูปแบบ URL เพื่อประเมินความเสี่ยงของเว็บไซต์ และใช้ WebSocket เพื่อแจ้งเตือนผู้ใช้แบบ Real-time ก่อนเข้าถึงเว็บไซต์ที่อาจเป็นอันตราย

วัตถุประสงค์

1. เพื่อศึกษาลักษณะและรูปแบบของ URL ที่เกี่ยวข้องกับเว็บไซต์อันตราย
2. เพื่อพัฒนาระบบตรวจจับเว็บไซต์อันตรายโดยใช้เทคนิค Machine Learning
3. เพื่อพัฒนาระบบรับส่งข้อมูลและตรวจสอบ URL แบบ Real-time ด้วยเทคโนโลยี WebSocket
4. เพื่อเพิ่มประสิทธิภาพในการตรวจจับเว็บไซต์อันตรายรูปแบบใหม่ที่ไม่อยู่ในฐานข้อมูล Blacklist

ขอบเขตการศึกษา

1. สถาปัตยกรรมระบบ: พัฒนาด้วย Python แบบ Client-Server ใช้ WebSocket เพื่อสื่อสารและตรวจสอบแบบ Real-time
2. การวิเคราะห์ URL: ใช้เทคนิค NLP เช่น TF-IDF เพื่อสกัดคุณลักษณะจากข้อความ URL
3. การจำแนกข้อมูล: ใช้ Machine Learning เพื่อแยก URL ปกติกับ URL อันตราย
4. ชุดข้อมูลฝึกโมเดล: ใช้ชุดข้อมูลจากแหล่งที่เชื่อถือได้ เช่น Kaggle สำหรับฝึกและทดสอบโมเดล

ผลทดสอบ

```
[+] Client connected from ::1
AI Checking URL: https://google.com
Safe: AI มองว่าปลอดภัย (1.00%)
AI Checking URL: http://kku.ac.th
Safe: AI มองว่าปลอดภัย (1.00%)
AI Checking URL: thecitysphere-montreal.blogspot.com/
Safe: AI มองว่าปลอดภัย (2.97%)
AI Checking URL: thecivilwarhomepagediscussion2824.yuku.com/topic/1265
Safe: AI มองว่าปลอดภัย (8.98%)
AI Checking URL: thecomedyschool.com/
Safe: AI มองว่าปลอดภัย (20.21%)
```

ผลทดสอบเว็บไซต์ที่ปลอดภัย

เอกสารที่เกี่ยวข้อง

1. Socketry. (n.d.). protocol-http. GitHub. สืบค้นจาก อังอิง: <https://github.com/socketry/protocol-http>
2. Phannakon P. (n.d.). สร้างโมเดล Machine Learning ด้วย Scikit-Learn. Medium. สืบค้นจาก อังอิง: <https://medium.com/@phannakon/p/สร้างโมเดล-machine-learning-ด้วย-scikit-learn-2f4c27cfbcd0>

```
[+] Client connected from ::1
AI Checking URL: http://192.168.1.99/login_page.php
ALERT! AI มองว่าเป็นเว็บอันตราย (99.00%)
AI Checking URL: https://secure-update-banking-wallet.com
ALERT! AI มองว่าเป็นเว็บอันตราย (99.66%)
AI Checking URL: http://free-netflix-account.info/login
ALERT! AI มองว่าเป็นเว็บอันตราย (99.70%)
AI Checking URL: premierpaymentprocessing.com/includes/boletto-2via-07-2012.php
ALERT! AI มองว่าเป็นเว็บอันตราย (95.49%)
AI Checking URL: myxxxcollection.com/v1/js/jih321/bpd.com.do/do/l.popular.php
ALERT! AI มองว่าเป็นเว็บอันตราย (92.46%)
AI Checking URL: smilesvoegol.servebbs.org/voegol.php
ALERT! AI มองว่าเป็นเว็บอันตราย (92.83%)
```

ผลทดสอบเว็บไซต์ที่อันตราย

โปรแกรม ซอฟต์แวร์ และไลบรารีที่ใช้



Python



Visual Studio Code

รายวิชา SC362003 Introduction to Computer Networking
เครือข่ายคอมพิวเตอร์ขั้นแนะนำ
อาจารย์ผู้สอน ศ. ดร.จักรชัย ไสอินทร์

ผู้จัดทำ (กลุ่ม 06)

1. 663380563-9 นาย นภัสกรณ์ บุตรโคตร
2. 663380359-8 นางสาวพิพพรรณ ศรีบุญเรือง
3. 663380573-6 นางสาววิยดา มุลกัน
4. 633021103-2 นาย ศุภกร กงชา