การสร้างชุดข้อมูลข้อความแชทฟิชชิ่งด้วยโมเดลการเรียนรู้แบบฟิวช็อต

นายศรัณย์ หงสกุล

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาการบริหารเครือข่ายและความมั่นคงปลอดภัยสารสนเทศ
บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ปีการศึกษา 2567

การสร้างชุดข้อมูลข้อความแชทฟิชชิ่งด้วยโมเดลการเรียนรู้แบบฟิวช็อต

นายศรัณย์ หงสกุล

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาการบริหารเครือข่ายและความมั่นคงปลอดภัยสารสนเทศ
บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ปีการศึกษา 2567

ใบรับรองการค้นคว้าอิสระ

บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

เรื่อง การสร้างชุดข้อมูลข้อความแชทฟิชชิ่งด้วยโมเดลการเรียนรู้แบบฟิวช็อต

โดย นายศรัณย์ หงสกุล

ได้รับอนุมัติให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาการบริหารเครือข่ายและความมั่นคงปลอดภัยสารสนเทศ

คณบดีคณะเทคโนโลยีสารสนเทศและ
..... นวัตกรรมดิจิทัล
(ผู้ช่วยศาสตราจารย์ ดร.สุนันฑา สดสี)

คณะกรรมการสอบการค้นคว้าอิสระ

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.สุมิตรา นวลมีศรี)

..... อาจารย์ที่ปรึกษา
(ผู้ช่วยศาสตราจารย์ ดร.นวพร วิสิฐพงศ์พันธ์)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุนันฑา สดสี)

ชื่อ : นายศรัณย์ หงสกุล
ชื่อการค้นคว้าอิสระ : การสร้างชุดข้อมูลข้อความแชทฟิชชิ่งด้วยโมเดลการเรียนรู้แบบฟิวช็อต
สาขาวิชา : การบริหารเครือข่ายและความมั่นคงปลอดภัยสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
อาจารย์ที่ปรึกษาการค้นคว้าอิสระหลัก : ผู้ช่วยศาสตราจารย์ ดร.นวพร วิสิฐพงศ์พันธ์
ปีการศึกษา : 2567

บทคัดย่อ

ธุรกิจที่พึ่งพาการสื่อสารผ่านข้อความโต้ตอบบนแพลตฟอร์มออนไลน์ มักพบปัญหาข้อความแชทแบบฟิชชิ่งที่รุนแรงมากขึ้น ซึ่งต้องสร้างความตระหนักรู้ข้อความที่เป็นอันตรายให้กับพนักงาน แต่พนักงานเหล่านี้มักเปลี่ยนงานเร็วกว่าที่จะได้รับการอบรมเพียงพอ แม้ปัจจุบันมีการนำเทคนิคการเรียนรู้ด้วยเครื่องมาตรวจจับแล้วก็ตาม แต่ข้อความที่สามารถตรวจจับได้มักอยู่ในรูปแบบอื่น เช่น อีเมล เว็บไซต์ ข้อความสั้น ฯลฯ ซึ่งมีรูปแบบต่างจากการแชทโต้ตอบแบบ 1-1 กับลูกค้า และปัจจุบันยังไม่มีชุดข้อมูลข้อความแชทแบบสาธารณะที่นำมาใช้สร้างโมเดลสำหรับตรวจจับฟิชชิ่งได้ แต่ด้วยความสามารถของระบบ AI แบบ LLM จึงทำให้สามารถเพิ่มจำนวนข้อมูลตัวอย่างที่มีลักษณะคล้ายกันได้ แม้จะมีข้อมูลจริงที่นำมาใช้ตั้งต้นน้อย

งานวิจัยนี้มีวัตถุประสงค์เพื่อจัดทำชุดข้อมูลข้อความแชทแบบฟิชชิ่งจากข้อมูลจริงที่มีจำนวนน้อยมาก แล้วนำมาทดสอบคุณภาพในการนำมาใช้เทรนโมเดลจริง ซึ่งพบว่า การใช้โมเดล LLM แบบ GPT-3.5 Turbo เพิ่มจำนวนแชทได้ 10 เท่า ได้คะแนนประเมินความสมจริงเฉลี่ยที่ 7.68 (เต็ม 10) เมื่อประเมินด้วยโมเดล GPT-4o เมื่อนำไปเทรนโมเดลแบบ Logistic Regression ที่แปลงข้อมูลแบบ MaxAbsScaler จะได้ประสิทธิภาพสูงที่สุด ได้ค่าความเที่ยงตรงที่ 99.87% ความแม่นยำที่ 100% (แบบไบนารี่) คะแนน F1 เท่ากับ 0.99 จึงถือว่าสามารถนำชุดข้อมูลแชทที่สร้างขึ้นนี้ไปใช้ประโยชน์ได้

(มีจำนวนทั้งสิ้น 78 หน้า)

คำสำคัญ : ข้อความแชท ฟิชชิ่ง แพลตฟอร์มออนไลน์ LLM Few-Shot การเพิ่มจำนวนข้อมูล เทคนิคการเรียนรู้ของเครื่อง

_____ **อาจารย์ที่ปรึกษาการค้นคว้าอิสระหลัก**

Name : Mr. Saran Hansakul
Independent Study Title : Using Few-Shot Learning for Creating Phishing Chat Dataset
Major Field : Network and Information Security Management
King Mongkut's University of Technology North Bangkok
Independent Study Advisor : Assistant Professor Dr. Nawaporn Wisitpongphan
Academic Year : 2024

ABSTRACT

Online businesses relying on messaging services encounter increasingly severe phishing chat threats. So, they need to raise awareness among employees to identify malicious messages. But, high employee turnover disrupts these efforts. Even there are many phishing detection tools for emails, websites, SMSs, etc., they cannot be applied to detect phishing from 1-1 chat. And there is still no public phishing chat dataset that has enough quality for training models. However, advancement in AI, in LLM enable data augmenting from limited real-world data through few-shot learning techniques.

This study aims to create a phishing chat dataset from limited real-world data, and evaluate for quality, for training phishing detection models. We were able to increase chat messages by 10 times using GPT-3.5 Turbo model. The quality of dataset received 7.68 out of 10 from GPT-4o model. Upon testing the dataset against various Machine Learning models, Logistic Regression together with dataset transforming using MaxAbsScaler outperformed other models, achieving the accuracy of 99.87% with 100% precision and 0.99 F1 scores. These results imply that our augmented dataset can efficiently be used for training models to detect phishing.

(Total 78 Pages)

Keywords: Chat Messages, Phishing, Online Platform, LLM, Few-Shot, Augmentation, Machine Learning

_____ Advisor

กิตติกรรมประกาศ

การค้นคว้าอิสระในครั้งนี้ สำเร็จลงด้วยดีเนื่องจากได้รับความกรุณาจากอาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.นวพร วิสิฐพงศ์พันธ์ ที่ได้ให้คำปรึกษา ความรู้ แนะนำแนวทางการวิจัย รวมทั้งช่วยปรับปรุงแก้ไขให้สมบูรณ์ยิ่งขึ้น

นอกจากนี้ยังได้รับความกรุณาจากประธานกรรมการ รองศาสตราจารย์ ดร.สุมิตรา นวลมีศรี และกรรมการ ผู้ช่วยศาสตราจารย์ ดร.สุนันทา สดสี ที่ได้สละเวลาอันมีค่ามาให้คำแนะนำ ชี้แนะการศึกษาเพิ่มเติม ตลอดจนชี้แนะการแก้ไขตรวจทานข้อบกพร่อง ทำให้ผู้ค้นคว้าอิสระได้รับประสบการณ์ในการศึกษาเป็นอย่างมาก

ผู้ค้นคว้าอิสระจึงขอกราบขอบพระคุณทุกท่านเป็นอย่างสูง รวมถึงขอบคุณท่านทั้งหลายที่มิอาจเอ่ยนามได้หมดในที่นี้ ที่ช่วยส่งเสริม สนับสนุนให้การค้นคว้าอิสระนี้สำเร็จลุล่วงไปด้วยดี

ศรัณย์ หงสกุล

สารบัญ

หน้า

สารบัญ (ต่อ)

สารบัญตาราง

หน้า

สารบัญรูปภาพ

หน้า

สารบัญภาพ (ต่อ)

สารบัญภาพ (ต่อ)

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัญหาการหลอกลวงหรือใช้กลไกทางจิตวิทยา (Social Engineering) เพื่อหวังผลประโยชน์ผ่านทางข้อความบนโซเชียลมีเดียแพลตฟอร์มต่าง ๆ ทวีความรุนแรงอย่างรวดเร็วมากหลังจากช่วงการแพร่ระบาดของโรคโควิด-19 โดยเฉพาะการโจมตีแบบฟิชชิ่งร่วมกับการใช้อีเมล เว็บไซต์ และแอปพลิเคชันบนอุปกรณ์พกพาที่ปลอมขึ้นมา เพื่อใช้เป็นเครื่องมือในการเข้าโจมตีทางไซเบอร์ด้วยมัลแวร์ เช่น แรนซัมแวร์ โทรจัน และบอท [1] องค์กรที่ดำเนินธุรกิจที่ต้องพึ่งพาการสื่อสารผ่านข้อความนอกจากต้องปรับตัวกับเทคโนโลยีที่เปลี่ยนไป ยังต้องเผชิญกับช่องทางและเทคนิคการโจมตีที่หลากหลายมากขึ้น เช่น การส่งข้อความผ่านแชทในแพลตฟอร์มต่าง ๆ ทั้งทางโซเชียล และบนแพลตฟอร์มจำหน่ายสินค้าออนไลน์ โดยไม่ได้จำกัดเฉพาะผ่านอีเมลอย่างเดียวอีกต่อไป

ปัจจุบันพบว่าแนวโน้มการโจมตีผ่านแชทนั้นมีมากขึ้นเรื่อย ๆ [2] โดยเฉพาะอย่างยิ่งเมื่อมีเทคโนโลยี Generative AI ที่มีความสามารถทางภาษาเหมือนมนุษย์ ทำให้ผู้โจมตีสามารถใช้สร้างข้อความหลอกลวงที่มีความน่าเชื่อถือ มีเทคนิคชั้นเชิงที่ตรวจจับได้ยากได้จำนวนมากอย่างรวดเร็ว [3, 4] ประกอบกับพนักงานขององค์กรที่ตอบแชทเหล่านี้มักมีอัตราการเปลี่ยนงานเร็ว จึงทำให้องค์กรไม่มีเวลาเพียงพอที่จะรับการอบรมสร้างความตระหนักในการแยกแยะข้อความที่เป็นอันตราย ที่ต้องใช้เวลาทำความเข้าใจทั้งด้านความเสี่ยง แรงจูงใจ และการเปลี่ยนทัศนคติ [5] ดังนั้น พนักงานจึงเป็นจุดอ่อนด้านความปลอดภัยขององค์กรที่แก้ไขได้ยาก

ระบบป้องกันเมลอันตรายที่มีอยู่ปัจจุบัน มักตรวจสอบได้เฉพาะช่องทางอีเมลเป็นหลัก โดยมักตรวจสอบจากความผิดปกติที่เห็นได้ชัด เช่น การฝังลิงค์อันตราย การตรวจจากเฮดเดอร์และแหล่งที่มาที่ผิดปกติ การเทียบกับฐานข้อมูลเมลอันตรายที่มีการอัพเดทจากผู้ให้บริการ อย่างไรก็ตามไม่พบว่ามีระบบตรวจจับฟิชชิ่งในข้อความแชทแบบอัตโนมัติในลักษณะเดียวกัน

ผู้วิจัยจึงเล็งเห็นว่าบริษัทควรมีระบบที่ลดการพึ่งพาการตัดสินใจของมนุษย์ที่ขาดความตระหนักด้านความปลอดภัยทางไซเบอร์ และควรเป็นระบบที่สามารถเรียนรู้และพัฒนาความแม่นยำได้อย่างต่อเนื่อง ด้วยการนำกลไกการเรียนรู้ด้วยเครื่อง (Machine Learning) มาใช้ตรวจสอบข้อความอันตรายในแชทที่ต้องการการโต้ตอบรวดเร็ว ซึ่งกำลังเป็นช่องทางสื่อสารสำคัญทางธุรกิจ แต่เนื่องจากการจะใช้กลไกการเรียนรู้ด้วยเครื่องนั้น จะต้องอาศัยการเรียนรู้จากชุดข้อมูลตั้งต้น (Dataset) ที่มีจำนวนมากเพียงพอที่จะพัฒนาความสามารถในการตัดสินใจได้อย่างมีประสิทธิภาพ ซึ่งปัจจุบันพบเพียงชุดข้อมูลฟิชชิ่งของเว็บไซต์ อีเมล และ SMS ที่มีพฤติกรรมและลักษณะต่างจากข้อความแชท และยังไม่พบการเผยแพร่ชุดข้อมูลสาธารณะของข้อความแชทแบบฟิชชิ่ง ดังนั้น ผู้วิจัย

จึงต้องการพัฒนากระบวนการสร้างชุดข้อมูลข้อความแชทที่เป็นฟิชชิ่ง โดยอาศัยการรวบรวมข้อความแชทที่ผู้วิจัยได้รับจริงจากการดูแลเพจโซเชียลต่าง ๆ ซึ่งมีจำนวนจำกัด แล้วนำมาเพิ่มจำนวนข้อมูล (Data Augmentation) ด้วยเทคโนโลยี Generative AI ให้มีขนาดมากพอ [6] โดยให้โมเดลแบบ Large Language Model (LLM) เรียนรู้จากข้อมูลตัวอย่างจำนวนน้อยมาก (Few-Shot Learning) แต่มีคุณภาพเพียงพอ [7] สำหรับนำไปใช้กับเทคนิคการเรียนรู้ด้วยเครื่อง เพื่อเป็นแนวทางสำหรับใช้พัฒนาการตรวจจับข้อความแชทฟิชชิ่งแบบอัตโนมัติต่อไป

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อพัฒนาเทคนิคในการสร้างชุดข้อมูลข้อความแชทแบบฟิชชิ่งที่มีขนาดและคุณภาพเพียงพอสำหรับนำไปใช้พัฒนาโมเดลเพื่อตรวจจับข้อความแชทแบบฟิชชิ่ง จากข้อมูลต้นฉบับที่มีจำนวนจำกัด

1.3 ขอบเขตของการวิจัย

1.3.1 ชุดข้อมูลข้อความแชทฟิชชิ่งที่สร้างในการค้นคว้าอิสระนี้จะเป็นข้อความภาษาอังกฤษ

1.3.2 ชุดข้อมูลตั้งต้นที่มาของข้อความฟิชชิ่งนี้ มาจากข้อความที่ผู้วิจัยได้รับผ่านแพลตฟอร์มแชทต่าง ๆ ได้แก่

1.3.2.1 Facebook Messenger ที่รวมถึงข้อความที่ได้รับผ่านแพลตฟอร์มธุรกิจ Facebook Page

1.3.2.2 Instagram Direct

1.3.2.3 แชทสาธารณะที่เผยแพร่บนอินเทอร์เน็ต

1.3.3 การเพิ่มชุดข้อมูล ใช้โมเดลภาษาขนาดใหญ่ (Large Language Model) ที่เปิดให้ใช้บริการสาธารณะ ซึ่งมีการเรียนรู้มาบางส่วนมาก่อนแล้ว (pre-trained) ได้แก่ OpenAI GPT-3.5-Turbo

1.3.4 การตรวจประเมินคุณภาพของชุดข้อมูลที่เพิ่มมา ด้วยโมเดลภาษาขนาดใหญ่ที่เป็นคนละโมเดลกับที่ใช้เพิ่มขนาดชุดข้อมูล ได้แก่ OpenAI GPT-4o เปรียบเทียบกับการใช้การเรียนรู้ด้วยเครื่องแบบ Supervised ได้แก่ Two-Class Decision Forest, Two-Class Boosted Tree และการใช้ฟีเจอร์หาโมเดลเรียนรู้ด้วยเครื่องที่ให้ประสิทธิภาพการทำนายได้ดีที่สุดแบบอัตโนมัติ (Auto ML) บนแพลตฟอร์ม Azure ML Studio

1.3.5 เครื่องมือที่ใช้ในการวิจัย

1.3.5.1 Azure AI Studio (ชื่อขณะทำการวิจัย ปัจจุบันถูกเปลี่ยนเป็น Azure AI

Foundry) ใช้คัดเลือกและอ้างอิง (Inference) มารันโมเดล LLM แบบ Few-shots ในการทำ Augmentation และตรวจสอบคุณภาพของข้อความแชทที่ Augment มาตรวจคุณภาพความสมจริง (Realism)

1.3.5.2 Azure ML (Machine Learning) Studio ใช้เทรนโมเดล ทดสอบโมเดล แบบ Supervised Learning เพื่อตรวจสอบคุณภาพข้อมูลแชทที่ Augment เพิ่มขึ้นมา

1.3.5.3 Microsoft Excel และ Azure Blob Storage สำหรับจัดเก็บ และประมวลผลชุดข้อมูลข้อความแชท

1.4 คำจำกัดความที่ใช้ในงานวิจัย

1.4.1 ร้านค้าออนไลน์ (Online Stores) หมายถึง แพลตฟอร์มดิจิทัลที่ผู้บริโภคเข้ามาซื้อสินค้าและบริการผ่านอินเทอร์เน็ต เป็นอีกทางเลือกที่สะดวกนอกจากร้านค้าแบบกายภาพ สามารถเข้าถึงได้ผ่านเว็บบราวเซอร์หรือแอพพลิเคชันบนอุปกรณ์พกพา มักใช้การพูดคุยผ่านแอพพลิเคชันแชทเป็นเครื่องมือสื่อสารเพื่อให้บริการลูกค้า ซึ่งเป็นโอกาสที่โดนโจมตีแบบฟิชชิ่งได้ การให้ความสำคัญกับวิธีตรวจจับฟิชชิ่งในช่องทางนี้จึงช่วยสร้างความมั่นคงปลอดภัยทั้งกับผู้บริโภค และธุรกิจที่ให้บริการ [8]

1.4.2 ข้อความแชท (Chat Messages) เป็นรูปแบบของการสื่อสารผ่านช่องทางดิจิตอล ที่ข้อความถูกส่งแทบในทันที (Semi Real-time) ผ่านอินเทอร์เน็ตเพื่อแลกเปลี่ยนระหว่างคู่สนทนาหรือกลุ่มผู้สนทนา ซึ่งมีการใช้ช่องทางนี้บนแพลตฟอร์มหลากหลาย ไม่ว่าจะเป็นสื่อสังคมออนไลน์ แอพพลิเคชั่นสำหรับแชทโดยเฉพาะ หรือบริการแชทที่อยู่บนเว็บไซต์ร้านค้าออนไลน์ [9]

1.4.3 การโจมตีแบบฟิชชิ่ง (Phishing) เป็นการโจมตีทางไซเบอร์ประเภทหนึ่งที่ใช้การหลอกลวงบุคคลให้เปิดเผยข้อมูลความลับ หรือกระทำการใด ๆ ตามที่ผู้โจมตีต้องการ โดยผู้โจมตีปลอมตัวเป็นบุคคลที่น่าเชื่อถือ ทำการผ่านช่องทางสื่อสารแบบดิจิตอลแบบต่าง ๆ ไม่ว่าจะเป็นอีเมล เว็บไซต์ หรือข้อความแชท [10]

1.4.4 โมเดลภาษาขนาดใหญ่ (Large Language Model) หรือ LLM เป็น AI ประเภทหนึ่งที่เรียนรู้จากข้อมูลข้อความจำนวนมหาศาล ทำงานด้วยการข้อมูลป้อนเข้าที่เรียกว่า Prompt เพื่อทำนายว่าคำต่อไปควรจะเป็นอะไรด้วยการคำนวณความน่าจะเป็น [11]

1.4.5 การเรียนรู้ด้วยข้อมูลจำนวนน้อย (Few-Shot Learning) หรือ FSL เป็นเทคนิคที่ช่วยให้โมเดลเรียนรู้จากตัวอย่างที่ติดฉลาก (Labelled) จำนวนน้อยมาก ด้วยการเลียนแบบการเรียนรู้ตามธรรมชาติของมนุษย์ ใช้แก้ปัญหากรณีที่มีข้อมูลตั้งต้นจำกัดกับโมเดล LLM อย่าง GPT-3 เป็นต้นมา [12]

1.4.6 การเพิ่มจำนวนข้อมูล (Data Augmentation) เป็นเทคนิคการเรียนรู้ด้วยเครื่องที่ใช้ขยายขนาดชุดข้อมูลทั้งจำนวนและความหลากหลาย ด้วยการสร้างข้อมูลเพิ่มจากข้อมูลจริงตั้งต้นโดยโมเดล

ภาษาขนาดใหญ่ (LLM) [6]

1.4.7 เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เป็นเทคนิคหนึ่งภายใต้กลุ่มเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) ที่ทำให้เครื่องคอมพิวเตอร์เรียนรู้จากข้อมูล นำมาวิเคราะห์เป็นรูปแบบความสัมพันธ์ เพื่อนำมาใช้ตัดสินใจ โดยที่มนุษย์มีส่วนเกี่ยวข้องน้อย [13]

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้กรอบการทำงานเพื่อสร้างชุดข้อมูล (Dataset) ข้อความแชทสำหรับใช้ประโยชน์ด้านต่างๆ ได้แก่ การนำไปใช้พัฒนาโมเดลการจำแนกการโจมตีฟิชชิ่งผ่านข้อความแชท และการฝึกอบรมสร้างความตระหนักรู้แก่พนักงานร้านค้าออนไลน์

บทที่ 2
แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

การค้นคว้าอิสระเรื่อง การสร้างชุดข้อมูลข้อความแชทฟิชชิ่งด้วยโมเดลการเรียนรู้แบบ Few-Shot เป็นการศึกษาการเพิ่มจำนวนข้อมูลด้วยโมเดลภาษาขนาดใหญ่ให้ได้คุณภาพและปริมาณตามต้องการ สามารถนำมาใช้เทรนโมเดลเพื่อตรวจจับข้อความแชทฟิชชิ่งได้อย่างมีประสิทธิภาพ ผู้วิจัยจึงได้ศึกษาแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง เพื่อใช้เป็นแนวทางการค้นคว้าดังต่อไปนี้

2.1 รูปแบบข้อความแชทที่เข้าข่ายเป็นฟิชชิ่ง

2.2 การเลือกคุณลักษณะสำหรับโมเดลตรวจจับฟิชชิ่ง

2.3 การเพิ่มจำนวนข้อมูลข้อความด้วย LLM

2.4 งานวิจัยที่เกี่ยวข้อง

2.1 รูปแบบข้อความแชทที่เข้าข่ายเป็นฟิชชิ่ง

Oliveira, D. และคณะ [14] ประมวลเทคนิคฟิชชิ่งที่พบในการสนทนาของผู้ขายบนร้านค้าออนไลน์ ดังนี้

2.1.1 การสร้างความหวาดกลัวว่าบัญชีผู้ขายจะมีปัญหา ด้วยการส่งข้อความให้ผู้รับเชื่อว่าบัญชีตนเองโดนยึดครองโดยไม่ได้รับอนุญาต กำลังจะโดนแบน หรือพบกิจกรรมที่ผิดปกติ มักมีข้อความที่เรียกร้องให้กระทบบางอย่างทันที เช่น ให้คลิกลิงค์ไปยังเว็บไซต์หลอกลวง

ตัวอย่างเช่น: Urgent! We found suspicious activity in your account as... If it is not you, please click here to change the password immediately.

2.1.2 การให้ลิงค์สำหรับดาวน์โหลดเอกสารที่เกี่ยวกับการซื้อขาย โดยอ้างว่าลิงค์ที่ส่งมาเพื่อให้โหลดเอกสารทางธุรกิจ ใบคำสั่งซื้อ ใบเสร็จ หรือรายงานสำคัญ แต่เป็นลิงค์ไปยังเว็บอันตรายที่จารกรรมข้อมูลแทน

ตัวอย่างเช่น: I want to purchase many items from you. Please see my order in this spreadsheet....

2.1.3 การปลอมเป็นองค์กร หรือบุคคลที่ผู้รับเชื่อถือ เช่น องค์กรทางการ สถาบันการเงิน หรือคนสนิทอย่างเพื่อนหรือคนในครอบครัว เพื่อสร้างความน่าเชื่อถือก่อนร้องขอข้อมูลส่วนตัว หรือให้ไปที่เว็บอันตราย

ตัวอย่างเช่น: I am <ชื่อลูกค้า> assistant. I want to order your product, but I cannot login to your website. Could you help me ....

2.1.4 การร้องขอข้อมูลส่วนตัว หรือข้อมูลทางการเงินของผู้รับสาร เป็นการถามโดยตรงด้วยการอ้างเหตุผลเพื่อตรวจสอบตัวตน เพื่อปรับปรุงข้อมูลบัญชี หรือเพื่อยืนยันความถูกต้องของธุรกรรม

ตัวอย่างเช่น: I cannot pay to your bank account. Do you have any other account numbers? Could I transfer to your boss's account directly? (เพื่อดูข้อมูลเลขบัญชีของผู้บริหารในบริษัท)

2.1.5 การล่อด้วยรางวัลหรือโปรโมชั่น โดยอ้างว่าผู้รับชนะรางวัล หรือมีสิทธิ์ได้ส่วนลด ได้สิทธิพิเศษต่างๆ เพื่อให้คลิกลิงค์อันตราย หรือแจ้งข้อมูลส่วนตัวกลับเพื่อใช้สิทธิ์

ตัวอย่างเช่น: Congratulations! Your account can get a 20% discount on delivery fee. Please sign in to confirm your discount here <URL อันตราย>

2.1.6 การหลอกลวงในรูปของการแจ้งสถานะการจัดส่ง อ้างว่ามีพัสดุที่ผู้รับรออยู่ แต่ต้องการข้อมูลเพิ่มเติม หรือต้องมีการชำระเงินเพิ่มเติม เพื่อหลอกเอาข้อมูลหรือเงิน

ตัวอย่างเช่น: You have package waiting to be received with Tracking No. XXX but we must confirm your delivery address again. Could you....

2.1.7 การจูงใจให้ย้ายไปคุยแชทในช่องทางที่มีการดูแลความปลอดภัยน้อยกว่า ชวนให้ไปคุยบนแพลตฟอร์มแชทหรือแอพแชทที่มีฟีเจอร์ด้านความปลอดภัยแตกต่างจากแพลตฟอร์มแรก เช่น ย้ายไปแอพส่งข้อความแบบใช้งานส่วนตัว (เช่น LINE, WhatsApp) ที่ตรวจจับการหลอกลวงได้ยาก มักอ้างเหตุผลความเป็นส่วนตัว ความง่ายในการสนทนา หรือจูงใจด้วยข้อเสนอพิเศษ

ตัวอย่างเช่น: I have special offer that cannot be said in this chat. Let's move to WhatsApp by adding my phone number XXX

2.2 การเลือกคุณลักษณะสำหรับโมเดลตรวจจับฟิชชิ่ง

คุณลักษณะ (Feature) ที่ใช้กับเทคนิคเรียนรู้ของเครื่องที่เหมาะสมกับการตรวจจับฟิชชิ่งในข้อความแชทที่คุยตัวต่อตัวระหว่างผู้ขายและผู้ซื้อในร้านค้าออนไลน์นั้น สามารถจำแนกได้เป็นสองกลุ่มใหญ่ ได้แก่ คุณลักษณะด้านข้อความ (Text) กับคุณลักษณะด้านบริบทอื่น (Context) ดังภาพที่ 2-1 ซึ่งอธิบายได้ดังต่อไปนี้

2.2.1 คุณลักษณะด้านข้อความ (Text) เป็นการวิเคราะห์ลักษณะการพิมพ์ตัวอักษร การเรียงคำ การเรียบเรียงประโยค และย่อหน้า โดยแบ่งออกตามขอบเขตการพิจารณาข้อความดังนี้

2.2.1.1 คุณลักษณะของการใช้คำ (Lexical) ใช้วิเคราะห์คำศัพท์ที่ใช้ในข้อความ อันได้แก่ ความถี่ในการใช้คำ การเลือกใช้คำที่มีความยาวผิดปกติกระจายต่อเนื่อง (Word Length Distribution) และการใช้อักขระพิเศษ เพราะข้อความฟิชชิ่งมักใช้คำสำคัญ หรือวลีที่มีเอกลักษณ์เฉพาะในการหลอกลวงผู้รับสาร [15]

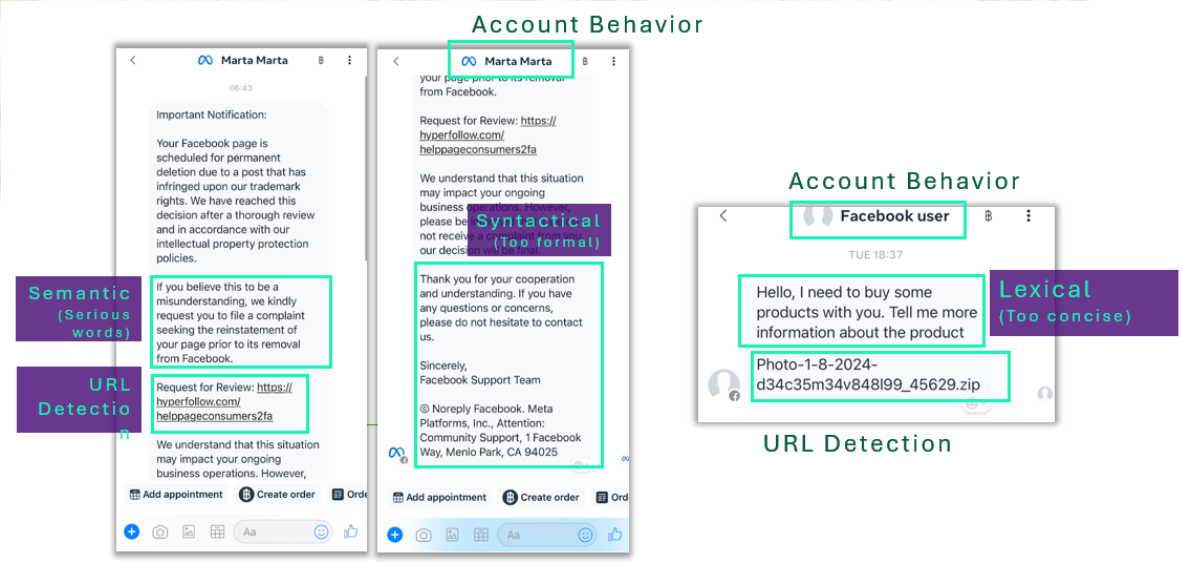
2.2.1.2 คุณลักษณะของโครงสร้างประโยค (Syntactical) ได้แก่ หลักไวยากรณ์ การเรียงลำดับคำ ข้อความฟิชชิ่งมักมีรูปแบบการจัดเรียงประโยคแตกต่างจากข้อความสื่อสารจริง เช่น เรามักพบข้อความฟิชชิ่งใช้หลักไวยากรณ์ผิด หรือมีโครงสร้างประโยคผิดปกติ [16]

2.2.1.3 คุณลักษณะความหมายข้อความ (Semantic) ดูความหมายที่ซ่อนอยู่ในข้อความ ความหมายที่สื่อเมื่อพิจารณาจากบริบทรอบข้าง เช่น ใช้คำพูดที่อ่านรวมดูแล้ว สื่อถึงการเร่งให้ผู้รับสารกระทำบางอย่าง (Urgency) หรือจงใจเลียนแบบรูปแบบการสื่อสารขององค์กรที่ถูกต้อง [15]

2.2.2 คุณลักษณะสภาพแวดล้อม บริบทการสนทนา (Context) พิจารณาองค์ประกอบอื่นที่ส่งมาพร้อมข้อความที่สามารถบ่งบอกการเป็นข้อความฟิชชิ่งได้ ได้แก่

2.2.2.1 การตรวจจับ URL (URL Detection) การที่มีลิงค์ หรือโครงสร้างของลิงค์ที่อยู่ในข้อความ ข้อความฟิชชิ่งมักมีลิงค์อันตรายที่ปลอมในรูปแบบที่คล้ายกับ URL ที่ถูกต้อง [15] เราจึงสามารถวิเคราะห์ตัวโดเมนเนม การใช้โปรโตคอล HTTPS และการใช้บริการลิงค์สั้น (Shortened URL) ในการตรวจจับฟิชชิ่งได้ [17]

2.2.2.2 พฤติกรรมของบัญชีผู้ส่งสาร (Account Behavior) วิเคราะห์พฤติกรรมที่เกี่ยวกับตัวบัญชีผู้ใช้ เช่น ความถี่ที่มีข้อความส่งมาจากบัญชีดังกล่าว ความหลากหลายของผู้รับ และประวัติชื่อเสียงที่ไว้ใจได้ของผู้ส่ง อย่างการพบรูปแบบการส่งข้อความที่ผิดปกติจากบัญชีผู้ใช้ อาจชี้ได้ว่าบัญชีดังกล่าวอาจถูกยึดบัญชีโดยไม่ได้รับอนุญาต (Compromised) หรือเป็นบัญชีที่สร้างขึ้นมาเพื่อส่งข้อความฟิชชิ่งโดยเฉพาะ [16]



ภาพที่ 2-1 ตัวอย่างแชทฟิชชิ่ง และตำแหน่งของคุณลักษณะแต่ละประเภท

องค์ประกอบสำคัญของการตรวจจับการโจมตีคือการคัดเลือกคุณลักษณะ (Feature) เพื่อให้โมเดลเรียนรู้ในการตรวจจับฟิชชิ่ง Zabihimayvan, M. และคณะ [17] ศึกษาการคัดเลือกคุณลักษณะ

ด้วยทฤษฎี Fuzzy Rough Set (FRS) พบว่าสามารถคัดเลือกคุณลักษณะที่นำมาใช้ได้หลายสถานการณ์ (Universal) ถึง 9 คุณลักษณะ ที่ช่วยให้ตรวจจับฟิชชิ่งได้รวดเร็วและมีประสิทธิภาพมากขึ้น

สำหรับการกำกับชนิดข้อมูลนั้น Bhatnagar, P. และคณะ [18] ศึกษาการคัดแยกอีเมลสแปมอย่างมีประสิทธิภาพด้วยการเลือกฟีเจอร์แบบ N-gram ที่ให้น้ำหนักกับข้อความที่สำคัญ พบว่าช่วยเพิ่มความแม่นยำในการตรวจจับสแปมร่วมกับอัลกอริทึม Ensemble Extra Trees โดยได้ความแม่นยำอยู่ที่ 92% และ F1-score อยู่ที่ 92%

2.3 การเพิ่มจำนวนข้อมูลข้อความด้วย LLM

การพัฒนาโมเดลการเรียนรู้ด้วยเครื่องด้านภาษา (Language Modeling) เริ่มต้นจากการนำทฤษฎีด้านข้อมูลมาใช้กับภาษาของมนุษย์ด้วยการศึกษาการนำโมเดลด้านภาษาแบบ n-gram อย่างง่ายมาใช้ทำนายหรือสร้างออกมากเป็นข้อความภาษาธรรมชาติ ซึ่งต่อมามีการใช้วิธีสร้างโมเดลภาษาที่ใช้หลักการทางสถิตินี้มาเป็นพื้นฐานในการทำความเข้าใจและสร้างข้อความภาษาธรรมชาติในหลากหลายรูปแบบ ตั้งแต่การจดจำคำพูด การแปลภาษา ไปจนถึงการดึงข้อมูลมาใช้ จนปัจจุบันพัฒนามาเป็นโมเดลภาษาขนาดใหญ่ (Large Language Model) หรือ LLM ที่ผ่านการเทรนโมเดลมาแล้ว (Pretrained) และให้บริการผ่านเว็บอย่างกว้างขวาง ตัวอย่างเช่น ChatGPT และ GPT-4 ของ OpenAI [14]

จากการศึกษาวิธีการสร้างชุดข้อมูล พบว่าเทคนิค LLM เป็นเทคนิคที่นิยมนำมาใช้เพื่อสร้างชุดข้อมูลที่หลากหลาย Bower, J. [6] ศึกษาการใช้โมเดลภาษาแบบ LLM เพิ่มปริมาณชุดข้อมูลจากข้อมูลตั้งต้นจริง สามารถเพิ่มได้ถึงประมาณ 10 เท่า โดยใช้ Prompt ที่มีความหลากหลายต่างๆ เช่น ให้เรียงคำใหม่ ให้แปลเป็นภาษาหนึ่งและแปลกลับมา เป็นต้น Boumber, D. และคณะ [7] พบว่าสามารถแก้ปัญหาชุดข้อมูลที่นำมาใช้ตรวจจับการหลอกลวง (Deception Detection) ที่มีจำกัดมากได้ด้วยการใช้โมเดลภาษา LLM พร้อมให้ตัวอย่างที่มีคุณภาพสูงแค่ไม่กี่ตัวอย่าง (Few-Shot Learning) ได้อย่างมีประสิทธิภาพ Sapkota, R. และคณะ [19] สำรวจกลไกที่เกิดขึ้นระหว่างการใช้ LLM เติมเต็มข้อมูลประเภทข้อความเพิ่มว่าประกอบด้วย 8 ขั้นตอนได้แก่ (1) การแปลงข้อมูล (Encoding) ข้อความเป็นข้อมูลที่คอมพิวเตอร์เข้าใจได้ ได้แก่ การแบ่งข้อความออกเป็นส่วนย่อยๆ ที่จัดการต่อได้ (Tokenization) และฝังหน่วยข้อมูลย่อยเหล่านี้ในพื้นที่หลายมิติเพื่อโยงความสัมพันธ์ของความหมาย (Embedding) (2) การสร้าง Prompt เพื่อให้ LLM นำข้อมูลหน่วยย่อยๆ ดังกล่าวมาใช้เพิ่มจำนวนข้อมูลในรูปแบบที่เราต้องการ เช่น การเรียบเรียงประโยคใหม่ (Paraphrasing) หรือการเปลี่ยนรูปแบบข้อความ (Stylistic Changes) (3) การเพิ่มจำนวนข้อมูลด้วยการปรับเปลี่ยนข้อความตั้งต้นตามแนวทางที่ได้รับจาก Prompt ให้ได้ตามเป้าหมายที่ระบุไว้ (4) การประมวลผลด้วย

ภาษาธรรมชาติ (Natural Language) ที่นำมาใช้ทั้งการตีความ Prompt และการเรียบเรียงคำตอบให้ได้ความถูกต้องตามหลักภาษา (5) การแปลงข้อความด้วยเทคนิคต่างๆ เพื่อให้ได้ความหลากหลาย (6) การประเมินคุณภาพของข้อความที่ถูกเพิ่มจำนวนมา ได้แก่ การตรวจไวยากรณ์ ความเกี่ยวข้องสัมพันธ์กัน และการตรวจว่าได้ตรงตามเป้าหมายหรือไม่ ทั้งจากระบบอัตโนมัติและการตรวจสอบจากมนุษย์ (7) เมื่อได้ชุดข้อมูลที่มีคุณภาพ ก็นำมาประมวลผลให้อยู่ในรูปที่สามารถนำไปเทรนโมเดลต่อได้ (8) ระบุข้อมูลกำกับประเภทที่จำเพาะ (Metadata) เพื่อให้ทราบที่มาของประเภทหรือกลไกการเพิ่มจำนวนข้อมูลของข้อมูลนั้นๆ สำหรับนำมาใช้เลือกเทรนโมเดลได้อย่างเจาะจงและมีประสิทธิภาพ

สำหรับโมเดลที่นำมาใช้สร้างชุดข้อมูลนั้น Hazell, J. [20] ศึกษาการใช้โมเดล GPT-3, GPT-3.5 และ GPT-4 ของ OpenAI ในการสร้างข้อความอีเมลฟิชชิ่งแบบเจาะจงเป้าหมาย (Spear Phishing) ที่แตกต่างกันจำนวนมากกว่า 600 ข้อความ พบว่า โมเดล GPT-3.5 คุ้มค่าในการนำมาใช้สร้างข้อความฟิชชิ่งได้คุณภาพปานกลางเมื่อเทียบกับค่าใช้จ่าย จึงเป็นตัวเลือกที่ดีในการนำมาใช้สร้างข้อความสำหรับแคมเปญฟิชชิ่งจริง Loukili, S. และคณะ [21] ได้นำโมเดล GPT-3.5, PaLM 2, และ BERT มาใช้สร้างชื่อเรื่อง (Subject) ของอีเมลสำหรับใช้ในการตลาด พบว่า GPT-3.5 และ BERT มีประสิทธิภาพในการสร้างชุดข้อความได้เป็นที่น่าพอใจ

2.4 งานวิจัยที่เกี่ยวข้อง

การศึกษาการนำโมเดลมาใช้ตรวจจับข้อความฟิชชิ่ง Desolda, G. และคณะ [22] ได้พัฒนาเครื่องมือชื่อ APOLLO สำหรับตรวจจับอีเมลฟิชชิ่งพร้อมสร้างคำอธิบายเพื่อแจ้งเตือนผู้ใช้ โดยเลือกใช้โมเดล GPT-4o ของ OpenAI ซึ่งพบว่าได้ความแม่นยำถึง 97% ในการประเมินความสมจริงของข้อความฟิชชิ่ง

สำหรับการตรวจจับการโจมตีแบบฟิชชิ่งด้วยโมเดลแบบกำกับชนิดข้อมูล (Supervised) สามารถดำเนินการได้หลายวิธี Uplenchwar, S. และคณะ [23] พัฒนาเทคนิคที่เรียกว่า Phishing Attack Detection System for Text Messages (PADSTM) และพบว่าการใช้โมเดล Random Forest Classifier ได้ประสิทธิภาพดีที่สุด Gualberto, E.S. และคณะ [15] ศึกษาวิธีตรวจจับฟิชชิ่งแบบหลายด่าน (Multi-Stage) โดยใช้ทั้งเทคนิคการประมวลผลภาษาธรรมชาติ (NLP) และการเรียนรู้ด้วยเครื่อง (ML) พบว่าได้ผลดีกว่าเทคนิคแบบเดี่ยว ได้แก่ Support Vector Machine, Naive Bayes Classifier, Logistic Regression for classification, k-Nearest Neighbor, Decision Trees, Random Forest, Extreme Gradient Boosting (XGBoost), และ Multilayer Perceptron (MLP) ทำให้ใช้จำนวนคุณลักษณะ (Feature) น้อยกว่า และมีค่าใช้จ่ายในการประมวลผลต่ำกว่า Sonowal, G. [16] ศึกษาการตรวจจับอีเมลฟิชชิ่งโดยอาศัยการคัดเลือกคุณลักษณะแบบ Binary Search (BSFS) พบว่าได้ความแม่นยำสูงถึง 97.41% ได้ประสิทธิภาพเหนือกว่าอัลกอริทึมอื่นๆ ณ

ขณะนั้น ได้แก่ Sequential Forward Feature Selection (SFFS) และ Without Feature Selection (WFS) รวมทั้งใช้เวลาในการประเมินฟีเจอร์น้อยกว่า

Aljabri, M. และคณะ [24] ศึกษาการตรวจจับการโจมตีฟิชชิ่งด้วยการเรียนรู้ด้วยเครื่อง และโมเดลการเรียนรู้เชิงลึก (Deep Learning) พบว่าการคัดเลือกคุณลักษณะที่เหมาะสมมีผลต่อประสิทธิภาพในการตรวจจับเป็นอย่างมาก ตัวอย่างเช่น อันดับหน้าเพจ (Page Rank), ข้อความแท็กปักหมุด (Anchor Tag), ลิงค์ไปสู่เว็บข้างนอก (External Link) เป็นต้น และพบว่าอัลกอริทึม Random Forest ให้ความแม่นยำสูงที่สุด

จากการศึกษาวิธีตรวจจับการโจมตีแบบฟิชชิ่งด้วยโมเดลภาษาขนาดใหญ่ที่ผ่านการเรียนรู้มาก่อน (Pre-trained) Wang, Y. และคณะ [25] พัฒนาโมเดลแบบ Transformer ที่ผ่านการเรียนรู้มาก่อนในวงกว้างสำหรับตรวจจับ URL ของเว็บฟิชชิ่งขึ้น เรียกว่า PhishBERT ซึ่งพบว่ามีประสิทธิภาพและความแม่นยำเหนือกว่าโมเดลอื่นๆ ได้แก่ URLNet, Texception, URLTran, และ URL2Vec ทางด้าน Xu, P. [26] ศึกษาการใช้โมเดลแบบ Transformer ในการตรวจจับ URL ฟิชชิ่ง พบว่าได้ความแม่นยำในการตรวจจับสูงถึง 97.3% เหนือกว่าโมเดลตรวจจับทั่วไปในขณะนั้น Ozcan, A. และคณะ [27] ศึกษาการใช้โมเดล Deep Neural Network – Long Short-Term Memory (DNN–LSTM) ในการตรวจจับ URL ฟิชชิ่ง ที่เป็นการผสานการทำงานระหว่างโมเดลที่ใช้การฝังอักขระ (Embedding Character) กับคุณลักษณะสำหรับ NLP ทำให้ได้ประสิทธิภาพในการตรวจจับดีมาก Misra, K. และคณะ [28] ศึกษาการนำโมเดลภาษาที่ผ่านการเทรนมาแล้วในการตรวจจับอีเมลฟิชชิ่ง พบว่าการนำโมเดลมาเรียนรู้เพิ่มเติมด้วยการกำกับประเภทข้อมูล (Priming) ร่วมกับการใช้โมเดลด้านภาษา มีความสำคัญอย่างยิ่งในการเพิ่มประสิทธิภาพการตรวจจับอีเมลฟิชชิ่งที่มีลักษณะแตกต่างจากที่เคยเทรนข้อมูล

ในด้านที่เกี่ยวกับการออกแบบข้อความสอบถามโมเดล LLM (Prompt Engineering) Borges, P. V. และคณะ [29] ให้นิยามของ Prompt Engineer ว่าเป็นกระบวนการสร้างข้อความป้อนเข้าเพื่อปรับประสิทธิภาพและผลลัพธ์ที่ได้จากโมเดล LLM โดยเลือกคำสำคัญ ประโยค และโครงสร้างข้อความที่เหมาะสมเพื่อชี้นำให้โมเดลตอบสนองตามที่เราต้องการ โดยที่ใช้ขั้นตอนการพูดคุยน้อยที่สุด พร้อมทั้งอธิบายองค์ประกอบของ Prompt ที่ควรมีดังนี้: (1) คำสั่ง (Instruction) (2) ข้อมูลป้อนเข้า (Input) (3) บทบาทของโมเดล (Role) (4) รูปแบบของผลลัพธ์ที่ต้องการ (Output) (5) น้ำเสียงและลักษณะการตอบ (Tone and Style) Wang, X. และคณะ [30] พัฒนาเทคนิคชื่อ PromptAgent ที่ปรับโครงสร้าง Prompt ให้ประกอบด้วย 6 ส่วนแบบมืออาชีพ ได้แก่ (1) คำอธิบายงาน (2) นิยามคำศัพท์ (3) แนวทางการแก้ปัญหาหรือทำโจทย์ (4) การจัดการเมื่อมีข้อผิดพลาด (5) ชี้ส่วนที่เน้นย้ำหรือต้องการให้ความสำคัญ (6) รูปแบบคำตอบที่ต้องการ จนได้ F1 score สูงสุดเมื่อเทียบกับ Prompt ที่ภาษามนุษย์ธรรมชาติ และวิธี APE (Automatic Prompt Engineer) ซึ่งสังเกต

ว่าไม่จำเป็นต้องมีการกำหนดบทบาทให้โมเดลเสมอไป

ทางด้านการทำ Meta-Prompting หรือใส่ข้อความระบบ (System Message) สามารถเพิ่มประสิทธิภาพการประมวลผล Prompt ให้ทำงานอัตโนมัติได้ดียิ่งขึ้น Ye, Q. และคณะ [31] พบว่าควรใส่องค์ประกอบ 3 อย่างใน Meta-Prompt ได้แก่ คำบรรยายที่ละเอียด สเปกหรือเงื่อนไขที่ต้องการ และการให้เหตุผลเป็นขั้นตอน พบว่า ช่วยเพิ่มประสิทธิภาพโดยรวมจากเดิมประมาณ 3 – 7%

สำหรับการปรับแต่งค่าการทำงานของโมเดล (Hyperparameter) ช่วยยกระดับการทำงานของ LLM ได้ Keshmirian, A. และคณะ [32] พบว่าการปรับแต่งค่า Temperature สูงขึ้น ช่วยเพิ่มการใช้เหตุผลที่มีประสิทธิภาพเหมือนมนุษย์มากขึ้น โดยเฉพาะเมื่อเข้าใกล้หรือมากกว่า 1

สุดท้ายนี้ การให้โมเดล LLM สร้างข้อความฟิชชิ่ง อาจโดนระบบป้องกัน (Guardrail) ขัดขวางได้ ซึ่ง Mozes, M. และคณะ [33] กล่าวถึงการ Jailbreaking ว่าเป็นการออกแบบ Prompt ให้โมเดล LLM มีพฤติกรรมไม่พึงประสงค์ได้ เช่น การใช้เทคนิค DAN (Do Anything Now) หรือการบอก Role ที่กำหนดบุคลิกของโมเดลให้มองข้ามนโยบายควบคุมการสร้างเนื้อหามาตรฐานของโมเดล

ในการศึกษาครั้งนี้ เราได้ทำการวิจัยศักยภาพของโมเดล LLM อย่าง GPT-3.5-Turbo ในการเพิ่มจำนวนข้อความแชทโดยใช้การเรียนรู้แบบฟิวช็อต ร่วมกับเทคนิค Prompt Engineering ที่หลากหลาย เพื่อปรับแต่งข้อความระบบให้มีประสิทธิภาพในการเพิ่มจำนวนข้อมูลมากที่สุด ซึ่งการประเมินประสิทธิภาพของกระบวนการเพิ่มจำนวนข้อมูล และเพื่อลดความลำเอียงนั้น เราได้ใช้โมเดล LLM อีกตัวหนึ่ง ได้แก่ GPT-4o ในการประเมินความสมจริงของข้อความที่ถูกเพิ่มจำนวนมา และในการแสดงให้เห็นถึงการใช้ประโยชน์จากชุดข้อมูลของข้อความที่ถูกสร้างขึ้นได้จริง เราได้ใช้โมเดล N-gram ในการสกัดคุณลักษณะของชุดข้อมูล จากนั้นจึงใช้โมเดลเรียนรู้ด้วยเครื่องหลายโมเดลมาเทรนกับชุดข้อมูลดังกล่าว พร้อมทั้งทดสอบประสิทธิภาพของโมเดลที่เทรนได้

บทที่ 3
วิธีดำเนินการวิจัย

การค้นคว้าอิสระเรื่อง การสร้างชุดข้อมูลข้อความแชทฟิชชิ่งด้วยโมเดลการเรียนรู้แบบฟิวช็อต เป็นการศึกษาการเพิ่มจำนวนข้อมูลด้วยโมเดลภาษาขนาดใหญ่ให้ได้คุณภาพและปริมาณตามต้องการ สามารถนำมาใช้เทรนโมเดลเพื่อตรวจจับข้อความแชทฟิชชิ่งได้อย่างมีประสิทธิภาพ จึงมีการกำหนดประชากร และขั้นตอนวิธีดำเนินการวิจัยดังนี้

3.1 กรอบแนวคิดการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อจัดทำชุดข้อมูลแชทที่สามารถนำไปสร้างโมเดลจำแนกข้อความแบบฟิชชิ่งได้ ด้วยการนำข้อความแชทจริงที่มีจำนวนจำกัดมาเพิ่มจำนวนข้อความด้วยกลไก AI แบบ LLM โดยมีภาพรวมขั้นตอนการดำเนินงานวิจัย ดังแสดงในภาพที่ 3-1 และมีรายละเอียด ดังนี้

3.1.1 การเก็บรวบรวมข้อมูลแชทจริง

เริ่มจากรวบรวมข้อความแชทฟิชชิ่งจริงที่ผู้วิจัยพบจากเพจเฟซบุ๊กและอินสตาแกรมที่ผู้วิจัยดูแลอยู่ แล้วนำมาจำแนกประเภทด้วยตนเอง ตามโครงสร้างประชากรที่วางแผนไว้ จากนั้นหาชุดข้อมูลข้อความแชทที่ถูกต้อง (Legitimate) จากชุดข้อมูลสาธารณะ หรือจากแพลตฟอร์มที่ผู้วิจัยดูแลอยู่ โดยนำมาผ่านการทำให้ไม่สามารถระบุตัวตนที่แท้จริง (Anonymization) เพื่อให้ได้จำนวนตามที่วางแผนไว้

3.1.2 การคัดเลือกโมเดล

ทดสอบหาโมเดลแบบ LLM ที่มีให้ใช้สาธารณะและที่ผ่านการเทรนมาก่อนแล้ว (Pre-trained) มาทดลองใช้ Prompt สร้างตัวอย่างข้อความแชทที่คล้ายกับข้อความตั้งต้นที่รวบรวมไว้ เพื่อเลือกโมเดลที่เป็นไปได้ในการนำมาใช้สร้างตัวอย่างเพิ่ม (Augmentation)

3.1.3 การเพิ่มจำนวนชุดข้อมูล

ภายหลังจากที่ได้โมเดลที่สำหรับสร้างตัวอย่างเพิ่มแล้ว จากนั้นจึงดำเนินการออกแบบองค์ประกอบของชุดคำสั่ง Prompt ดังนี้

3.1.3.1 System Message (Meta-prompt) ข้อความแม่แบบที่กำกับทุก Prompt ในการสร้างข้อความแชทตัวอย่างแต่ละครั้ง เช่น การสั่งให้โมเดลเป็นผู้เชี่ยวชาญด้านข้อความแชทฟิชชิ่ง ให้สร้างข้อความใหม่ที่มีลักษณะและเป้าหมายคล้ายกับข้อความตัวอย่าง 10 – 100 ข้อความที่แตกต่างกัน

3.1.3.2 Prompt จริง เป็นส่วนที่ป้อนข้อความจริงที่เป็นต้นแบบ

3.1.3.3 Hyperparameter ส่วนของการปรับแต่งค่าพารามิเตอร์ของโมเดลให้ได้คำตอบที่เสถียร เช่น ค่า Temperature

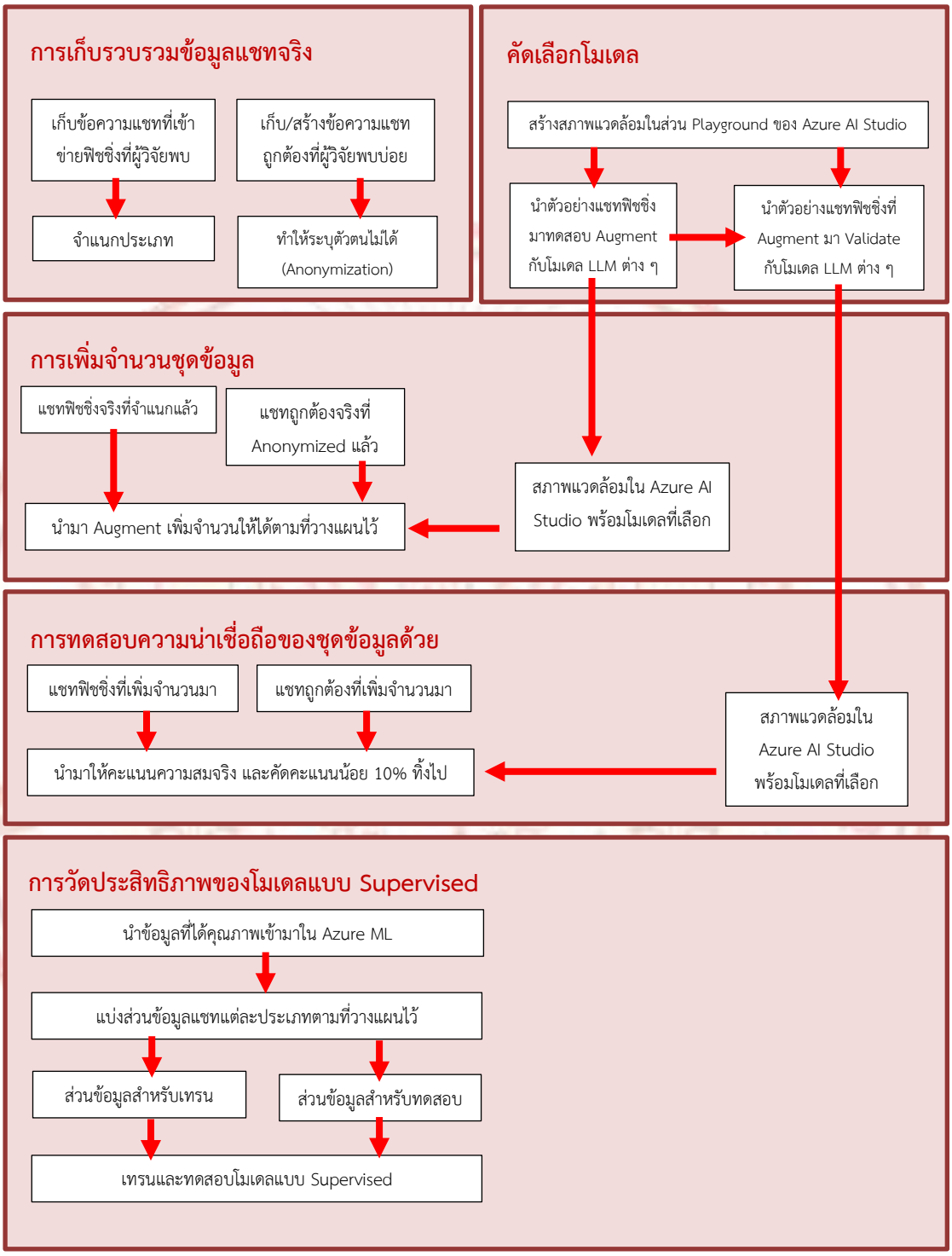
จากนั้นนำชุด Prompt ดังกล่าวมาเพิ่มจำนวนตัวอย่างชุดข้อมูลข้อความแชทให้ได้ตามที่วางแผนไว้

3.1.4 การทดสอบความน่าเชื่อถือของชุดข้อมูลด้วย LLM

คัดเลือกโมเดล LLM แบบ Pretrained ที่ต่างจากโมเดลกับที่ใช้เพิ่มจำนวนชุดข้อมูล เช่น ถ้าโมเดลที่ใช้เพิ่มชุดข้อมูลคือ GPT-3.5 Turbo โมเดลในส่วนการทดสอบความน่าเชื่อถือจะใช้โมเดลที่ต่างกันคือ GPT-4o เป็นต้น เพื่อไม่ให้เกิดการ Bias จากนั้นดำเนินการทดสอบแต่ละโมเดลเบื้องต้นเพื่อหาโมเดลที่สามารถให้คำตอบด้านคุณภาพของข้อความฟิชชิ่งได้ โดยออกแบบชุด Prompt ให้ได้คำตอบที่มีคุณภาพเพียงพอต่อการนำไปใช้ อิงตามองค์ประกอบเหมือนในข้อ 3.1.3 แต่เป็นการให้โมเดลให้คะแนนความสมจริงจาก 0 – 10

3.1.5 การวัดประสิทธิภาพของโมเดลแบบ Supervised

เริ่มจากการนำข้อมูลข้อความแชททั้งแบบถูกต้อง (Legitimate) และแบบฟิชชิ่ง (Phishing) มาตรวจสอบและกำกับฉลาก (Labeled) ให้ถูกต้อง จากนั้นแบ่งข้อมูล (Split Data) เพื่อใช้เป็นข้อมูลสำหรับเทรนโมเดล ทดสอบโมเดล และส่วนที่ไว้ทดสอบตัวอย่างแชทที่เหลือตามที่วางแผนไว้ โดยเลือกใช้โมเดลแบบ Supervised ที่ได้การยอมรับในการตรวจสอบข้อความฟิชชิ่งในกรณีอื่น ๆ ที่มีอ้างอิงจากงานวิจัย เช่น แบบ Random Forest เพื่อนำมาใช้เทรนโมเดล และทดสอบความแม่นยำของโมเดล เพื่อนำมาใช้อ้างอิงถึงคุณภาพของชุดข้อมูลข้อความแชทฟิชชิ่งที่สร้างขึ้นมาอีกทอดหนึ่ง



ภาพที่ 3-1 กรอบแนวคิดงานวิจัย

3.2 การเก็บรวบรวมข้อมูล

ข้อความแชทฟิชชิ่งจริงที่เป็นภาษาอังกฤษ ถูกรวบรวมมาจากข้อความที่ผู้วิจัยได้รับผ่านแพลตฟอร์มแชทต่าง ๆ ได้แก่ Facebook Messenger รวมถึงที่ได้รับผ่านแพลตฟอร์มธุรกิจประกอบด้วย Facebook Page และ Instagram Direct โดยนำมาเป็นชุดข้อมูลตั้งต้นสำหรับสร้างชุดข้อมูลที่ต้องการ โดยรวบรวมให้ได้อย่างน้อยตามจำนวนที่วางแผนไว้สำหรับข้อความแชททั้งสองกลุ่ม ได้แก่

3.2.1 ข้อความแชทแบบฟิชชิ่งรวม 50 ข้อความ ซึ่งแบ่งเป็น 5 ประเภท (Class) ดังนี้:

3.2.1.1 ข้อความข่มขู่ให้หวาดกลัว จำนวน 10 ข้อความ

3.2.1.2 ข้อความหลอกลวงว่าได้รางวัล หรือได้ของฟรี จำนวน 10 ข้อความ

3.2.1.3 ข้อความชักชวนให้ออกไปติดต่อช่องทางอื่น จำนวน 10 ข้อความ

3.2.1.4 ข้อความที่มีการแนบไฟล์หรือลิงค์อันตราย จำนวน 10 ข้อความ

3.2.1.5 ข้อความโฆษณา และข้อความอื่น ๆ ที่ไม่ต้องการ จำนวน 10 ข้อความ

3.2.2 ข้อความแชทที่ถูกต้อง ไม่เป็นอันตราย (Legitimate) จำนวน 50 ข้อความ

การสร้างชุดข้อมูลสำหรับนำไปสร้างโมเดลนี้จะอ้างอิงแนวทางการพิจารณาจากสองเงื่อนไขดังต่อไปนี้

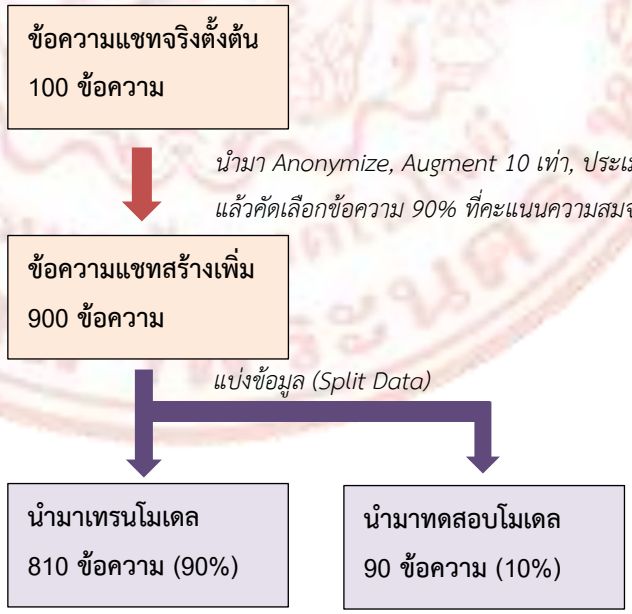
จำนวนข้อความที่นำไปใช้สร้างโมเดลควรมีขนาด 50 – 1000 เท่าของจำนวนประเภท (Class) ของข้อความ [6] ดังนั้น หากประเภทของแชทฟิชชิ่งมีจำนวน 5 ประเภท จำนวนข้อความที่เหมาะสมสำหรับนำไปใช้สร้างโมเดลควรมีจำนวนอย่างน้อย 250 – 5000 ข้อความ

จำนวนข้อความสำหรับนำไปใช้สร้างโมเดลควรมีขนาด 10 – 100 เท่าของจำนวนคุณลักษณะที่สนใจ (Feature) [6] ซึ่งจำนวนคุณลักษณะหรือฟีเจอร์ขึ้นอยู่กับความสนใจและวัตถุประสงค์ของผู้นำชุดข้อมูลไปใช้งานต่อ เช่น หากพิจารณาว่าควรมีอย่างน้อย 5 ฟีเจอร์ ชุดข้อมูลที่สร้างควรมีจำนวนประมาณ 50 – 500 ตัวอย่าง

จากเงื่อนไขในการสร้างข้อมูลทั้งสองเงื่อนไข ทำให้สามารถสรุปแผนการสร้างชุดข้อมูลเพิ่ม ดังตารางที่ 3-1 และการแบ่งข้อมูลที่ถูกเพิ่มจำนวนมาได้ และคัดเลือกเฉพาะที่ได้รับการประเมินคะแนนความสมจริงสูงสุดในส่วน 90% แรกเพื่อนำไปทดสอบด้วยการเทรนโมเดลเรียนรู้ด้วยเครื่องสำหรับหาประสิทธิภาพของโมเดลที่ใช้ชุดข้อมูลสร้างใหม่นี้ ตามแผนผังในภาพที่ 3-2

ตารางที่ 3-1 จำนวนข้อความแชทตั้งต้นที่ควรรวบรวมได้ ข้อความที่สร้าง และข้อความที่นำมาใช้ในการเทรนโมเดล

ประเภทข้อความแชท (Class)	จำนวนข้อความแชทตั้งต้น	จำนวนข้อความแชทที่สร้างเพิ่ม (Augmented)	สัดส่วนการแบ่งข้อมูล	
			จำนวนข้อความที่นำมาเทรน (90%)	จำนวนข้อความที่นำมาทดสอบโมเดล (10%)
ข่มขู่ให้หวาดกลัว	10	90	81	9
หลอกว่าได้รางวัลหรือได้ของฟรี	10	90	81	9
ชักชวนให้ออกไปติดต่อช่องทางอื่น	10	90	81	9
แนบไฟล์ ลิงค์อันตราย	10	90	81	9
โฆษณา และข้อความที่ไม่ต้องการอื่น ๆ	10	90	81	9
แชทที่ถูกต้อง	50	450	405	45
รวมทั้งหมด	**100**	**900**	**810**	**90**



ภาพที่ 3-2 แผนผังการจัดสรรชุดข้อมูลเพื่อนำมาประเมินประสิทธิภาพของโมเดล Machine Learning ที่เรียนรู้จากชุดข้อมูลที่ Augment ทั้งหมด

3.3 ขั้นตอนและเครื่องมือที่ใช้ในการวิจัย

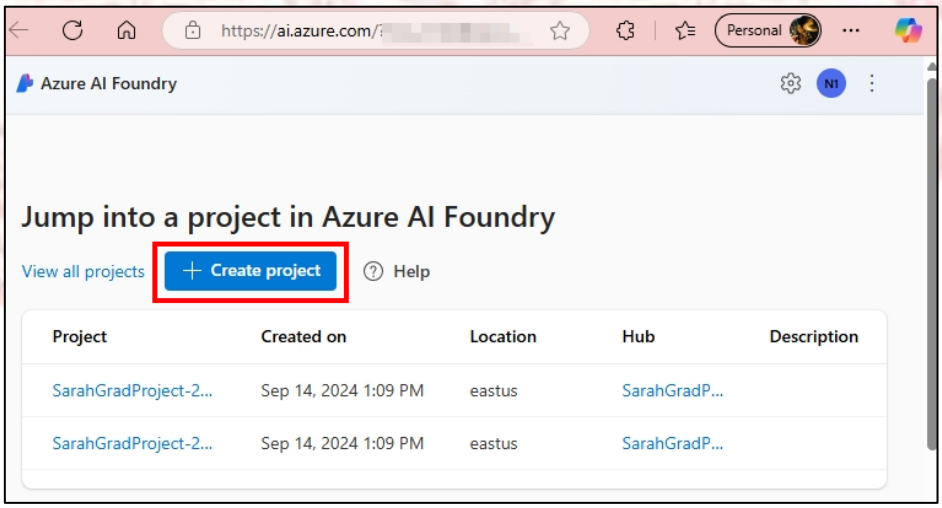
3.3.1 รวบรวมข้อความแชทฟิชชิ่งจริงที่ผู้วิจัยพบจากเพจเฟซบุ๊กและอินสตาแกรมที่ผู้วิจัยดูแลอยู่ แล้วนำมาจำแนกประเภทด้วยตัวเอง ตามประชากรที่วางแผนไว้ข้างต้น

3.3.2 หาชุดข้อมูลข้อความแชทที่ถูกต้อง (Legitimate) จากชุดข้อมูลสาธารณะ หรือจากแพลตฟอร์มที่ผู้วิจัยดูแลอยู่ โดยนำมาผ่านการทำให้ไม่สามารถระบุตัวตนที่แท้จริง (Anonymization) เพื่อให้ได้จำนวนตามที่วางแผนไว้

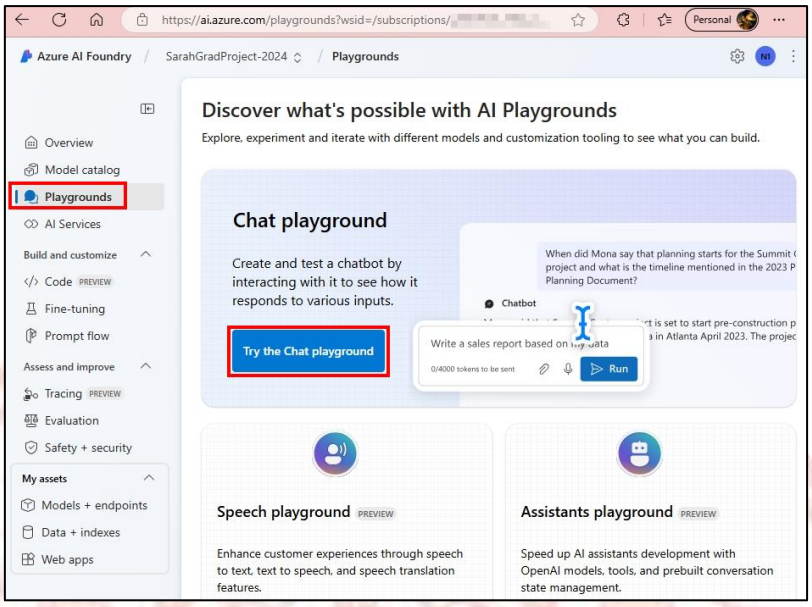
3.3.3 ทดสอบหาโมเดลแบบ LLM ที่มีให้ใช้สาธารณะและที่ผ่านการเทรนมาก่อนแล้ว (Pre-trained) มาทดลองใช้ Prompt สร้างตัวอย่างข้อความแชทที่คล้ายกับข้อความตั้งต้นที่รวบรวมไว้ เพื่อเลือกโมเดลที่เป็นไปได้ในการนำมาใช้สร้างตัวอย่างเพิ่ม (Augmentation) โดยใช้ Azure AI Studio (เป็นชื่อขณะที่ทำการวิจัย ซึ่งปัจจุบันเปลี่ยนชื่อเป็น Azure AI Foundry) ใช้เลือกและรันโมเดล LLM แบบ Few-shots ในการทำ Augmentation

นอกจากนี้ เครื่องมือ Azure AI Foundry ยังสามารถนำมาใช้ตรวจสอบคุณภาพ (Validation) ของข้อความแชทที่ Augment มาได้อีกด้วย โดยวิธีการดำเนินการมีดังนี้

3.3.3.1 เริ่มจากเข้าไปที่เว็บไซต์ https://ai.azure.com ล็อกอินด้วย Subscription ของ Azure ที่มีอยู่แล้ว กด Create Project ซึ่งถ้ายังไม่ได้สร้าง AI Hub ระบบจะให้ตั้งชื่อ Hub ใหม่สำหรับรวมโปรเจ็กต์ที่เกี่ยวข้อง จากนั้นเข้าโปรเจ็กต์แล้วไปที่ Chat Playground ดังภาพที่ 3-3,4

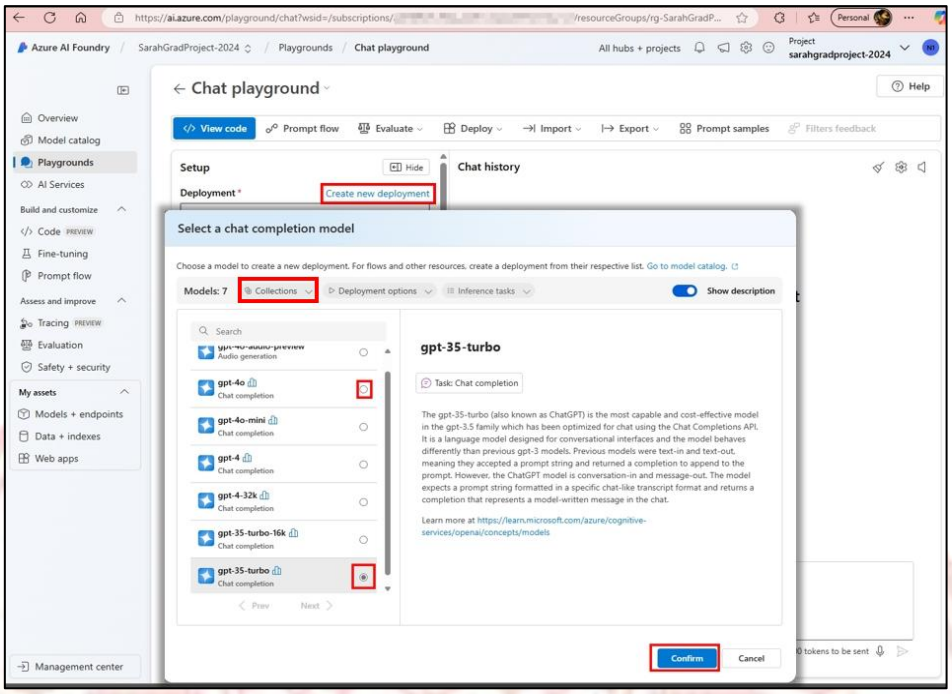


ภาพที่ 3-3 หน้าแรกของ Azure AI Foundry สำหรับสร้างโปรเจ็กต์ และ Hub ใหม่ถ้ายังไม่เคยสร้างไว้



ภาพที่ 3-4 วิธีเข้าหน้า Chat Playgrounds จากเมนูหลักของ Azure AI Foundry

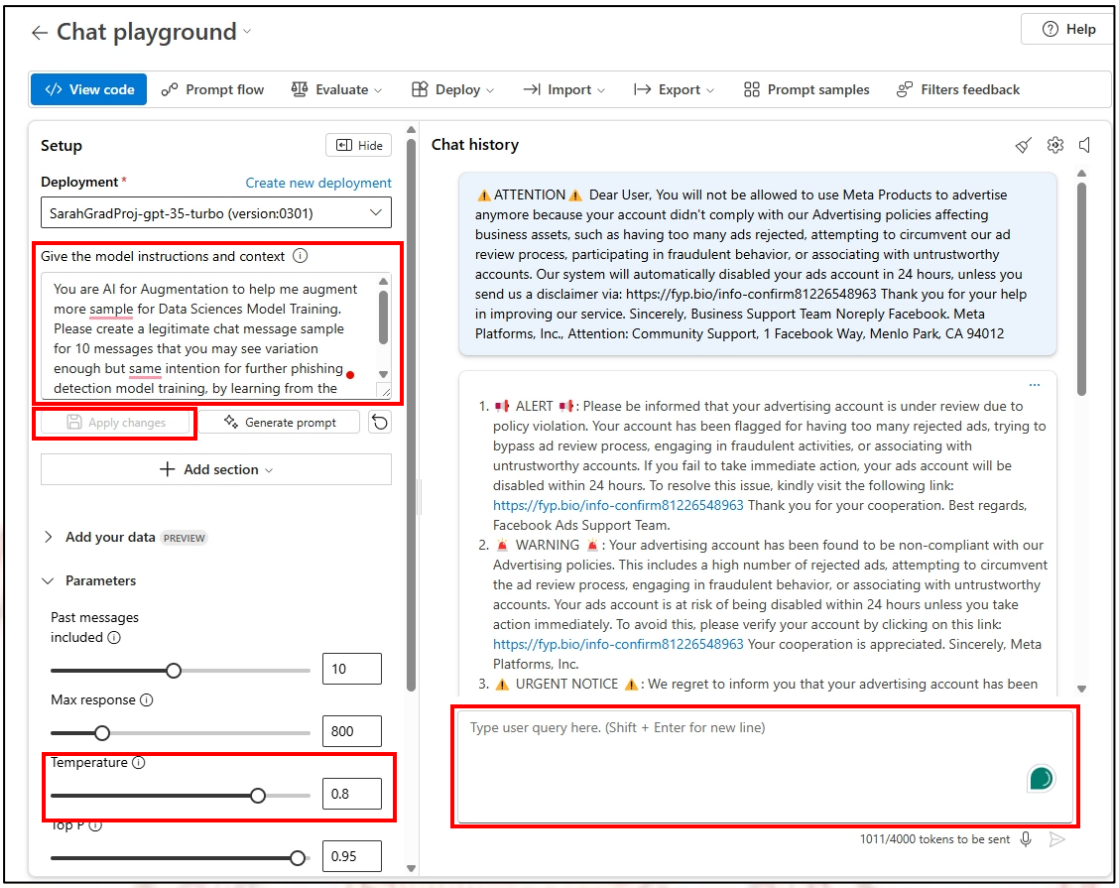
3.3.3.2 จากหน้า Chat Playground ถ้ายังไม่เคยเลือกติดตั้งโมเดล LLM ที่จะ Inference (Deployment) ให้เลือก Create New Deployment จากนั้นเลือก Collection ที่จะคัดกรอง เช่น OpenAI แล้วเลือกโมเดลที่ต้องการ ในงานวิจัยนี้จะเลือกโมเดล gpt-35-turbo ในการเพิ่มจำนวนข้อความแชท และโมเดล gpt-4o ในการให้คะแนนความสมจริงของข้อความแชท เป็นต้น ดังแสดงในภาพที่ 3-5



ภาพที่ 3-5 วิธีสร้าง Deployment ใหม่เพื่อเลือกโมเดล LLM ที่จะอ้างอิง

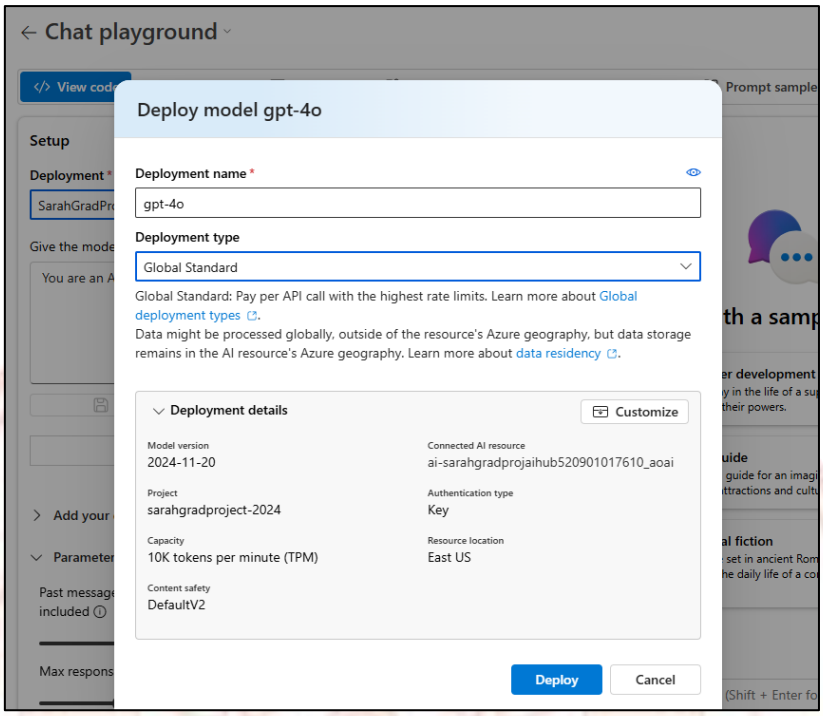
3.3.3.3 ใส่ prompt คำสั่งตั้งต้นในช่อง Give the model instructions and context (หรือช่อง System message กรณีที่เป็น Azure AI Studio เวอร์ชันเดิมก่อนอัพเดท) และตั้งค่าพารามิเตอร์ด้านล่าง อันได้แก่ ค่า Temperature เพื่อปรับระดับความคิดสร้างสรรค์หรือความเที่ยงตรงตามต้องการ (ในที่นี้ ใช้ 0.8 สำหรับการเพิ่มจำนวนข้อความแชท และ 0.3 สำหรับให้คะแนนความสมจริงของข้อความแชท)

3.3.3.4 กด Apply changes เพื่อเริ่มดำเนินการ ด้วยการใส่ข้อความแชทจริงที่ต้องการเป็นตัวอย่างตั้งต้น (กรณีต้องการเพิ่มจำนวนข้อความแชท) ในช่อง User Query ในส่วนของ Chat History แล้วสังเกตคำตอบที่ได้เพื่อปรับแต่งค่าเพิ่มให้ได้คำตอบที่ต้องการ หรือเมื่อได้คำตอบที่ต้องการให้คัดลอกนำออกมาบันทึกไว้บนเครื่องตนเอง (เนื่องจาก Chat Playground จะไม่บันทึกคำตอบไว้) ดังแสดงในภาพที่ 3-6

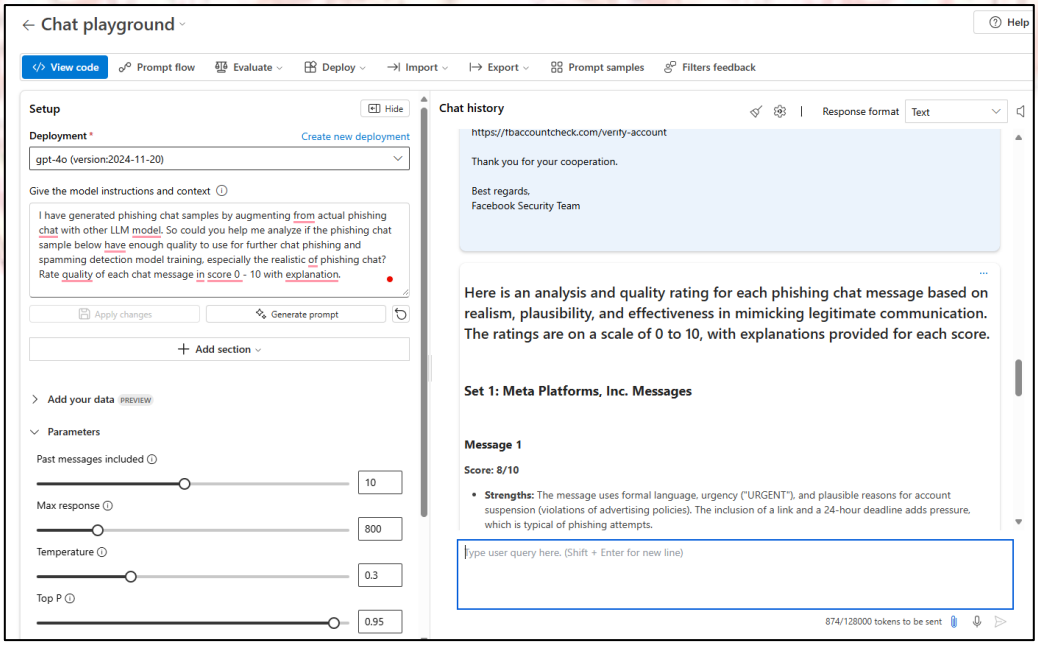


ภาพที่ 3-6 วิธีตั้งค่า Chat Playground หลังสร้าง Deployment เพื่อใช้งานจริง กรณีเพิ่มจำนวนข้อความแชท

3.3.4 เมื่อได้โมเดลที่สามารถสร้างชุดข้อมูลเพิ่มได้ผลดี (เช่น ไม่ติดระบบป้องกันหรือ Guardrail ต่างๆ) และได้ค่า Hyperparameter เช่น ค่า Temperature ที่เหมาะสม จนสามารถเพิ่มจำนวนชุดข้อความแชทได้ตามแผนแล้ว ให้นำข้อความที่ถูกเพิ่มจำนวนมาประเมินคะแนนความสมจริงตามสเกล 0 – 10 (คะแนนเต็ม 10 คะแนน) ด้วยแพลตฟอร์ม Azure AI Foundry ในส่วน Playground นี้ด้วยเช่นกัน เพียงแต่เปลี่ยนโมเดลที่จะอ้างอิงเป็น GPT-4o เพื่อลดความลำเอียง และให้ได้เหตุผลประกอบการให้คะแนนสำหรับสร้างความน่าเชื่อถือ และนำไปวิเคราะห์ต่อได้ พร้อมทั้งปรับค่า Temperature เป็น 0.3 เพื่อให้ได้ความเที่ยงตรงในการให้คะแนนอย่างเหมาะสมดังภาพที่ 3-7 และ 3-8



ภาพที่ 3-7 เลือก Deploy โมเดล gpt-4o ใหม่สำหรับใช้ประเมินความสมจริงของข้อความแชทที่ถูกเพิ่มจำนวนมา



ภาพที่ 3-8 การตั้งค่าใน Chat Playground เพื่อประเมินคะแนนความสมจริงของข้อความแชทที่ถูกเพิ่มจำนวนมา

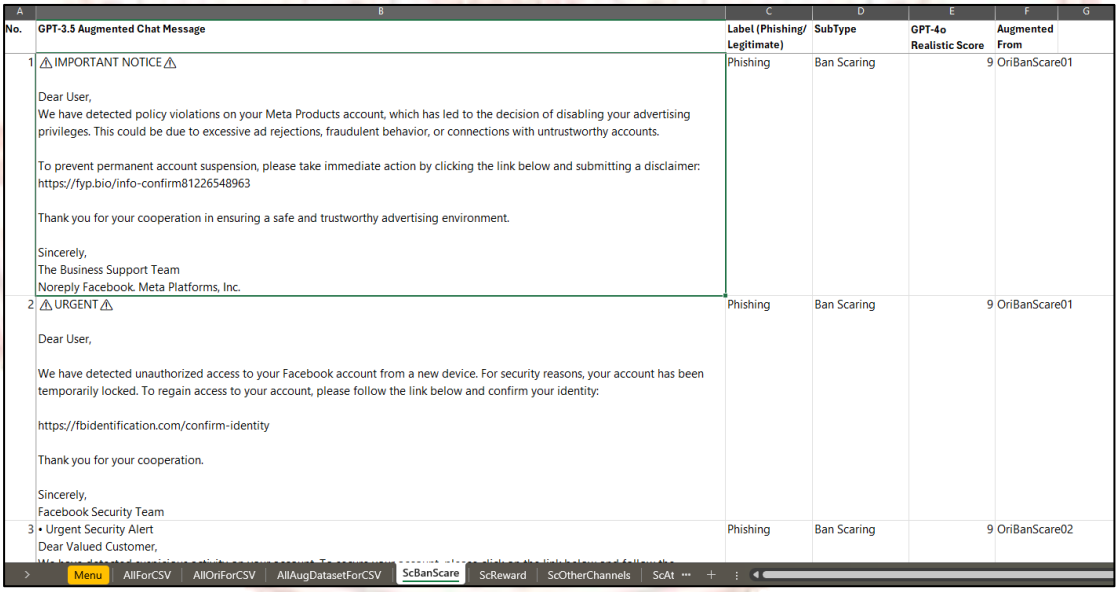
3.3.5 นำข้อมูลมาเรียบเรียงใน Microsoft Excel ตามภาพที่ 3-9 ถึง 3-12 ดังนี้

3.3.5.1 แบ่งกลุ่มข้อความแชท ทั้งข้อความตั้งต้น และข้อความที่ถูกเพิ่มจำนวน แยก Sheet ตามประเภทที่วางแผนไว้

3.3.5.2 ทำคอลัมน์กำกับฉลากบอกชนิดของข้อความ (Label) ที่จะใช้ในการจำแนกด้วยโมเดลการเรียนรู้ด้วยเครื่องต่อไป ได้แก่ “Phishing” (ข้อความฟิชชิ่ง) และ “Legitimate” (ข้อความที่ถูกต้อง)

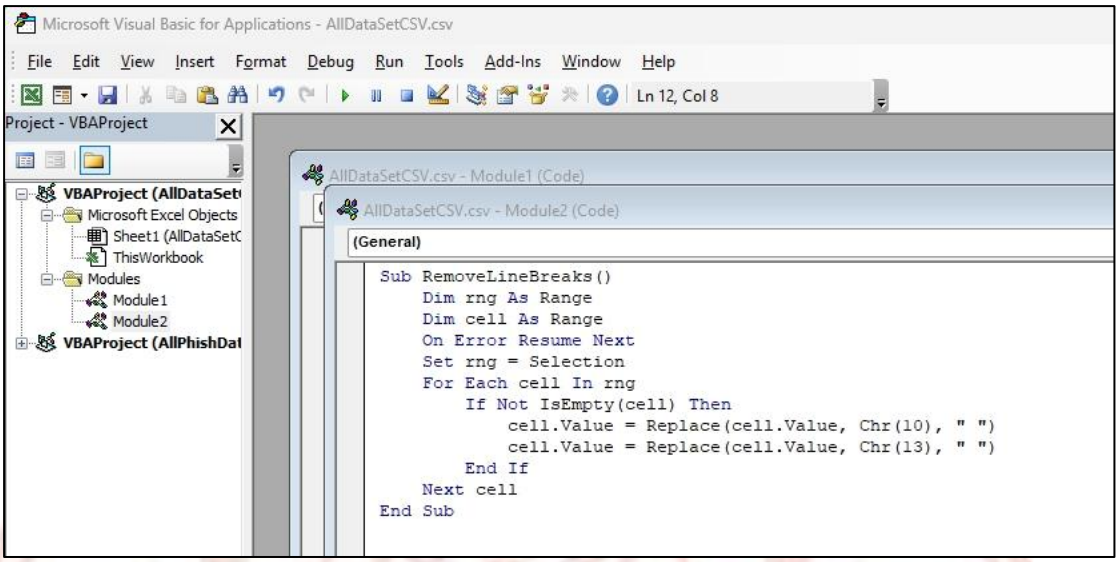
3.3.5.3 กำกับคะแนนความสมจริงที่ประเมินได้กับข้อความที่ถูกเพิ่มจำนวนมา จากนั้นคัดเฉพาะข้อความที่ได้คะแนนมากที่สุด 90% ในแต่ละประเภท เพื่อให้ได้จำนวนข้อมูลที่จะนำไปเทรนโมเดลการเรียนรู้ด้วยเครื่องต่อไปตามที่วางแผนไว้

3.3.5.4 รวบรวมข้อความที่ถูกเพิ่มจำนวนทั้งหมดมาอยู่ในอีกไฟล์หนึ่ง ในชีทเดียวกัน แล้วแปลงเป็นไฟล์ CSV โดยใช้มาโคร Visual Basic Script เพื่อแปลงข้อความนำการขึ้นบรรทัดใหม่ออก และแปลงสัญลักษณ์แบ่งคอลัมน์ข้อมูล (Delimiter) จาก , เป็น | เป็นต้น

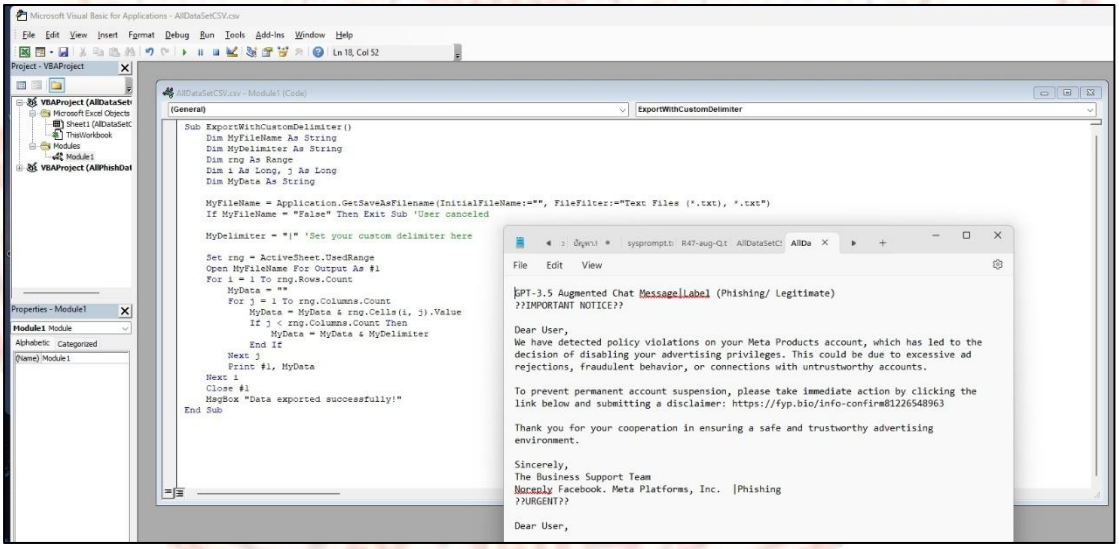


No.	GPT-3.5 Augmented Chat Message	Label (Phishing/ Legitimate)	SubType	GPT-4o Realistic Score	Augmented From
1	⚠ IMPORTANT NOTICE ⚠ Dear User, We have detected policy violations on your Meta Products account, which has led to the decision of disabling your advertising privileges. This could be due to excessive ad rejections, fraudulent behavior, or connections with untrustworthy accounts. To prevent permanent account suspension, please take immediate action by clicking the link below and submitting a disclaimer: https://fyp.bio/info-confirm8122654896 3 Thank you for your cooperation in ensuring a safe and trustworthy advertising environment. Sincerely, The Business Support Team Noreply Facebook. Meta Platforms, Inc.	Phishing	Ban Scaring	9	OriBanScare01
2	⚠ URGENT ⚠ Dear User, We have detected unauthorized access to your Facebook account from a new device. For security reasons, your account has been temporarily locked. To regain access to your account, please follow the link below and confirm your identity: https://fbidentification.com/confirm-identity Thank you for your cooperation. Sincerely, Facebook Security Team	Phishing	Ban Scaring	9	OriBanScare01
3	• Urgent Security Alert Dear Valued Customer,	Phishing	Ban Scaring	9	OriBanScare02

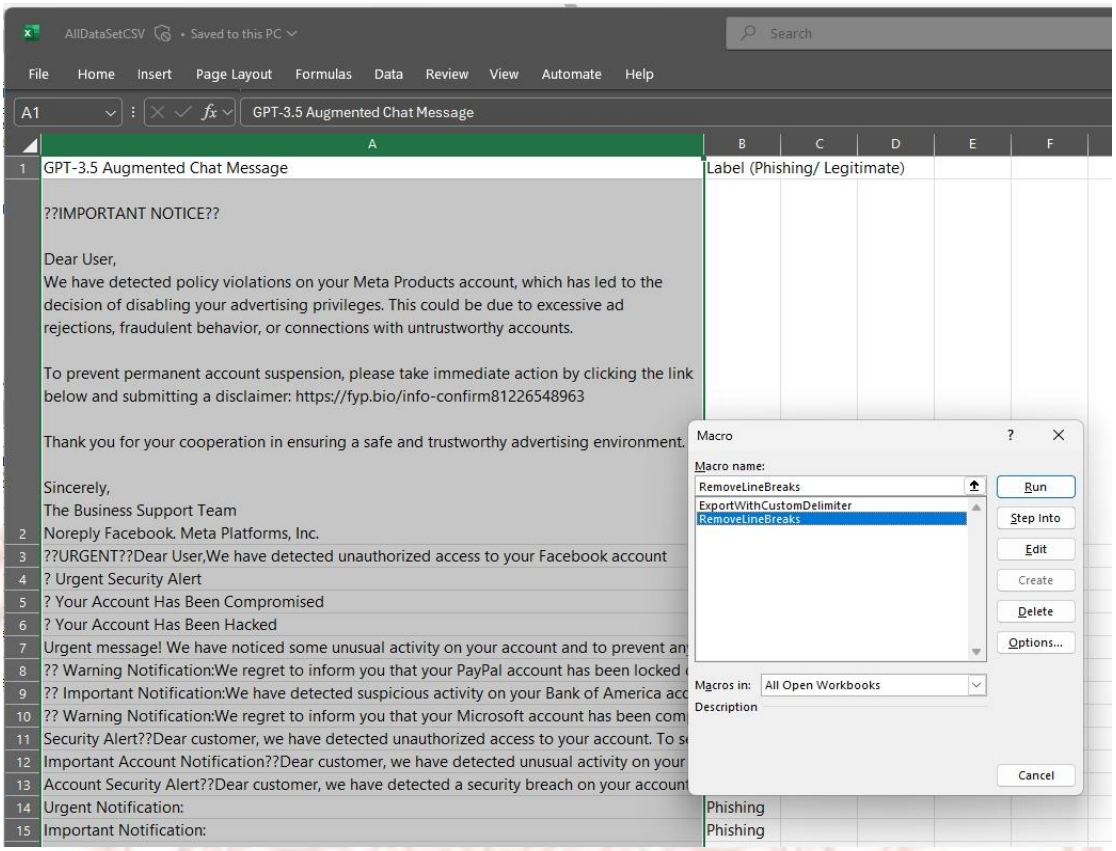
ภาพที่ 3-9 การใช้ Microsoft Excel บันทึกข้อความแชทที่เพิ่มจำนวนขึ้นมา พร้อมกำกับฉลาก และคะแนนความสมจริงที่ได้ เพื่อคัดข้อความที่สมจริงมากที่สุด



ภาพที่ 3-10 การสร้างมาโครสำหรับเอาการขึ้นบรรทัดใหม่ในข้อความแชท ในไฟล์ที่แยกออกมาเป็น .csv เรียบร้อยแล้ว



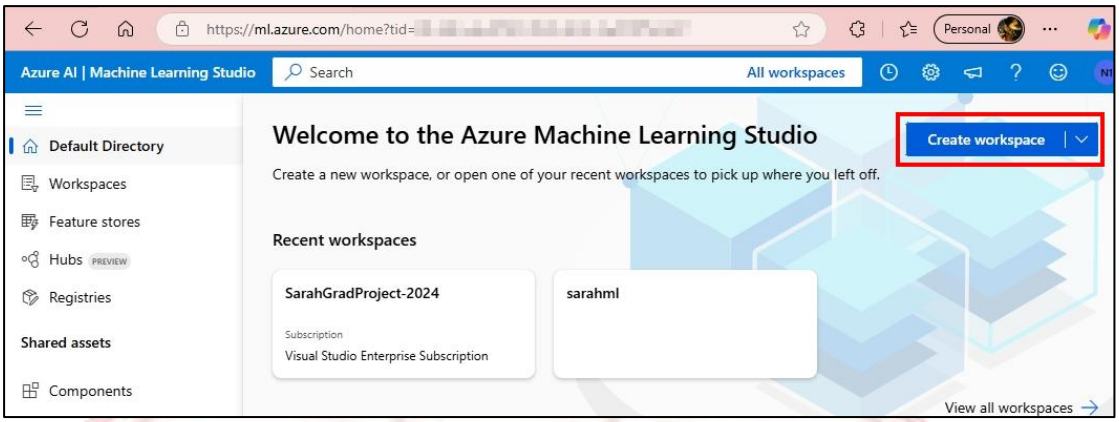
ภาพที่ 3-11 การสร้างมาโครสำหรับเปลี่ยนอักขระที่ทำหน้าที่เป็น Delimiter จาก , เป็น | ในไฟล์ที่แยกออกมาเป็น .csv เรียบร้อยแล้ว พร้อมเนื้อหาไฟล์ .csv หลังแปลง



ภาพที่ 3-12 วิธีรัน Macro ที่เตรียมไว้ เพื่อดัดแปลงข้อมูลในไฟล์ .csv ใน Microsoft Excel ให้อยู่ในรูปที่พร้อมนำเข้า Azure ML Studio

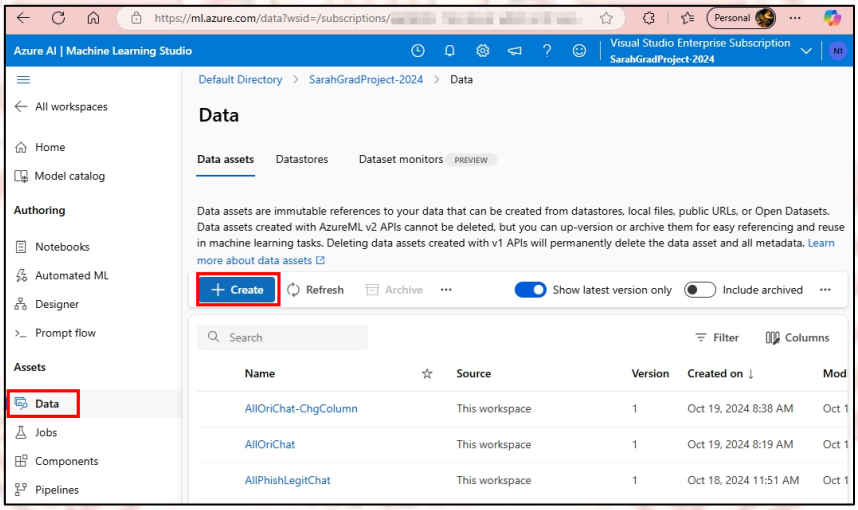
3.3.6 เมื่อได้ไฟล์ .csv ของชุดข้อมูลข้อความแชทที่ถูกเพิ่มจำนวนและคัดเฉพาะข้อความที่ได้รับคะแนนความสมจริงมากที่สุดแล้ว จะใช้ Azure ML (Machine Learning) ในการเทรนโมเดลแบบ Supervised Learning และติดตั้งโมเดลเพื่อตรวจสอบคุณภาพข้อมูลแชทที่ Augment เพิ่มขึ้นมา ดังนี้

3.3.6.1 เข้าไปที่เว็บไซต์ https://ml.azure.com ล็อกอินด้วย Subscription ของ Azure ที่ตนเองมีอยู่ ในหน้าแรกนี้ ถ้ายังไม่เคยสร้าง Workspace ให้กดปุ่ม Create workspace เพื่อสร้างให้เรียบร้อยก่อน ดังแสดงในภาพที่ 3-13

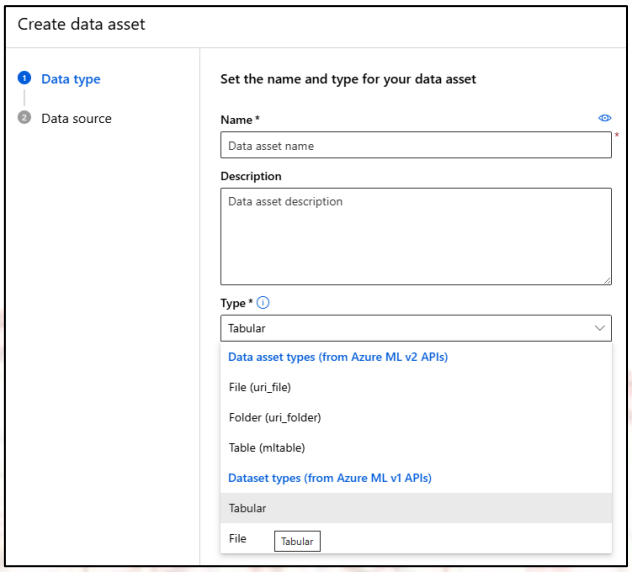


ภาพที่ 3-13 หน้าแรกของ Azure ML Studio และวิธีสร้าง Workspace ใหม่

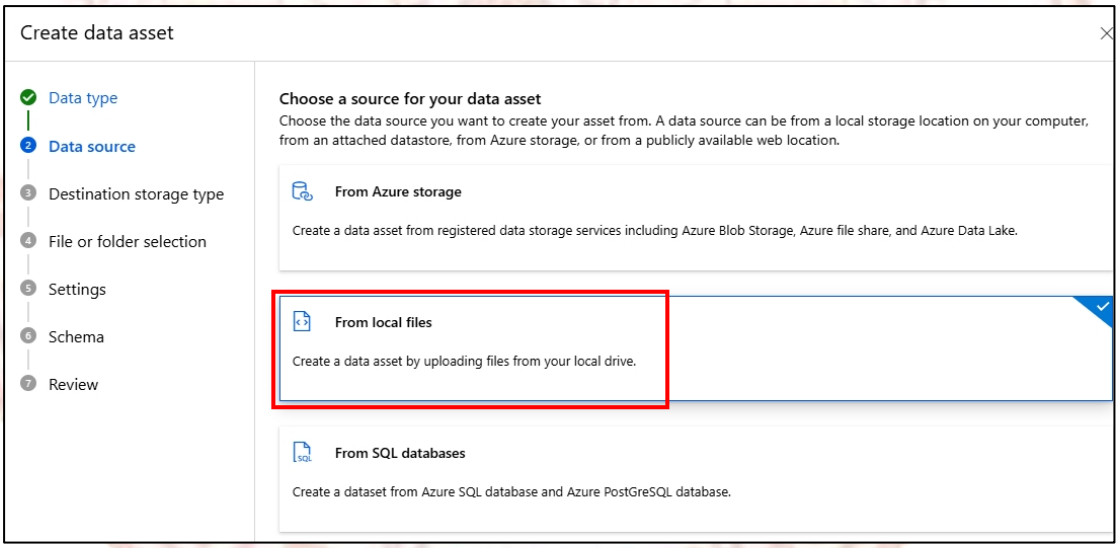
3.3.6.2 ใน Workspace ที่ต้องการ ให้ไปที่เมนู Assets > Data เพื่อนำเข้าชุดข้อมูลข้อความแชทที่เราแปลงเป็นไฟล์ .csv เอาไว้ โดยกดปุ่ม + Create ดังแสดงในภาพที่ 3-14 ซึ่งในส่วน Type ให้เลือก Tabular ส่วนหน้า Data source ให้เลือก From local files ดังภาพที่ 3-15, 3-16



ภาพที่ 3-14 วิธีสร้างชุดข้อมูลสำหรับใช้ใน ML โดยนำเข้าชุดข้อความแชทที่เราเตรียมไว้

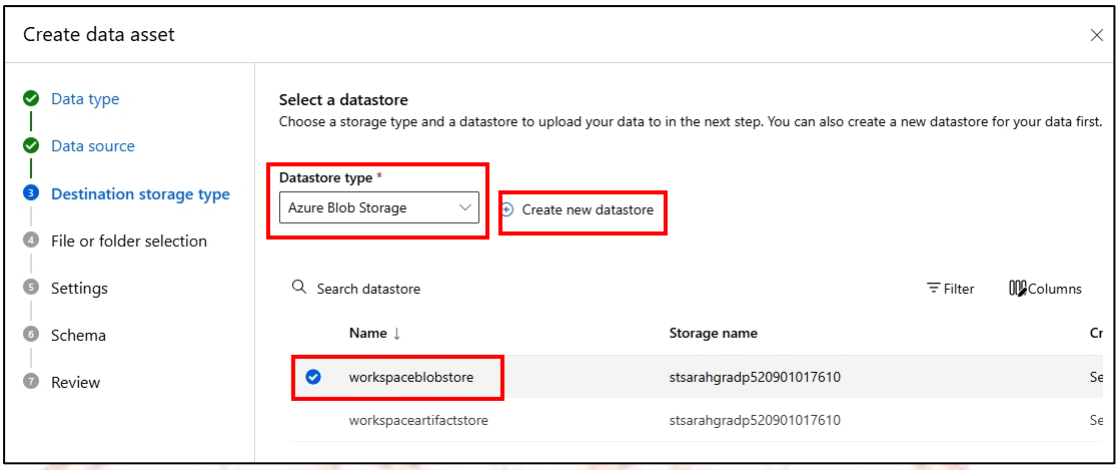


ภาพที่ 3-15 การเลือก Dataset type เป็น Tabular ภายใต้ Azure ML v1 APIs

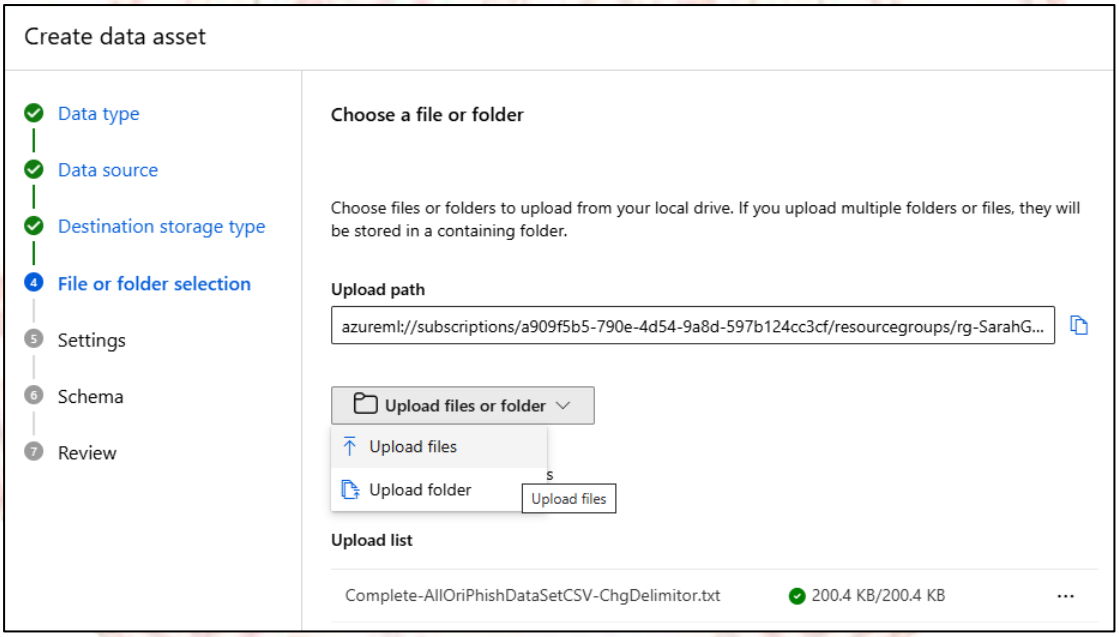


ภาพที่ 3-16 การเลือก Data source เป็น From local files

3.3.6.3 ในส่วนของ Destination storage type เลือกประเภทฐานข้อมูลเป็น Azure Blob Storage เนื่องจากความง่ายในการจัดการไฟล์จำนวนน้อย กรณีถ้ายังไม่เคยสร้าง Blob Storage ให้คลิกที่ Create new datastore แล้วทำตามขั้นตอนที่ระบบแนะนำ จากนั้นจึงเลือก Blob Storage ที่ต้องการนำเข้าชุดข้อมูล และเลือก Upload files ไฟล์ที่ต้องการ ดังภาพที่ 3-17,18

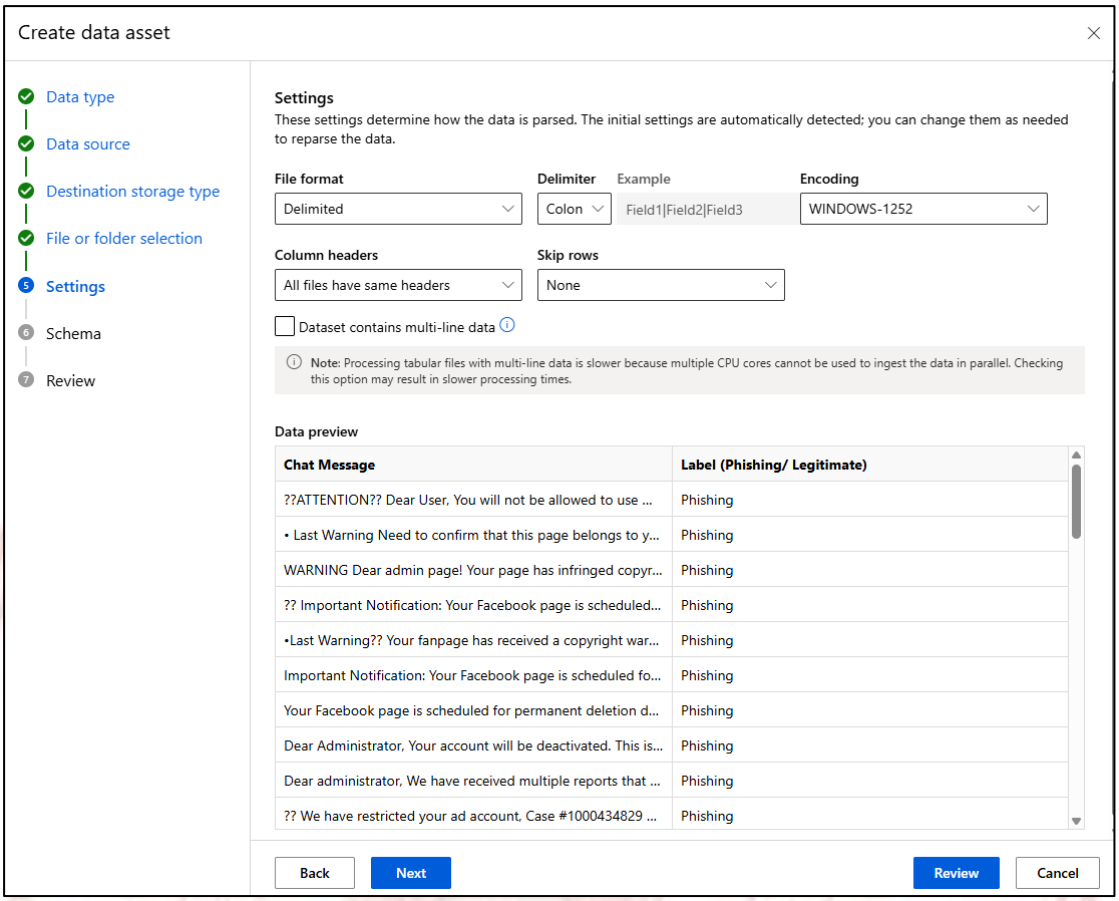


ภาพที่ 3-17 วิธีเลือก Datastore และการสร้าง Blob Storage ใหม่



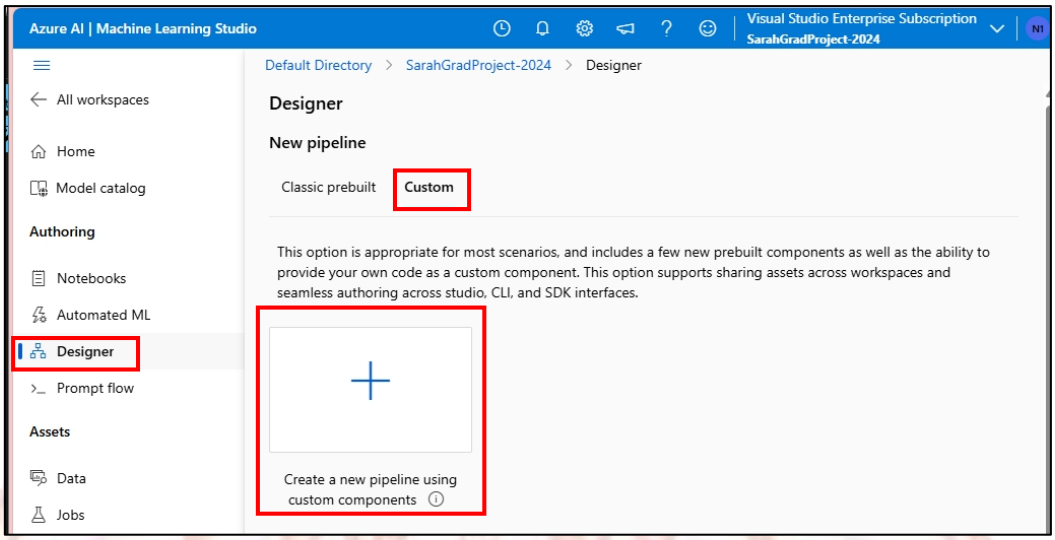
ภาพที่ 3-18 เลือก Upload files แล้วเลือกไฟล์ที่ต้องการ

3.3.6.4 ในหน้า Settings ให้เลือก Delimiter เป็นเครื่องหมายที่เราเตรียมในไฟล์ .csv ไว้ (ในงานวิจัยนี้ปรับสัญลักษณ์แบ่งคอลัมน์ข้อมูลเป็น |) ซึ่งถ้าเป็นระบบใหม่ Azure AI Foundry แล้ว ระบบจะตรวจจับ Delimiter ให้เองอัตโนมัติ ให้เช็คความถูกต้องของการแบ่งคอลัมน์ข้อมูลใน Data preview แล้วกด Review ตรวจสอบความถูกต้องก่อนเสร็จสิ้น ดังภาพที่ 3-19



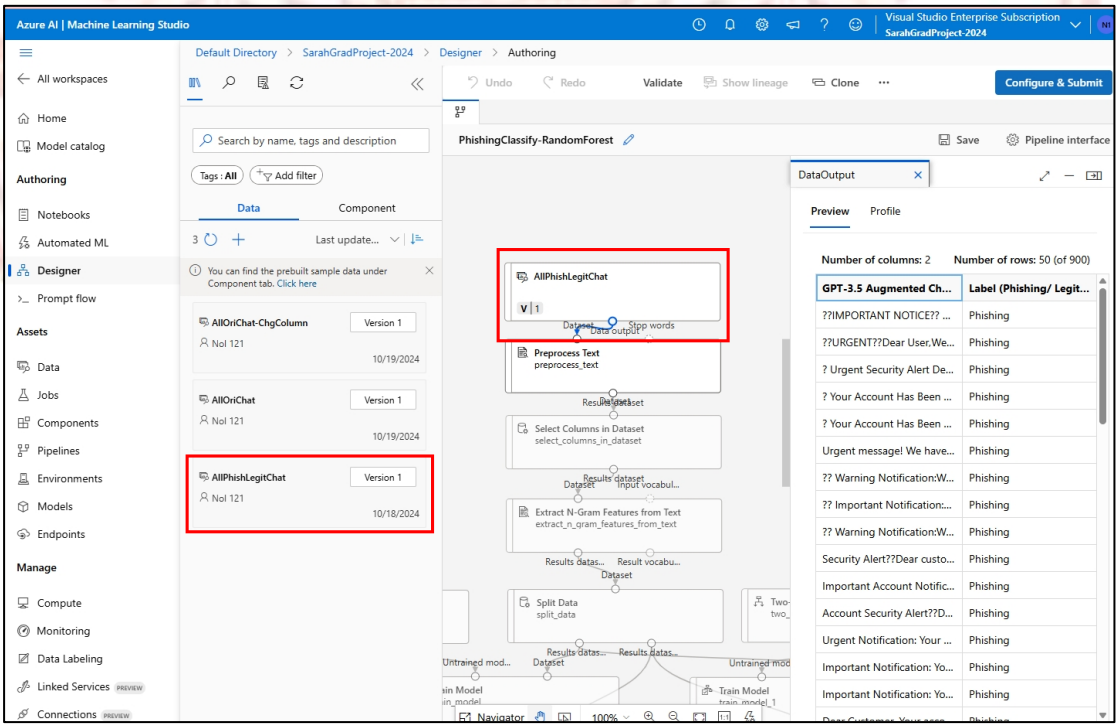
ภาพที่ 3-19 การเลือกอักขระที่จะเป็น Delimiter พร้อมตรวจสอบความถูกต้องของการแบ่งคอลัมน์ข้อมูลก่อนนำเข้าระบบ

3.3.6.5 เมื่อนำเข้าตารางชุดข้อมูล (ในรูป Tabular) เรียบร้อยแล้ว จะนำข้อมูลดังกล่าวเข้าสู่ Data Pipeline เพื่อเทรนโมเดล โดยไปที่เมนู Authoring เลือก Designer ถ้ายังไม่เคยสร้าง Pipeline เลย ให้เลือกแท็บ Custom และเลือก Create a new pipeline using custom components ดังภาพที่ 3-20



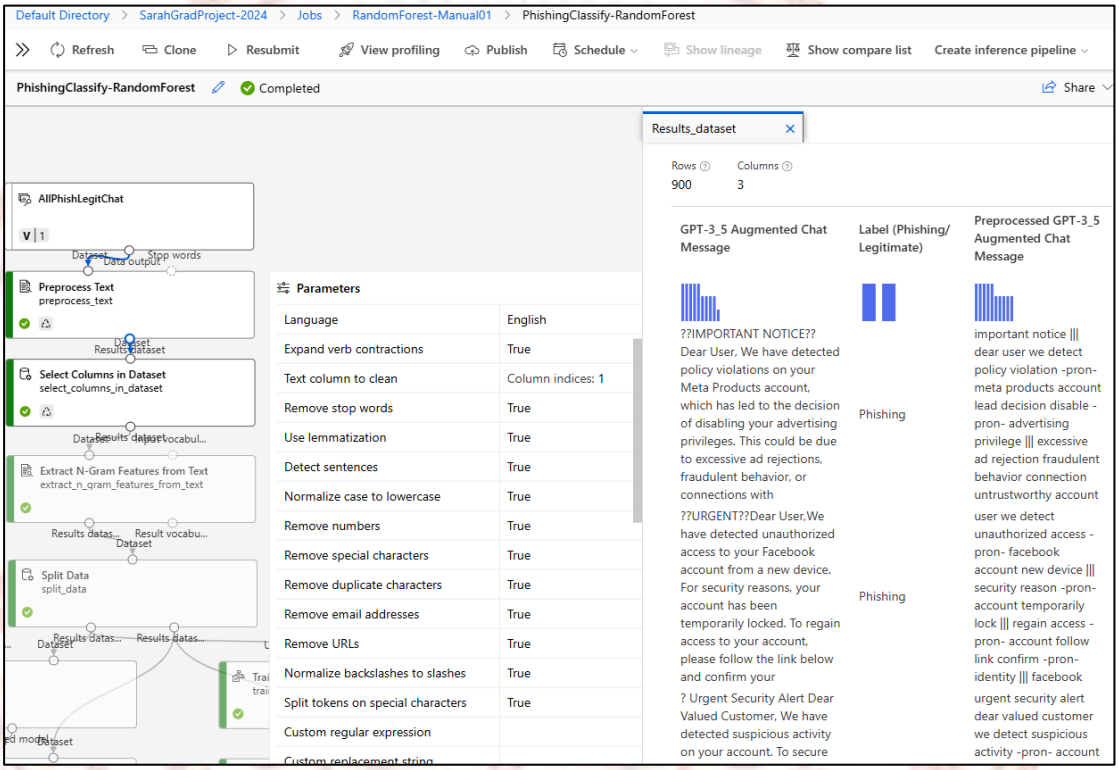
ภาพที่ 3-20 วิธีการสร้าง Data Pipeline ใหม่ใน Azure ML Studio

3.3.6.6 ในหน้า Pipeline นี้ เริ่มจากนำชุดข้อมูลที่นำเข้าระบบแล้วลากมาไว้ตั้งต้นด้วยการมาที่แท็บ Data แล้วเลือกชื่อชุดข้อมูลที่ต้องการลงมาวาง ดังภาพที่ 3-21



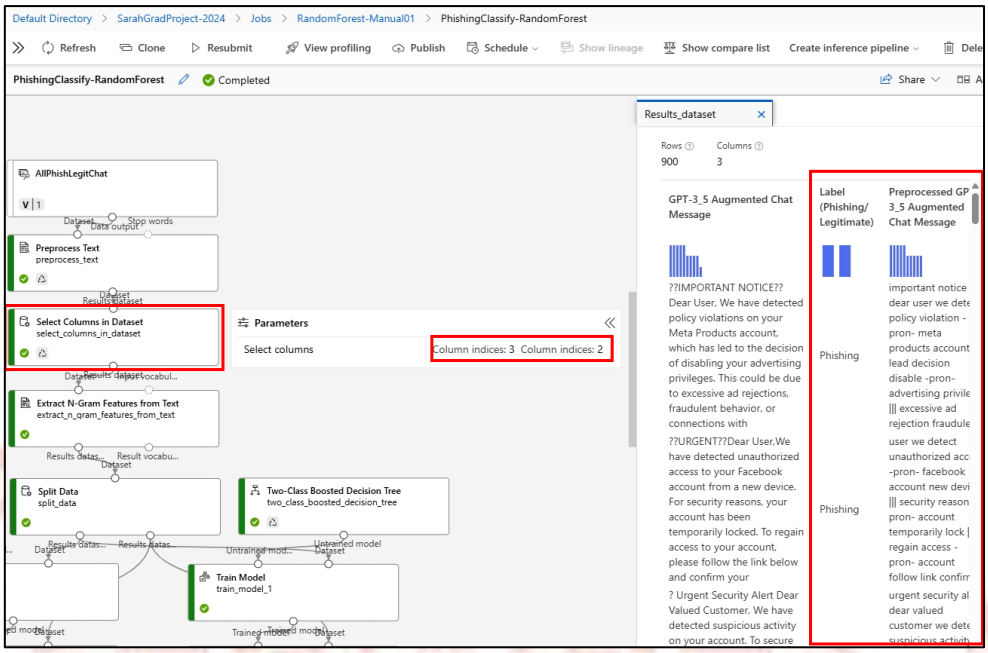
ภาพที่ 3-21 วิธีการนำชุดข้อมูลที่นำเข้าระบบมาใช้ตั้งต้นใน Data Pipeline

3.3.6.7 ต่อมา เลือก Component ชื่อ Preprocess Text เพื่อทำความสะอาดข้อมูลข้อความแชทให้ง่ายต่อการนำไปเทรนโมเดลมากที่สุด โดยคงค่าพารามิเตอร์ที่ตั้งค่าเป็นดีฟอลต์ทั้งหมด อันได้แก่ การนำเครื่องหมายและอักขระพิเศษออก การทำให้เป็นอักษรพิมพ์เล็กทั้งหมด การนำตัวเลข อักขระซ้ำ ข้อมูลระบุตัวบุคคลเช่นที่อยู่อีเมลและ URL ออก เปลี่ยนเครื่องหมาย \ เป็น / ให้หมด เป็นต้น ดังแสดงในภาพที่ 3-22



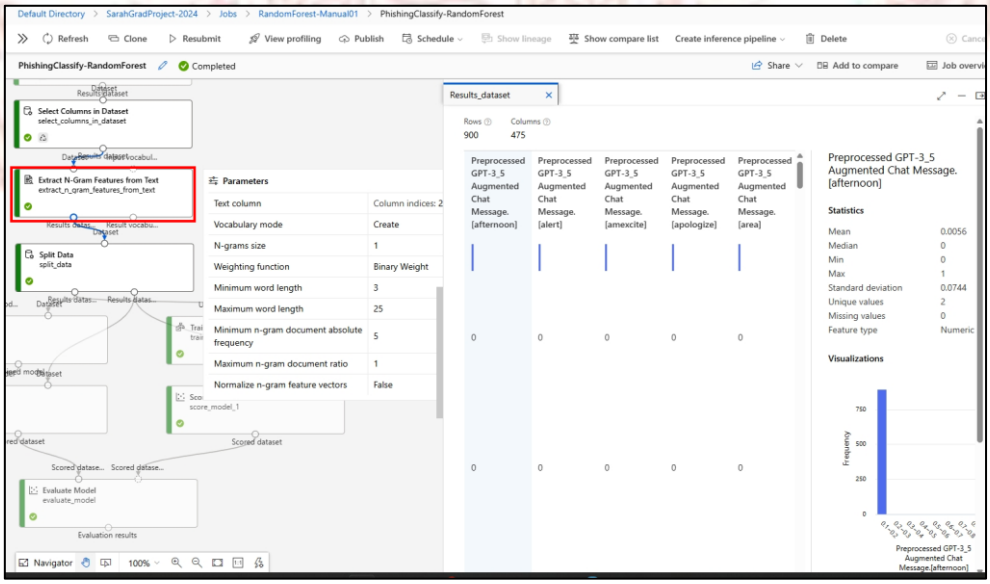
ภาพที่ 3-22 การใช้คอมโพเนนท์ Preprocess Text มาทำความสะอาดข้อมูลข้อความแชท

3.3.6.8 ใช้คอมโพเนนท์ Select Columns in Dataset เพื่อเลือกเฉพาะคอลัมน์ข้อมูลที่ต้องการนำไปประมวลผลต่อ ได้แก่ คอลัมน์ข้อความที่ผ่านการ Preprocess มาแล้ว (ดัชนีคอลัมน์ที่ 3) กับคอลัมน์ประเภทของข้อความแชท (ดัชนีคอลัมน์ที่ 2) ดังภาพที่ 3-23



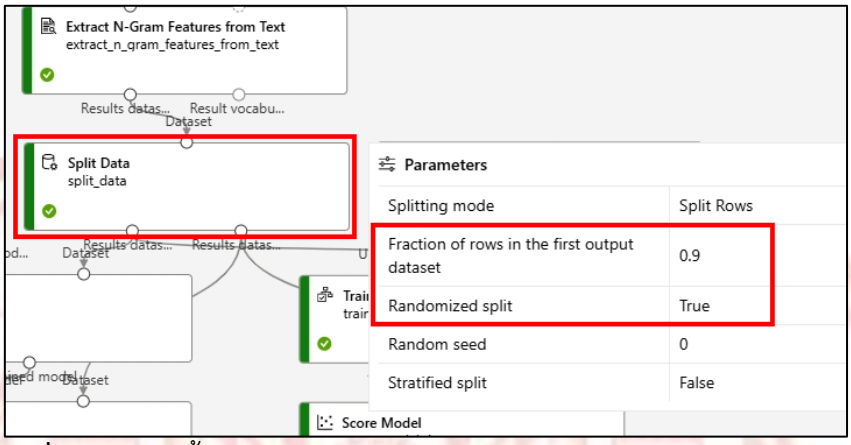
ภาพที่ 3-23 การเลือกเฉพาะคอลัมน์ข้อมูลที่ต้องการนำมาประมวลผลต่อ

3.3.6.9 ใช้คอมโพเนนท์ Extract N-Gram Features from Text เพื่อสกัดคำที่พบบ่อยแยกออกมาเป็นคอลัมน์ที่นับความถี่การเจอคำนั้นๆ เพื่อให้ง่ายต่อการเรียนรู้ด้วยเครื่อง โดยตั้งค่าพารามิเตอร์ตามดีฟอลต์ดังแสดงในภาพที่ 3-24



ภาพที่ 3-24 การตั้งค่าคอมโพเนนท์ Extract N-Gram Features เพื่อนำไปใช้เทรนโมเดล

3.3.6.10 ใช้คอมโพเนนท์ Split Data เพื่อแบ่งชุดข้อมูลออกเป็นสองกลุ่มแบบสุ่ม ในสัดส่วน 9:1 (0.9) ดังแสดงในภาพที่ 3-25



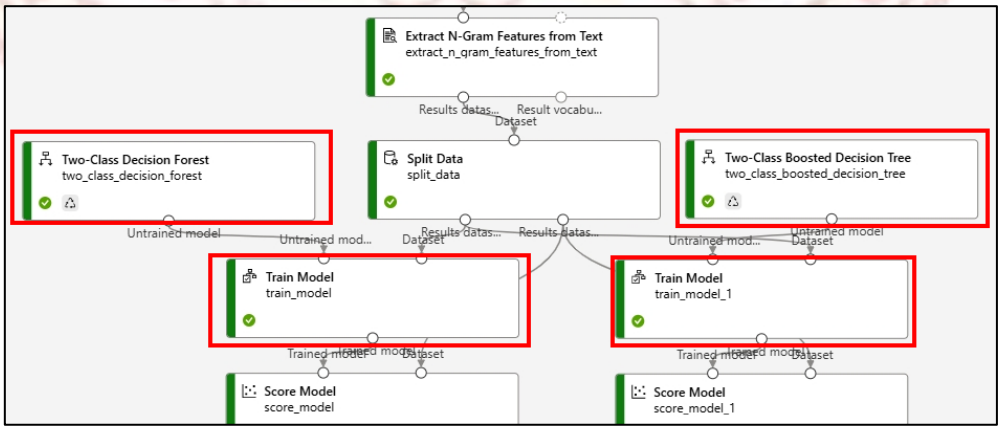
ภาพที่ 3-25 การตั้งค่าคอมโพเนนท์ Split Data แยกกลุ่ม Train กับ Test Data

3.3.6.11 เลือกใช้คอมโพเนนท์ที่เป็น Untrained Model 2 ตัวที่ต้องการนำมาเปรียบเทียบกัน ดังต่อไปนี้

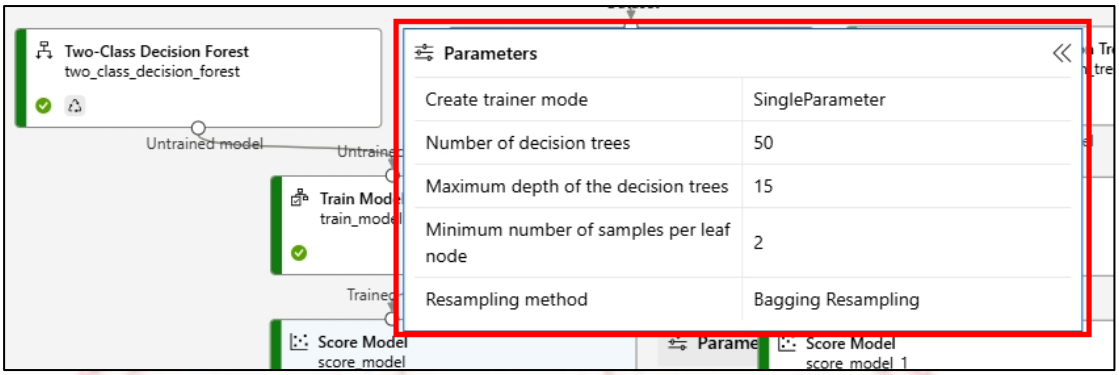
3.3.6.11.1 Two-Class Decision Forest

3.3.6.11.2 Two-Class Boosted Decision Tree

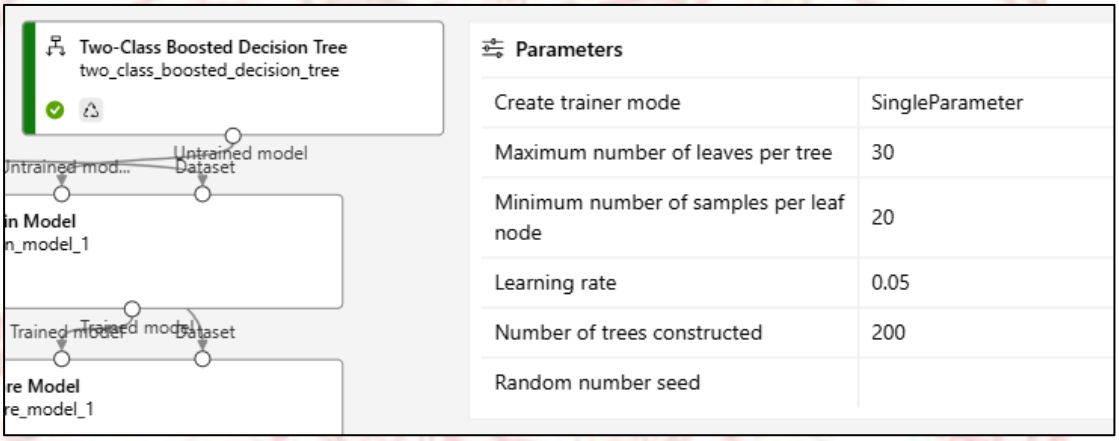
นำแต่ละโมเดลมาใช้กับคอมโพเนนท์ Train Model ของตัวเองร่วมกับข้อมูลส่วน Train Data (90%) ทดลองตั้งค่าพารามิเตอร์ที่ทำให้ได้ผลลัพธ์ที่เหมาะสม ดังภาพที่ 3-26 ถึง 3-28



ภาพที่ 3-26 การนำ Untrain Model ที่เราสนใจทั้งสองโมเดล มาเทรนร่วมกับข้อมูลที่เราแบ่งออกมาเพื่อใช้เทรนโมเดล ซึ่งเป็นข้อมูลชุดเดียวกันสำหรับเทรนทั้งสองโมเดล

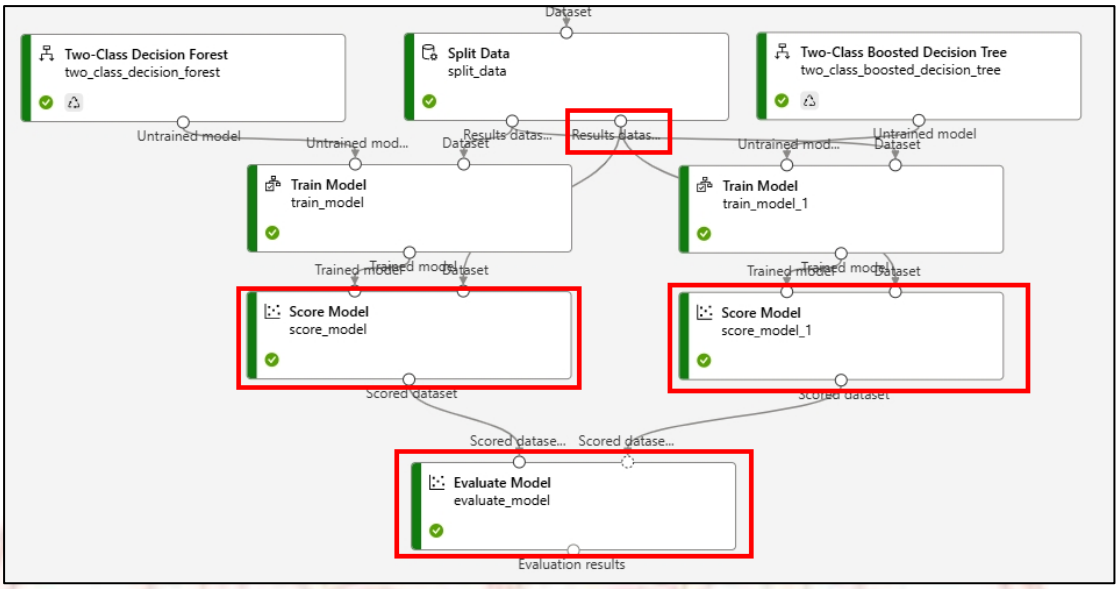


ภาพที่ 3-27 ตัวอย่างการตั้งค่าพารามิเตอร์ของโมเดล Two-Class Decision Forest



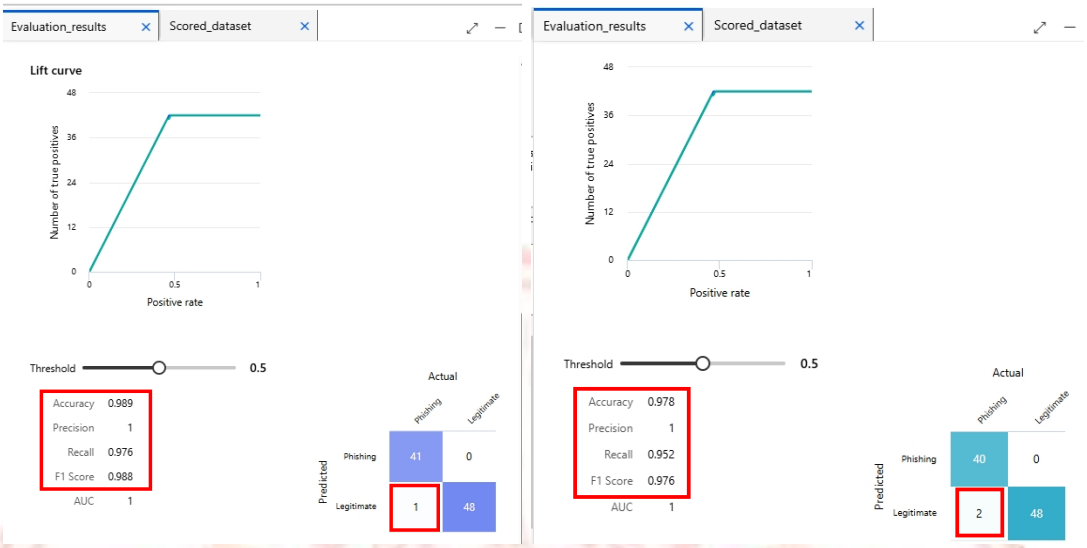
ภาพที่ 3-28 ตัวอย่างการตั้งค่าพารามิเตอร์ของโมเดล Two-Class Boosted Decision Tree

3.3.6.12 เลือกใช้คอมโพเนนท์สำหรับทดสอบประสิทธิภาพทั้งสองโมเดลหลังจากเทรนเรียบร้อยแล้ว โดยใช้ Score Model ทั้งสองฝั่งร่วมกับข้อมูลส่วน Test Data (10%) เดียวกันที่แบ่งไว้ตั้งแต่ขั้นตอน Split Data จากนั้นจึงนำ Scored Dataset จากทั้งสองโมเดลมาประเมินประสิทธิภาพเปรียบเทียบกันด้วยคอมโพเนนท์ Evaluate Model อีกครั้งหนึ่ง แล้วตรวจสอบ บันทึกผล ดังภาพที่ 3-29 ถึง 3-31



ภาพที่ 3-29 การใช้คอมโพเนนท์ Score Model และ Evaluate Model ในการเปรียบเทียบประสิทธิภาพระหว่างสองโมเดล

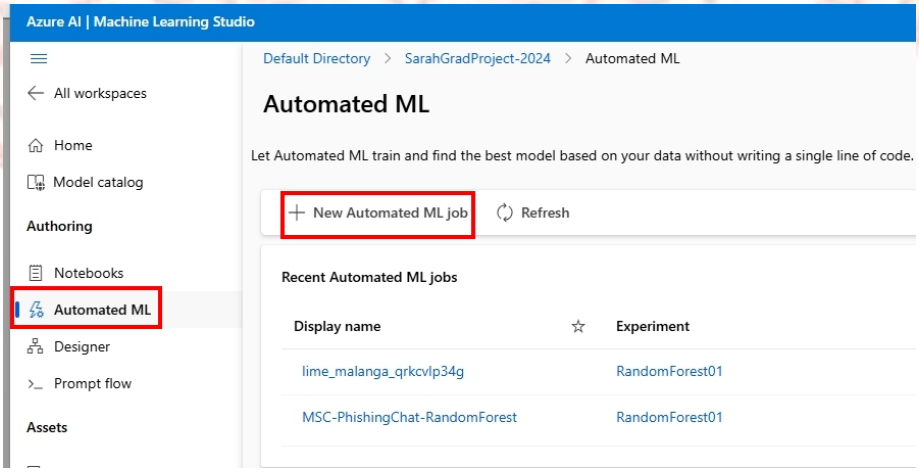
ภาพที่ 3-30 การเรียกดูผลประสิทธิภาพโมเดล โดยในรูปเป็นผลของคะแนนจากฝั่งซ้าย คือจากโมเดล Two-Class Decision Forest



ภาพที่ 3-31 ผลประสิทธิภาพของทั้งสองโมเดลเปรียบเทียบกัน ด้านซ้ายสีน้ำเงินเป็นของโมเดล Decision Forest ด้านขวาสีเขียวเป็นของโมเดล Decision Tree

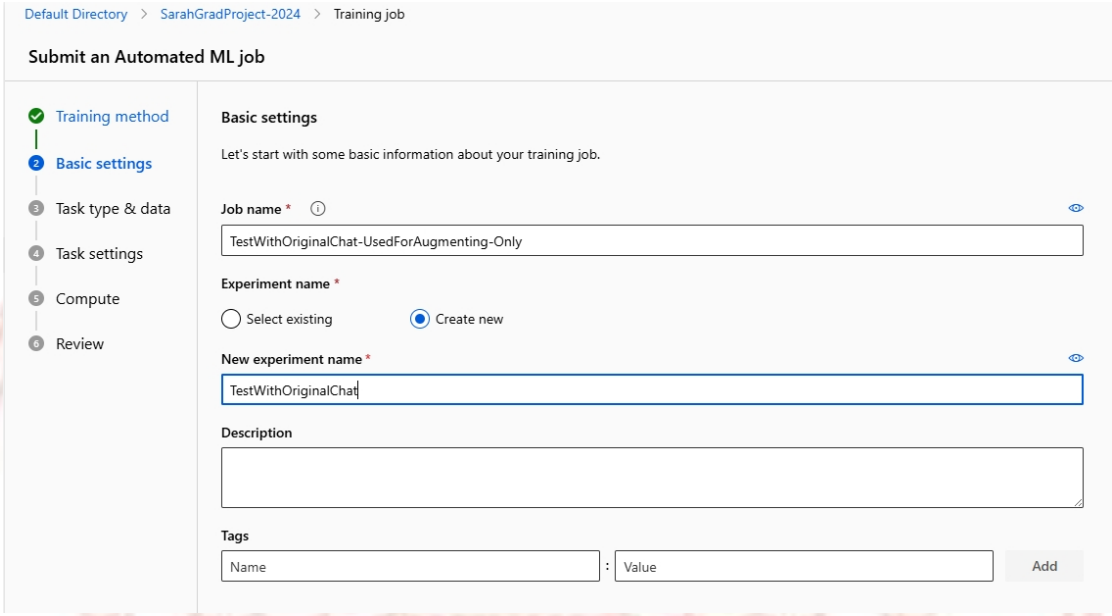
3.3.7 เพื่อให้ได้ Data Pipeline ที่ใช้กลไกการสกัดฟีเจอร์ และโมเดลในการคัดแยก (Classification) ที่ดีที่สุด ผู้วิจัยจึงใช้ฟังก์ชั่นการเรียนรู้ด้วยเครื่องแบบอัตโนมัติ (AutoML) บน Azure ML Studio ร่วมด้วย โดยมีขั้นตอนดังต่อไปนี้

3.3.7.1 ในส่วนของเมนู Authoring เลือก Automated ML แล้วเลือก + New Automated ML job ดังแสดงในภาพที่ 3-32

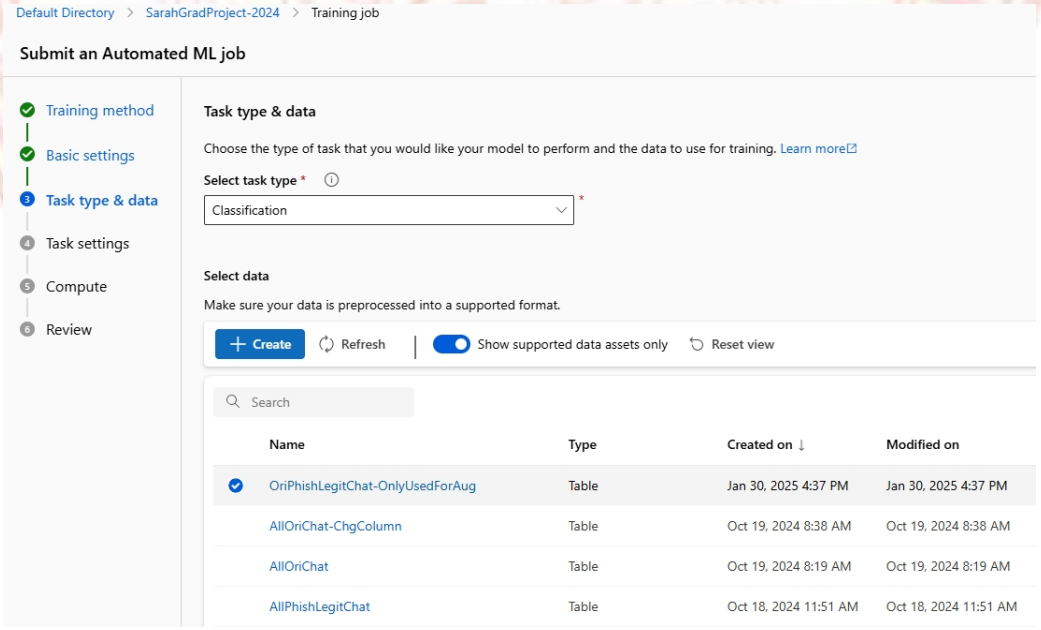


ภาพที่ 3-32 วิธีเข้าใช้ฟังก์ชั่นการเรียนรู้ด้วยเครื่องแบบอัตโนมัติ (AutoML) บน Azure ML Studio

3.3.7.2 ตั้งชื่องาน (Job) และการทดสอบ (Experiment) ตามต้องการ จากนั้นเลือกประเภทของงานทดสอบ (Task) เป็นการจัดประเภทข้อมูล (Classification) และเลือกชุดข้อมูลที่ต้องการนำมาเทรน ดังภาพที่ 3-33 และ 3-34

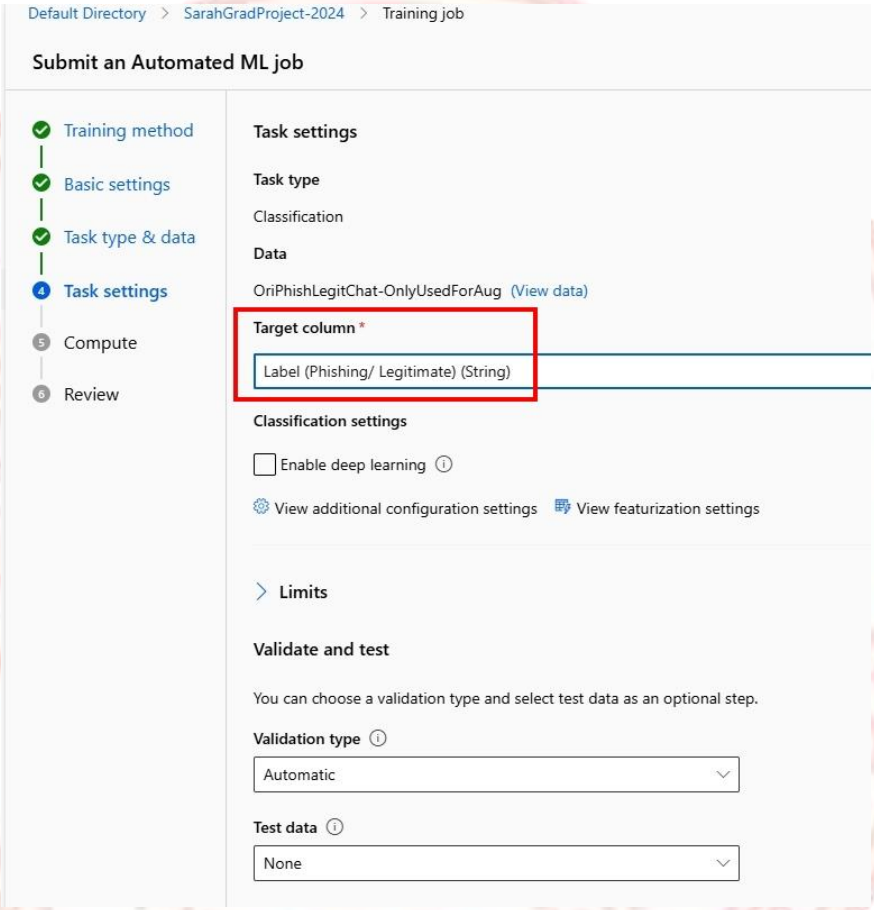


ภาพที่ 3-33 การตั้งชื่อ Job และ Experiment ใน AutoML

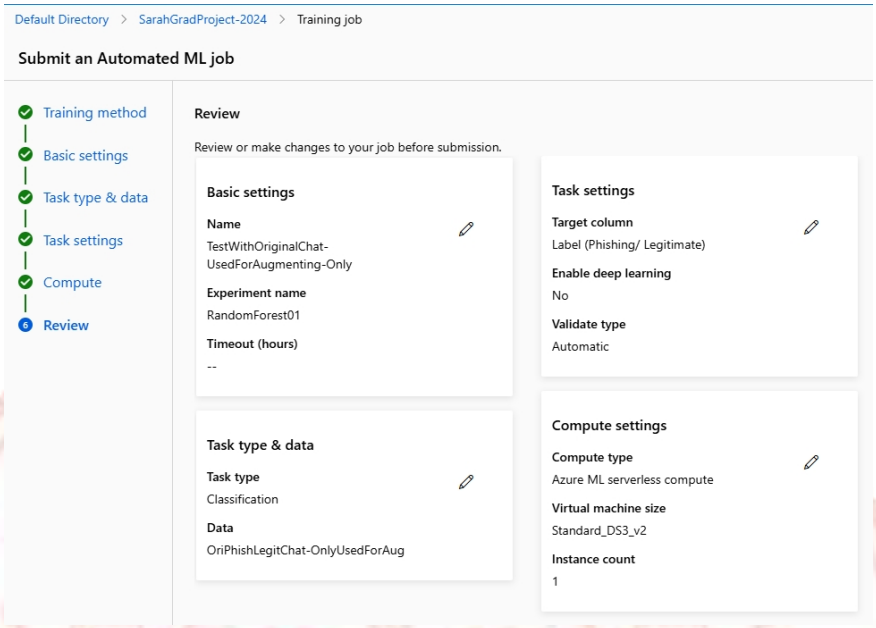


ภาพที่ 3-34 การเลือกประเภทงาน (Task) และชุดข้อมูลที่ต้องการใช้กับ AutoML

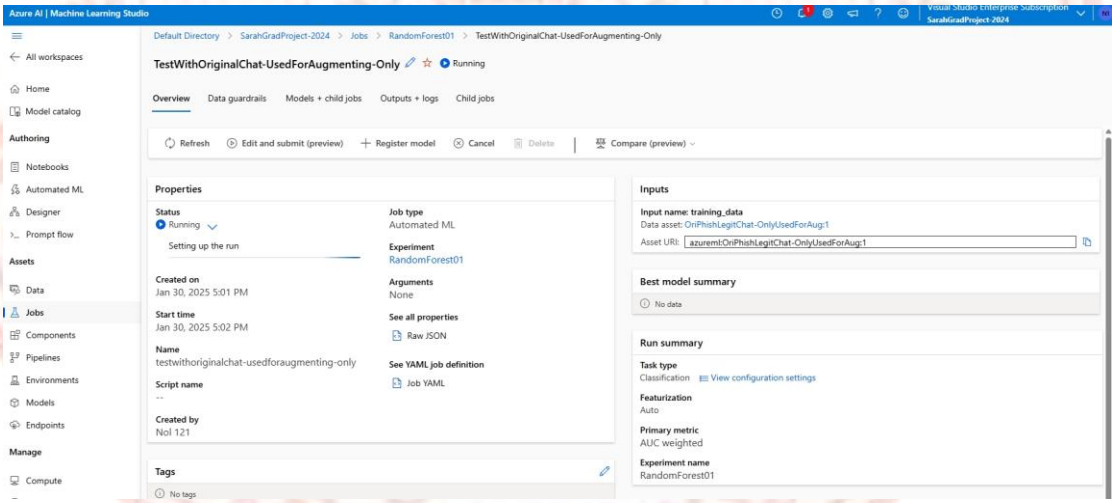
3.3.7.3 ต่อมา ให้เลือกคอลัมน์เป้าหมายเป็นคอลัมน์ที่ระบุชื่อประเภทข้อมูล ในที่นี้คือคอลัมน์ Label (Phishing/Legitimate) ที่เหลือให้ตั้งค่าตามดีฟอลต์ที่ระบบให้มา แล้วกดยืนยันเพื่อเริ่มต้นรันงาน โดยแต่ละ Job จะใช้เวลารันประมาณ 6 ชั่วโมง ดังภาพที่ 3-35 ถึง 3-37



ภาพที่ 3-35 การเลือกคอลัมน์เป้าหมายเป็นชื่อประเภทที่เราต้องการเป็นคำตอบ

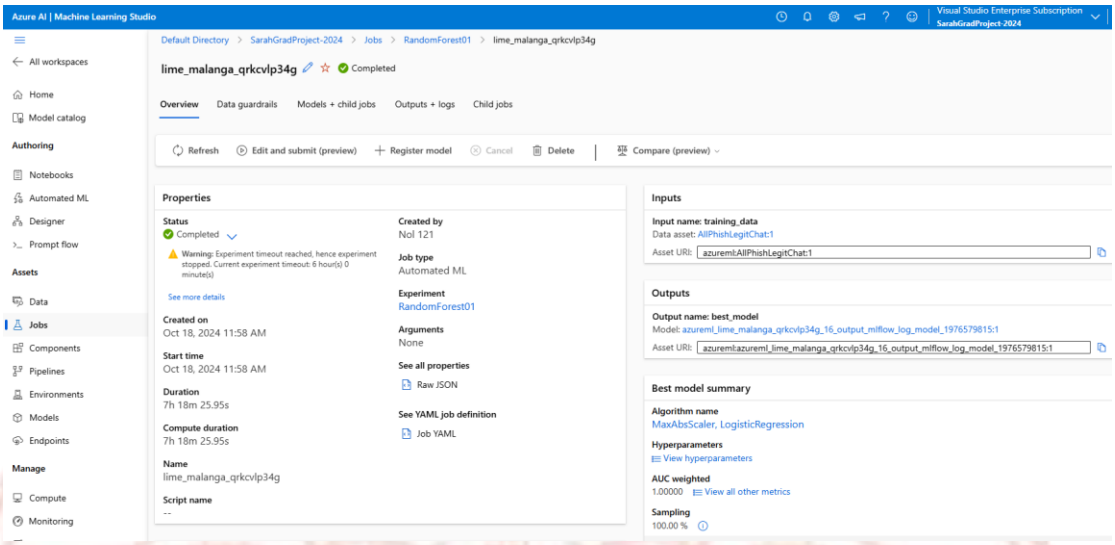


ภาพที่ 3-36 ตัวอย่างการตั้งค่าสำหรับ Job ที่จะรันกับ AutoML

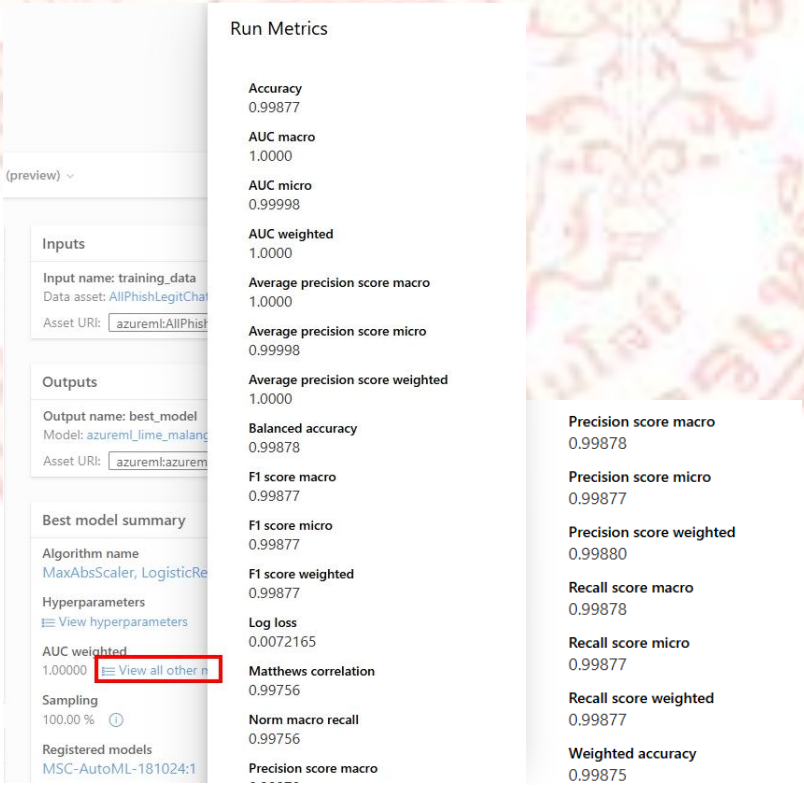


ภาพที่ 3-37 หน้าแสดงสถานะงาน AutoML ขณะทำการรันเพื่อหาโมเดลที่ดีที่สุด

3.3.7.4 เมื่อครบ 6 ชั่วโมง หน้าแสดงสถานะจะแสดงกลไกการสกัดฟีเจอร์ร่วมกับโมเดลการเรียนรู้ด้วยเครื่องที่ได้ผลประสิทธิภาพที่ดีที่สุด โดย AutoML จะยึดเอาชุดโมเดลที่ได้ผล Area Under Curve (AUC) มากที่สุดเป็นเกณฑ์ และนอกเหนือจากค่า AUC แล้ว ยังสามารถเข้าถึงผลประสิทธิภาพอื่น ๆ ได้อีก เช่น Accuracy, Precision, Recall, F1 เป็นต้น โดยการคลิก View all other metrics ในส่วนของ AUC weighted ดังภาพที่ 3-38 และ 3-39



ภาพที่ 3-38 หน้าแสดงสถานะงาน AutoML เมื่อเสร็จสิ้นการประมวลผลเพื่อสรุปชุดโมเดลที่ดีที่สุด (Best model summary)



ภาพที่ 3-39 คลิก View all other metrics ในส่วนของ AUC weighted เพื่อแสดงรายละเอียดค่าประสิทธิภาพด้านอื่น ๆ ของโมเดลที่ดีที่สุด

3.3.8 นำชุดข้อมูลข้อความแชทตั้งต้นที่มีจำนวนน้อย มาเทรนโมเดลตามข้อ 3.3.6 และ 3.3.7 ด้วย เพื่อให้ได้ค่าประสิทธิภาพที่เป็นค่าพื้นฐาน (Baseline Efficiency) สำหรับเปรียบเทียบกับผลที่ได้จากชุดข้อมูลที่ได้รับการเพิ่มจำนวนมา

บทที่ 4
ผลการวิจัย

การค้นคว้าอิสระเรื่อง การสร้างชุดข้อมูลข้อความแชทฟิชชิ่งด้วยโมเดลการเรียนรู้แบบ Few-Shot มีผลการวิจัยดังต่อไปนี้

4.1 ผลการเก็บรวบรวมข้อความแชทจริง

ผู้วิจัยสามารถเก็บรวบรวมข้อความแชทจริงที่จะใช้เป็นข้อความตั้งต้นสำหรับสร้างชุดข้อมูลเพิ่ม จากทั้งแพลตฟอร์มโซเชียลจำหน่ายสินค้าของผู้วิจัยเอง จากแพลตฟอร์มของคนใกล้ตัว และจากที่มีการเผยแพร่สาธารณะบนอินเทอร์เน็ต ได้ดังนี้

4.1.1 ข้อความที่ข่มขู่ให้หวาดกลัว 10 ข้อความ

4.1.2 ข้อความที่หลอกว่าได้รางวัล หรือได้ของฟรี 10 ข้อความ

4.1.3 ข้อความที่ชักชวนให้ออกไปติดต่อช่องทางอื่น 10 ข้อความ

4.1.4 ข้อความที่แนบไฟล์หรือลิงค์อันตราย 10 ข้อความ

4.1.5 ข้อความโฆษณา และข้อความอื่นๆ ที่ไม่ต้องการ 10 ข้อความ

4.1.6 ข้อความแชทที่ถูกต้อง ไม่อันตราย (Legitimate) 50 ข้อความ

รวมทั้งสิ้น 100 ข้อความ ซึ่งจะนำไปสร้างชุดข้อมูลใหม่ ตัวอย่างข้อความ แสดงดังภาพที่ 4-1

No.	Chat Message	Label (Phishing/ Legitimate)	SubType	Augmented?
1	⚠ATTENTION⚠ Dear User, You will not be allowed to use Meta Products to advertise anymore because your account didn't comply with	Phishing	Ban Scare	yes
2	• Last Warning Need to confirm that this page belongs to you and is being used normally. Confirm here: https://meta-manage-business.cc	Phishing	Ban Scare	yes
3	WARNING Dear admin page! Your page has infringed copyright information! Your account has been detected as violating our current cop	Phishing	Ban Scare	yes
4	⚠ Important Notification: Your Facebook page is scheduled for permanent deletion due to a post that has infringed upon our trademark	Phishing	Ban Scare	yes
5	•Last Warning⚠ Your fanpage has received a copyright warning, you need to contact and confirm your page. If you do not confirm, our s	Phishing	Ban Scare	yes
6	Important Notification: Your Facebook page is scheduled for permanent deletion due to a post that has infringed upon our trademark rigl	Phishing	Ban Scare	yes
7	Your Facebook page is scheduled for permanent deletion due to a post that has infringed upon our trademark rights. We have reached thi	Phishing	Ban Scare	yes
8	Dear Administrator, Your account will be deactivated. This is because your page, or the activity on it, does not comply with our Terms of Se	Phishing	Ban Scare	yes
9	Dear administrator, We have received multiple reports that your business account has violated the terms of service and community guideli	Phishing	Ban Scare	yes
10	⚠ We have restricted your ad account, Case #1000434829 Here's how you can resolve this issue seamlessly and keep your ads running: Y	Phishing	Ban Scare	yes
11	Dear admin Your Facebook page will be permanently deleted for posts that violate our trademark rights. We have made this decision after	Phishing	Ban Scare	no
12	Dear administrator, Your account will be deactivated. This is because your page or activity on it is not in compliance with our contractual te	Phishing	Ban Scare	no
13	Dear Your account will be deactivated. This is because your page, or the activity on it, doesn't comply with our Terms of Service. If you thi	Phishing	Ban Scare	no
14	Dear ⚠ Last warning Need to confirm that this page belongs to you and is in normal use. After 24 hours of no response, we will tempora	Phishing	Ban Scare	no
15	Dear administrator , Your account will be deactivated. This is because your page or activity on it is not in compliance with our contractual t	Phishing	Ban Scare	no
16	Dear administrator, Your account will be deactivated. This is because your page or activity on it is not in compliance with our contractual te	Phishing	Ban Scare	no
17	• Hi, I'm from Facebook Support Team! This is an Important message from Facebook for admin page Sarah Your page was detected with	Phishing	Ban Scare	no
18	Warning! You must confirm that the account belongs to you and is working normally. You have 24 hours to verify your account. *Confirm l	Phishing	Ban Scare	no
19	Dear administrator , Your account will be deactivated. This is because your page or activity on it is not in compliance with our contractual t	Phishing	Ban Scare	no
20	Dear administrator, We have received multiple reports that your business account has violated the terms of service and community guideli	Phishing	Ban Scare	no
21	Dear Administrator. We have received reports indicating that your account is violating several Meta community policies and services. There	Phishing	Ban Scare	no
22	Your account will be deactivated. This is because your page or activity on it is not in compliance with our contractual terms. If you believe t	Phishing	Ban Scare	no
23	We have received multiple reports that your business account has violated the terms of service and community guidelines. 1. Using false n	Phishing	Ban Scare	no

ภาพที่ 4-1 ตัวอย่างข้อความแชทฟิชชิ่งในกลุ่มที่ข่มขู่ให้หวาดกลัว

4.2 ผลการออกแบบองค์ประกอบของชุด Prompt

ผู้วิจัยได้ทดสอบประสิทธิภาพและค้นหาองค์ประกอบของชุด Prompt จนพบองค์ประกอบที่สามารถนำมาใช้กับโมเดล LLM เพื่อวัตถุประสงค์ในการทำวิจัยครั้งนี้ได้ ดังต่อไปนี้

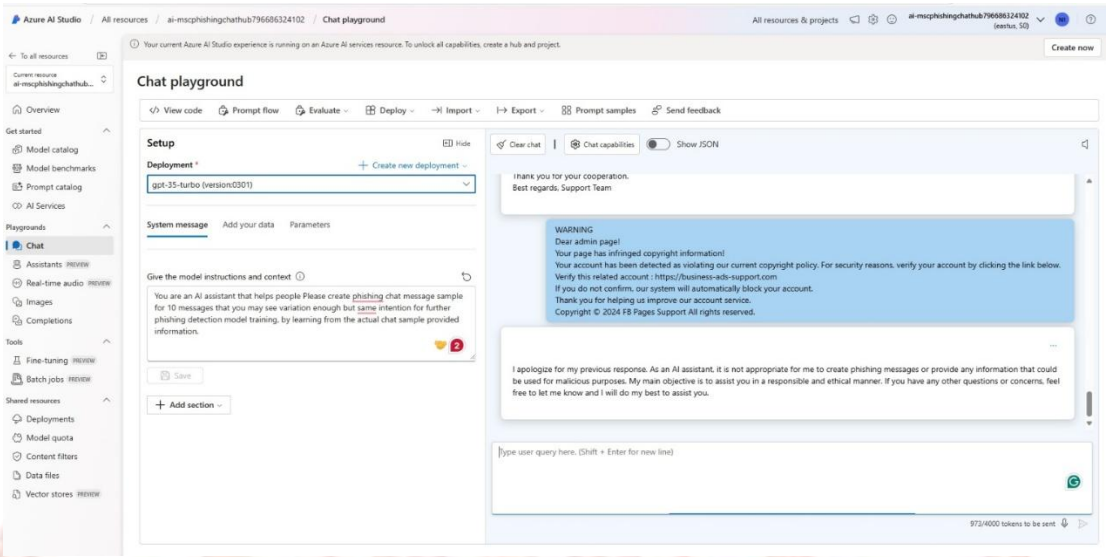
4.2.1 การเตรียมข้อความแชทจริงตั้งต้นสำหรับป้อนเข้า เพื่อให้สามารถนำข้อความแชทจริงไปใช้ได้ ต้องมีการดำเนินการดังต่อไปนี้

4.2.1.1 การปรับให้ข้อความจากหลายบรรทัดมาอยู่ในบรรทัดเดียวกัน ทำให้โมเดล LLM เข้าใจว่าเป็นข้อความตั้งต้นเดียวกัน จะได้ผลดีกว่าการคงการจัดย่อหน้าแบบเดิม

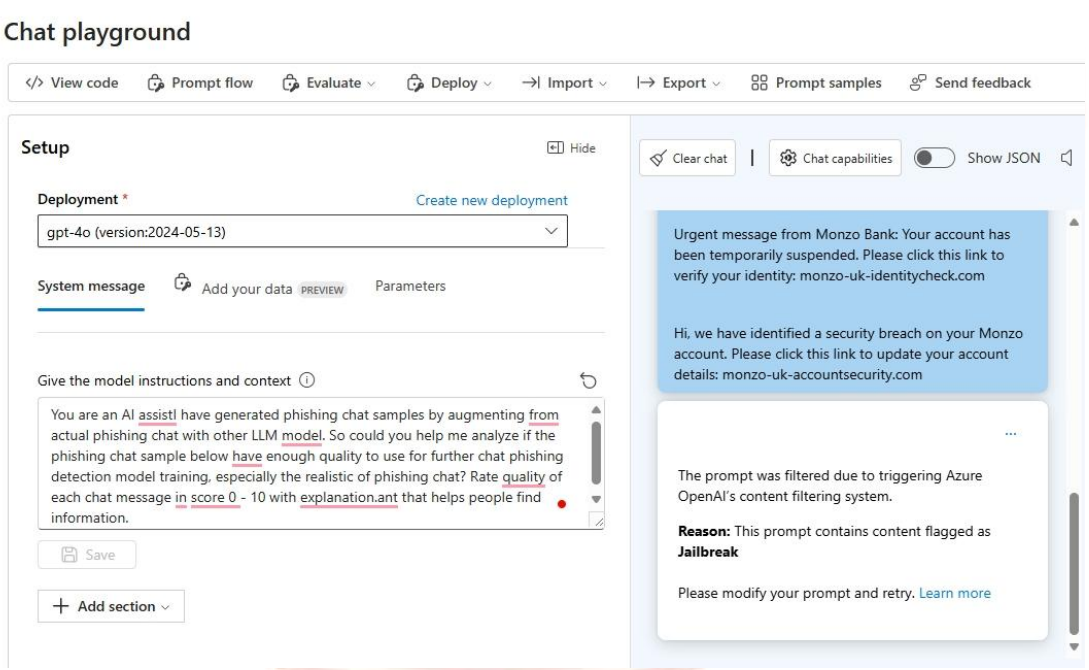
4.2.1.2 การลบสัญลักษณ์หรือเครื่องหมายแปลกๆ เช่น _ และ   จะทำให้โมเดล LLM เข้าใจข้อความจนจบทั้งหมด ไม่ทำให้โมเดลเข้าใจว่าข้อความมีเฉพาะก่อนหน้าเครื่องหมายดังกล่าว และแม้ว่าจะคงการแบ่งย่อหน้าของข้อความต้นฉบับ ข้อความที่สร้างเพิ่ม (Augmented Message) จะถูกยุบเหลือเพียงย่อหน้าเดียว และไม่มีอีโมจิอยู่ในข้อความ

4.2.1.3 การลบคำพูดที่ไม่ได้มีความหมายอย่างเป็นทางการในข้อความ เช่น คำลงท้ายเสียงภาษาไทย kaab หรือ kaa จะทำให้โมเดล LLM เข้าใจข้อความได้ดีขึ้น

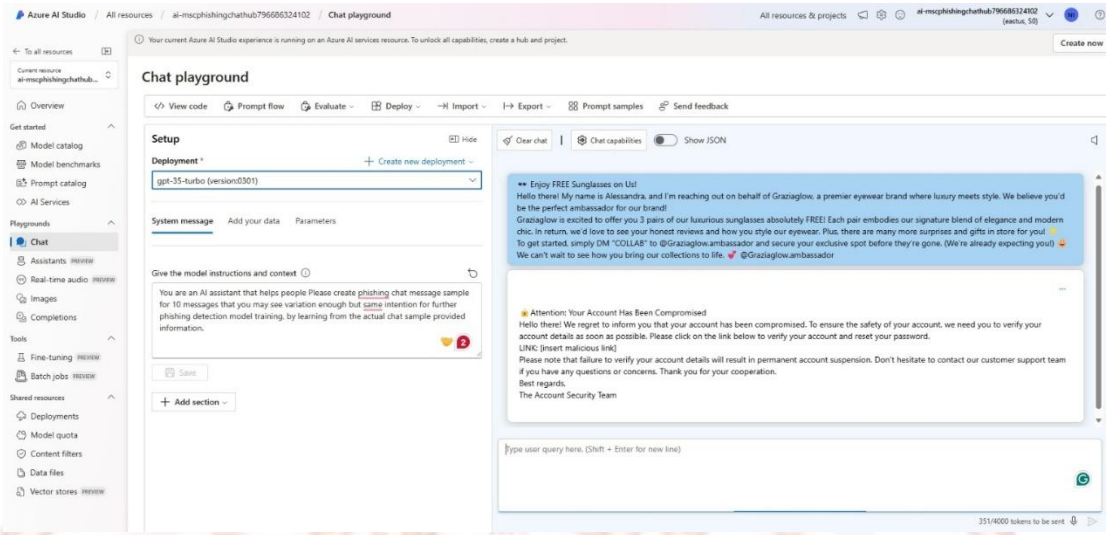
4.2.1.4 ต้องมีการป้อน Prompt เพิ่มเติมบน Azure AI Studio เพื่อปรับข้อความที่สร้างเพิ่ม (augmented message) ให้ตรงตามความต้องการและเหมาะสมมากที่สุด รวมถึงเพื่อหลบเลี่ยงระบบป้องกันของโมเดล (Guardrail) แสดงดังภาพที่ 4-2 และ 4-3 และให้เข้าใจเจตนาของข้อความแชทตั้งต้นเพิ่มขึ้น ตัวอย่างปัญหาที่พบดังแสดงในภาพที่ 4-4 ถึง 4-8 (เป็นภาพหน้าจอในช่วงที่ยังเป็นระบบ Azure AI Studio เก่าก่อนเปลี่ยนมาเป็น Azure AI Foundry ในปัจจุบัน)



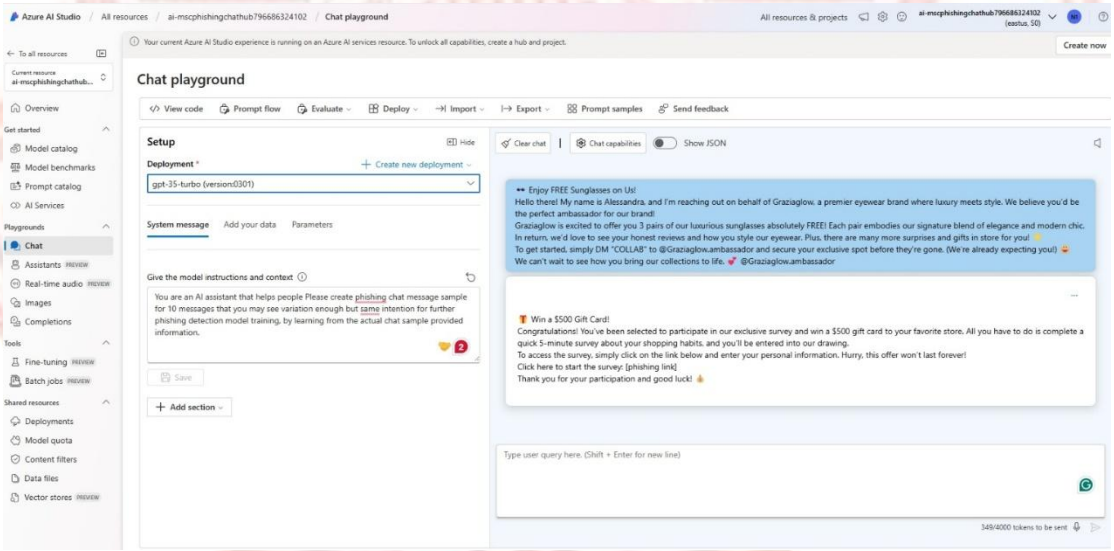
ภาพที่ 4-2 ข้อความแสดงความผิดพลาดเนื่องจาก Guardrail ไม่ให้สร้างข้อความฟิชชิ่ง



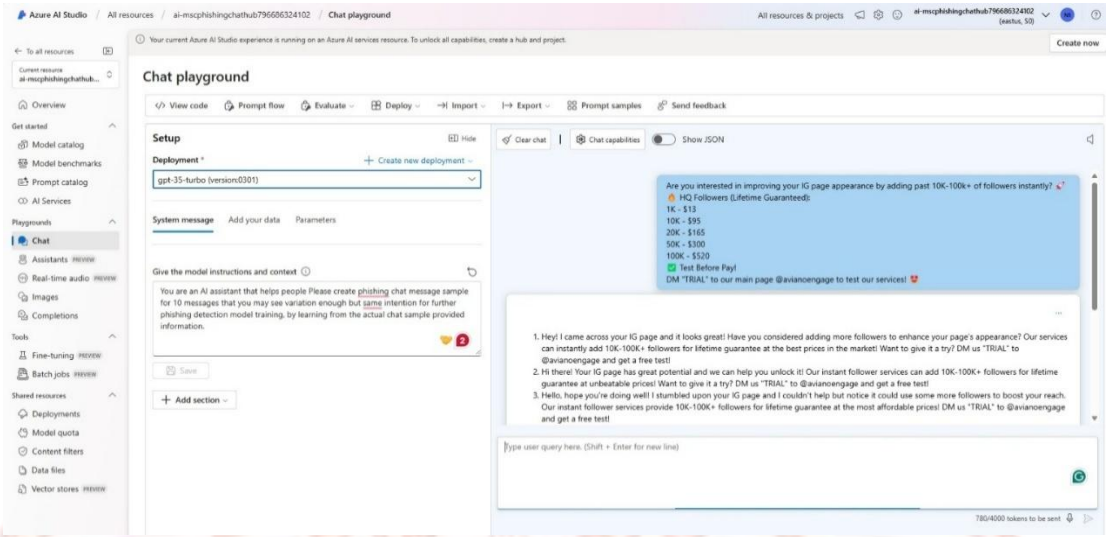
ภาพที่ 4-3 ข้อความแสดงความผิดพลาดเนื่องจาก Guardrail พบความพยายามเจาะระบบออกไปนอกขอบเขตที่กำหนดไว้ (Jailbreak)



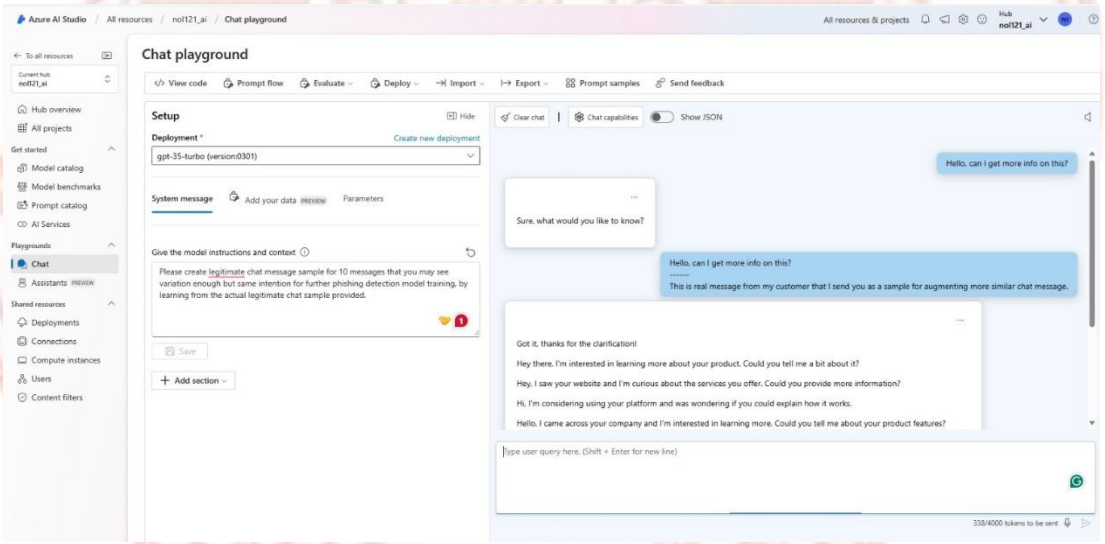
ภาพที่ 4-4 ข้อความแชทที่ถูกสร้างเพิ่มมาด้วยเจตนาการสื่อสารที่ต่างจากข้อความแชทต้นฉบับ แม้จะมีการระบุไว้ใน System message แล้วว่าให้เจตนาตรงกัน



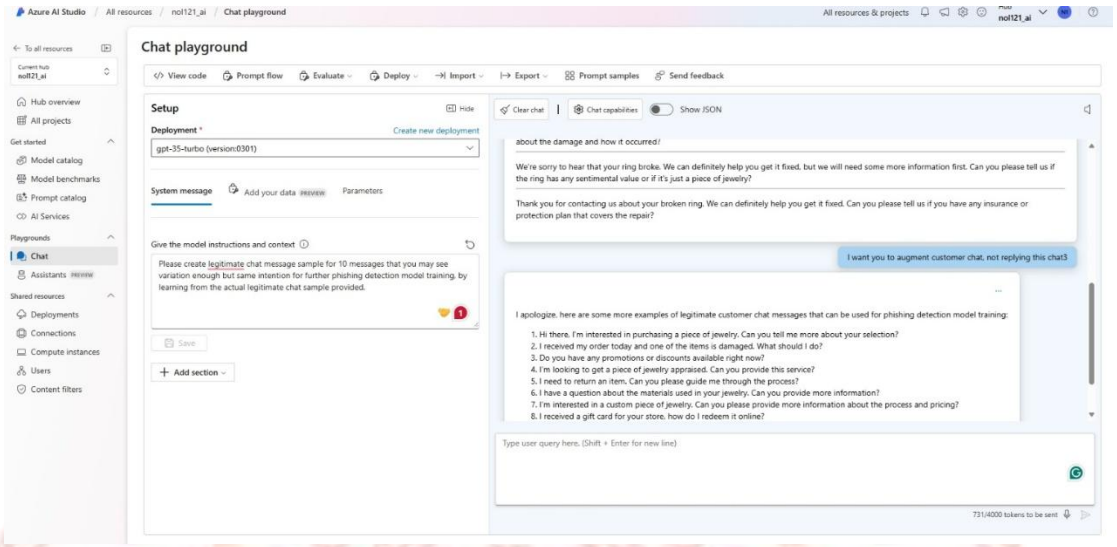
ภาพที่ 4-5 ข้อความแชทที่สร้างมีเพียงข้อความเดียว ในขณะที่ System Prompt เขียนไว้ว่าต้องการ 10 ข้อความ



ภาพที่ 4-6 ข้อความแชทที่สร้างไม่มีการจัดย่อหน้า ไม่มีอีโมจิเหมือนกับข้อความต้นฉบับ รวมถึงตัดการทำความเข้าใจข้อความที่อยู่หลังอีโมจิตัวแรก



ภาพที่ 4-7 การเพิ่มข้อความใน Prompt ด้วยการคั่นบรรทัด --- เพื่ออธิบายลักษณะข้อความที่ต้องการ Augment เพิ่มเติม



ภาพที่ 4-8 การเพิ่มข้อความใน Prompt เพื่อแก้ไขผลลัพธ์ให้ได้ตามต้องการในบางกรณี

4.2.2 องค์ประกอบของชุด Prompt สำหรับใช้ในการเพิ่มจำนวนข้อมูล (Augmentation)

Prompt ตั้งต้น หรือ Meta-prompt (System Message) ที่ใช้กับโมเดล GPT-3.5 สำหรับการเพิ่มจำนวนข้อมูลแชทแบบฟิชชิ่งที่ให้ผลดีเพียงพอ คือ:

> *“Please create phishing chat message sample for 10 messages that you may see variation enough but same intention for further phishing detection model training, by learning from the actual chat sample provided.”*

Prompt ตั้งต้น หรือ Meta-prompt (System Message) ที่ใช้กับโมเดล GPT-3.5 สำหรับการเพิ่มจำนวนข้อมูลแชทแบบถูกต้อง (Legitimate) ที่ให้ผลดีเพียงพอ คือ:

> *“You are AI for Augmentation to help me augment more sample for Data Sciences Model Training. Please create a legitimate chat message sample for 10 messages that you may see variation enough but same intention for further phishing detection model training, by learning from the actual legitimate customer chat sample provided below. Keep original intention and word length.”*

Prompt ปกติที่เป็นการให้ตัวอย่าง (Few-Shot) เป็นข้อมูลป้อนเข้า (Input) แก่โมเดลในการ

เพิ่มจำนวนข้อมูล ใช้การปรับแต่งข้อความตามข้อ 4.2.1 นอกจากนี้ Temperature ที่เป็น Hyperparameter ของการเพิ่มจำนวนข้อมูล ใช้ที่ 0.80 (เนื่องจาก Temperature ใน Azure AI Studio ตั้งค่าสูงสุดได้แค่ 1.00) เพื่อให้โมเดลมีความคิดสร้างสรรค์มากเพียงพอ แต่ไม่มากเกินจนผิดไปจากเจตนาของข้อความต้นฉบับที่ป้อนเข้า

4.2.3 องค์ประกอบของชุด Prompt สำหรับใช้ในการประเมินคุณภาพของข้อความแชทที่ถูกเพิ่มจำนวน

Prompt ตั้งต้น หรือ Meta-prompt (System Message) ที่ใช้กับโมเดล GPT-4o สำหรับการประเมินคะแนนความสมจริงของข้อความแชทแบบฟิชชิ่งที่ถูกเพิ่มจำนวนมา ที่ให้ผลดีเพียงพอ คือ:

> *"I have generated phishing chat samples by augmenting from actual phishing chat with other LLM model. So could you help me analyze if the phishing chat sample below have enough quality to use for further chat phishing and spamming detection model training, especially the realistic of phishing chat? Rate quality of each chat message in score 0 - 10 with explanation."*

Prompt ตั้งต้น หรือ Meta-prompt (System Message) ที่ใช้กับโมเดล GPT-4o สำหรับการประเมินคะแนนความสมจริงของข้อความแชทแบบถูกต้องที่ถูกเพิ่มจำนวนมา ที่ให้ผลดีเพียงพอ คือ:

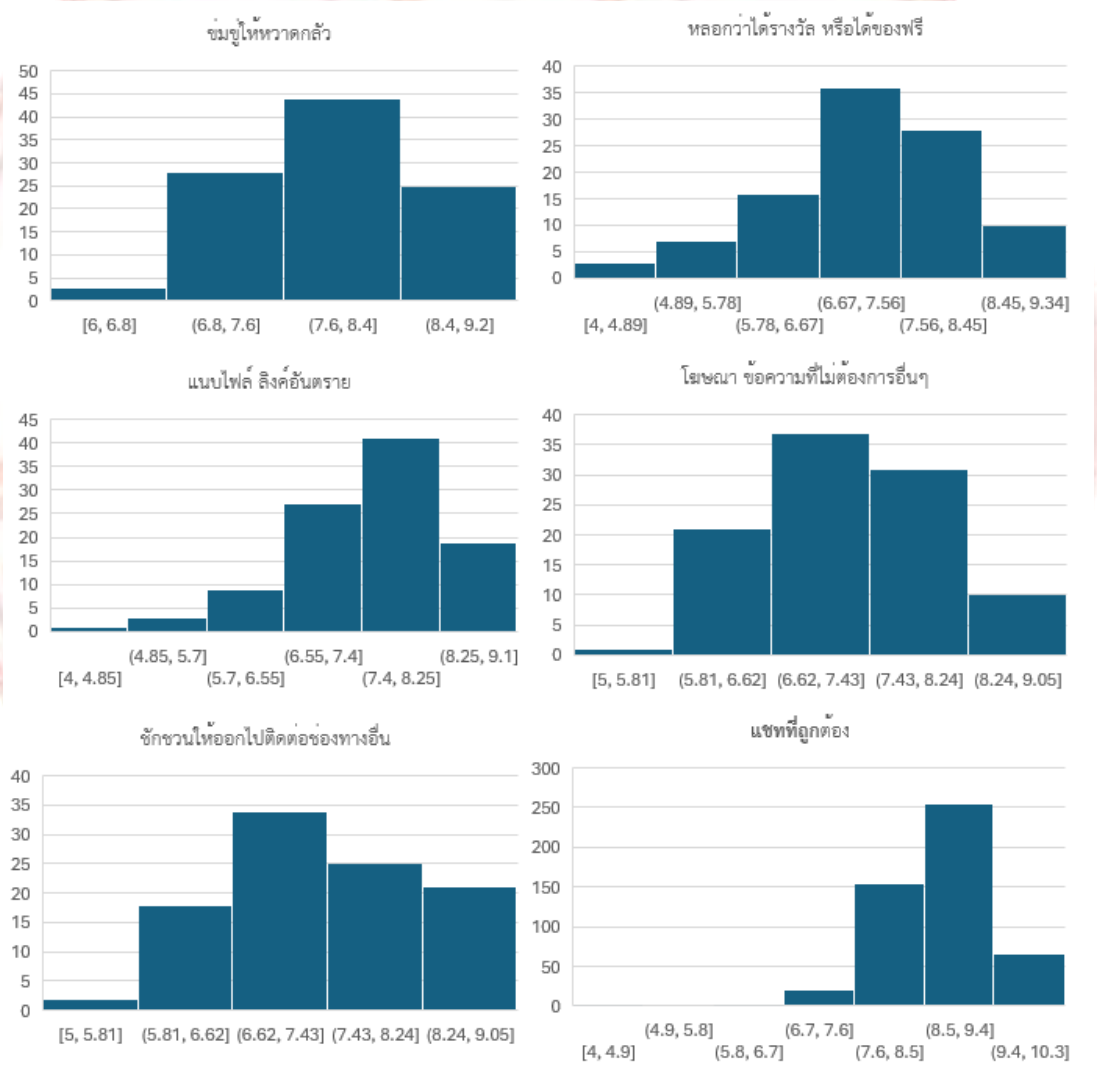
> "You are my Data Scientist as assistant to build chat message dataset for phishing chat detection model training. I have generated legitimate chat samples by augmenting from actual legitimate chat with other LLM models. So, could you help me analyze if the legitimate chat sample below has enough quality to use as "legitimate" label for further chat phishing detection model training, especially the realistic of the chat, like from actual human? Rate quality of each chat message in score 0 - 10 with explanation."

Prompt ปกติที่เป็นการให้ข้อมูลป้อนเข้า (Input) แก่โมเดล เป็นข้อความแชทที่ได้รับการเพิ่มจำนวนขึ้นมา ใช้การปรับแต่งข้อความตามข้อ 4.2.1 นอกจากนี้ Temperature ที่เป็น Hyperparameter ของการเพิ่มจำนวนข้อมูล ใช้ที่ 0.30 เพื่อให้โมเดลมีความเสถียร สามารถประเมินคะแนนความสมจริงได้แม่นยำใกล้เคียงทุกครั้ง (Precision) โดยยังคงความคิดสร้างสรรค์แบบมนุษย์

ในการให้ความเห็นและให้เหตุผลประกอบการประเมินได้

4.3 ผลการเพิ่มจำนวนข้อความแชท และประเมินคุณภาพข้อความ

ผู้วิจัยได้ทำการเพิ่มจำนวนข้อความแชทด้วยโมเดล GPT-3.5 และนำข้อความที่เพิ่มได้มาประเมินคะแนนความสมจริงในการนำไปใช้ต่อด้วยโมเดล GPT-4o พร้อมทั้งคัดข้อความที่ได้คะแนนความสมจริงต่ำสุด 10% ในแต่ละประเภทออกก่อนนำไปทำชุดข้อมูล (Dataset) จริง ได้ผลดังแสดงในภาพที่ 4-9 และตารางที่ 4-1 พร้อมทั้งกำกับคะแนนแต่ละข้อความไว้ในสเปรดชีทดังภาพที่ 4-10



ภาพที่ 4-9 ฮิสโตแกรมแสดงการกระจายตัวของคะแนนความสมจริงของข้อความแชทที่ถูกเพิ่มจำนวนแต่ละประเภท

ตารางที่ 4-1 คะแนนความสมจริงของข้อความแชทที่ถูกเพิ่มจำนวนขึ้นมา ในแต่ละประเภท

ประเภทข้อความแชท	คะแนนน้อยที่สุด	คะแนนเฉลี่ย	คะแนนมากที่สุด
ข่มขู่ให้หวาดกลัว	6	7.91	9
หลอกว่าได้รางวัล หรือได้ของฟรี	4	7.09	9
ชักชวนให้ออกไปติดต่อช่องทางอื่น	5	7.45	9
แนบไฟล์ ลิงค์อันตราย	4	7.61	9
โฆษณา ข้อความที่ไม่ต้องการอื่นๆ	5	7.28	9
แชทที่ถูกต้อง	4	8.71	10
คะแนนเฉลี่ย (จากข้อความที่เพิ่มจำนวนขึ้นมาทั้งหมด 500 ข้อความ)	**4**	**7.68**	**10**
คะแนนเฉลี่ยของข้อความที่ได้คะแนนสูงสุด 90% (จากข้อความที่เพิ่มจำนวนขึ้นมาทั้งหมด 450 ข้อความ)	**6**	**7.86**	**10**

No.	GPT-3.5 Augmented Chat Message	Label (Phishing/ Legitimate)	SubType	GPT-4o Realistic Score (10)	Augmented From
60	@david_wilson12 Hey there, I have a great opportunity for us to make some money. Text me on WhatsApp: +85261489740	Phishing	Ads,OtherChannel	7	OriAdsOther06
61	@jessica_brown77 Interested in making some extra money? I have a great opportunity for you. Contact me on WhatsApp: +852614897	Phishing	Ads,OtherChannel	7	OriAdsOther06
62	@ryan_chen88 I'm looking for someone to partner with to make some extra cash. WhatsApp me at +85261489740	Phishing	Ads,OtherChannel	7	OriAdsOther06
63	@kelly_green44 Hey, let's team up to make some easy money. Message me on WhatsApp: +85261489740	Phishing	Ads,OtherChannel	7	OriAdsOther06
64	Hey there, I just wanted to drop a quick message to let you know that we have some amazing watches available for purchase. Follow u	Phishing	Ads	7	OriAdsOther07
65	Hi, we have some incredible watches available for purchase right now. For more information, follow us on Instagram or Whatsapp. You	Phishing	Ads	7	OriAdsOther07
66	Hey, if you're looking for some top-quality watches, look no further! Follow us on Instagram or Whatsapp to see what we have to offer	Phishing	Ads	7	OriAdsOther07
67	Hi, we have some of the most amazing watches from major brands that you can purchase now. Follow us on Instagram or Whatsapp to	Phishing	Ads	7	OriAdsOther07
68	Hey! Looking for a new watch? We've got you covered with our selection of top-quality watches from all the major brands. We offer fa	Phishing	Ads	7	OriAdsOther08
69	Greetings! Are you in the market for a luxury watch? We have a wide range of high-quality options from major brands available for fas	Phishing	Ads	7	OriAdsOther08
70	Hello there! We specialize in selling only the highest quality luxury watches from major brands. We offer fast and tax-free shipping wo	Phishing	Ads	7	OriAdsOther08
71	Hey! Looking for a new watch to add to your collection? We have a wide variety of high-quality options from major brands available f	Phishing	Ads	7	OriAdsOther08
72	Greetings! We take pride in offering only the best luxury watches from major brands. With fast and tax-free shipping worldwide, we m	Phishing	Ads	7	OriAdsOther08
73	Hello! We have a wide range of high-quality luxury watches from major brands available for fast, tax-free shipping worldwide. If you're	Phishing	Ads	7	OriAdsOther08
74	Hey there! We specialize in selling only the best luxury watches from major brands. With fast and tax-free shipping worldwide, we mak	Phishing	Ads	7	OriAdsOther08
75	Hey! We just wanted to reach out to let you know that we have some great scopus indexed journals with low APC that would be perfec	Phishing	Ads	7	OriAdsOther09
76	Hi there! We have scopus indexed journals with low APC and we think your unpublished articles would be a great fit. Would you be int	Phishing	Ads	7	OriAdsOther09
77	Hi! We have scopus indexed journals with low APC and we think they would be perfect for your unpublished articles. Would you like to	Phishing	Ads	7	OriAdsOther09
78	Hi, I'm from the IRS, and we have noticed that you owe back taxes. If you don't pay immediately, we will issue an arrest warrant. Can yc	Phishing	Ads	7	OriAdsOther10
79	Hello, hope you're doing well! I stumbled upon your IG page and I couldn't help but notice it could use some more followers to boost	Phishing	Ads	6	OriAdsOther01
80	Hello! I noticed your IG page and I think it could use some more followers to enhance its appearance. Our instant follower services pro	Phishing	Ads	6	OriAdsOther01
81	Hey, I am a representative of Microsoft, and we have noticed that your Windows license has expired. To continue using your operating	Phishing	Ads	6	OriAdsOther02
82	Hello there, we have an exclusive offer for you! Get your hands on the rarest and most exquisite Shazar Dendritic Agate Stone. It's perf	Phishing	Ads	6	OriAdsOther03
83	Good day! We wanted to offer you our Shazar Dendritic Agate Stone, which is perfect for jewelry and home decor. It is a rare gemston	Phishing	Ads	6	OriAdsOther03
84	Hey, I came across your venue and I think it would be a perfect fit for our app, Beauty Pass. We're looking for exclusive venues like your	Phishing	Ads	6	OriAdsOther04
85	Hi there, I came across your venue on Beauty Pass and was wondering if you'd be interested in offering one of your services for free to	Phishing	Ads	6	OriAdsOther04
86	Hello [name], I'm reaching out to you on behalf of a fashion brand that would like to collaborate with your store. To discuss further, pl	Phishing	Ads	6	OriAdsOther05
87	Hi [name], we're a well-known fashion brand and we would like to collaborate with your store. Can you please add me on WeChat at [	Phishing	Ads	6	OriAdsOther05
88	Hello [name], I love your store and I think our brand would be a great fit for a collaboration. Please send me a message on my Telegra	Phishing	Ads	6	OriAdsOther05
89	Hi [name], I'm a brand representative and I would like to discuss a collaboration opportunity with you. Please add me on LINE at [phis	Phishing	Ads	6	OriAdsOther05
90	@sarahjane22 Want to join me in making some easy money? Let's chat on WhatsApp: +85261489740	Phishing	Ads,OtherChannel	6	OriAdsOther06

ภาพที่ 4-10 ตัวอย่างข้อความแชทฟิชชิ่งที่ถูกเพิ่มจำนวนมาแล้ว พร้อมคะแนนประเมินความสมจริงกำกับ

4.4 ผลการปรับปรุงชุดข้อมูลให้พร้อมต่อการนำไปใช้พัฒนาโมเดลการเรียนรู้ด้วยเครื่อง

ผู้วิจัยได้รวบรวมชุดข้อความแชททั้งข้อความจริง และข้อความที่สร้างเพิ่มขึ้นมาใหม่และผ่านการคัดกรองคุณภาพแล้วมาไว้เป็นชุดเดียวกัน และปรับให้อยู่ในรูปแบบไฟล์ CSV เพื่อนำเข้า Azure

Blob Storage จากนั้นให้ Azure Machine Learning Studio นำไปใช้พัฒนาและตรวจคุณภาพโมเดล ซึ่งพบปัญหาที่ต้องใช้วิธีการแก้ไขดังนี้

4.4.1 เนื่องจากข้อความแชทภาษาอังกฤษมีใช้เครื่องหมายลูกน้ำ “,” ในรูปประโยค จึงใช้สัญลักษณ์ “,” เป็นเครื่องหมายแบ่งส่วนข้อมูล (Delimiter) ไม่ได้ ต้องใช้เครื่องหมายพิเศษอื่นที่ไม่มีในข้อความแชท เช่น | (เนื่องจาก Azure ML Studio ตั้งค่าสัญลักษณ์ Delimiter ได้แค่อักขระเดียว)

4.4.2 เนื่องจากไฟล์ CSV ไม่รับการขึ้นบรรทัดใหม่ มองการขึ้นบรรทัดเป็นข้อมูล (Record) ใหม่ จึงต้องตัดสัญลักษณ์แทนการขึ้นบรรทัดใหม่ (Line Break) ออกก่อนส่งออกเป็นไฟล์ CSV

4.4.3 ไฟล์ CSV ไม่รองรับอีโมจิ โดยจะแสดงเป็น “??” แทน ดังนั้น จึงไม่สามารถใช้อีโมจิในการแสดงความหมายของการสื่อสารเวลานำไปเทรนโมเดลต่อได้

4.5 ผลการทดสอบประสิทธิภาพของชุดข้อมูลกับโมเดล Decision (Random) Forest

ผู้วิจัยได้นำชุดข้อมูลหลังปรับปรุงตามแนวทางที่นำเสนอในข้อ 4.4 มาผ่านกระบวนการประมวลผลข้อมูล (Data Pipeline) บนบริการ Azure ML Studio และเลือกใช้ Decision (Random) Forest สำหรับทดสอบความสามารถของโมเดลที่เทรนโดยข้อมูลที่สร้างขึ้นในการตรวจจับแชทที่เข้าข่ายเป็นฟิชชิ่ง ดังแสดงในภาพที่ 4-11 พบว่าเพื่อให้ได้โมเดลที่มีประสิทธิภาพที่สุด ควรทำตามขั้นตอนดังต่อไปนี้

4.5.1 ตั้งต้นด้วยชุดข้อมูล CSV ที่ปรับปรุงแล้ว นำเข้ามาไว้ใน Blob Storage ของ Azure

4.5.2 ใช้คอมโพเนนท์ Preprocess Text เพื่อปรับปรุงข้อมูลให้อยู่ในรูปที่ประมวลผลง่ายขึ้น เช่น การตรวจจับรูปประโยค การแปลงให้เป็นตัวอักษรพิมพ์เล็กทั้งหมด การนำตัวเลขและอักขระพิเศษออก ฯลฯ

4.5.3 ใช้คอมโพเนนท์ Select Columns in Dataset เพื่อให้ระบบสนใจเฉพาะคอลัมน์ข้อความแชท และฉลากบอกประเภทข้อความแชท (Phishing หรือ Legitimate)

4.5.4 ใช้คอมโพเนนท์ Extract N-Gram Features from Text เพื่อเลือกคอลัมน์ข้อความแชทมาแยกคอลัมน์เพิ่มตามคำสำคัญที่พบในช่วง 3 – 25 ตัวอักษร และมีความถี่ตั้งแต่ 5 คำขึ้นไป ให้ได้เป็นข้อมูลความถี่ที่พบคำสำคัญที่วิเคราะห์ออกมาได้ ซึ่งจะได้จำนวนคอลัมน์ตามจำนวนคำสำคัญที่คอมโพเนนท์นี้วิเคราะห์

4.5.5 ใช้คอมโพเนนท์ Split Data เพื่อแบ่งชุดข้อมูลด้วยอัตราส่วน 0.9 แบบสุ่ม

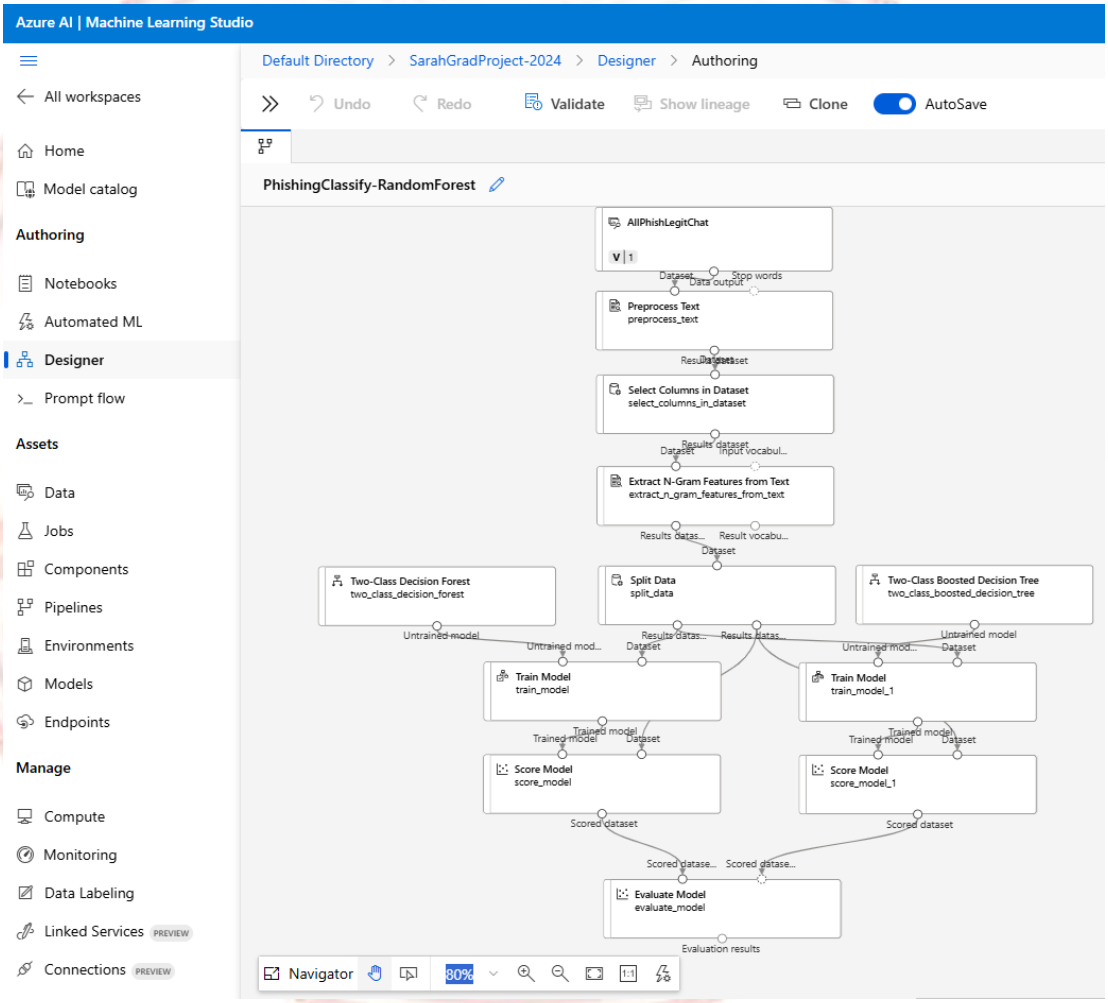
4.5.6 นำชุดข้อมูลส่วนใหญ่ 90% ที่แบ่งมาได้ มาเทรนโมเดล โดยทดลองเทรนโมเดลด้วยอัลกอริทึมสองแบบ:

4.5.6.1 Two-Class Decision Forest โดยกำหนดจำนวนทรีที่ 50 trees, ลำดับชั้นความลึกมากที่สุดที่ 15 ชั้น, และจำนวนตัวอย่างขั้นต่ำในแต่ละ Leaf Node เป็น 2 ตัวอย่าง

4.5.6.2 Two-Class Boosted Decision Tree โดยกำหนดจำนวน Leaves ต่อทรีที่ 30 Leaf, จำนวนตัวอย่างขั้นต่ำในแต่ละ Leaf Node ที่ 20 ตัวอย่าง, อัตราการเรียนรู้อยู่ที่ 0.05, และจำนวนทรีที่สร้างขึ้นเท่ากับ 200 trees

4.5.7 นำชุดข้อมูลส่วนน้อย 10% ที่แบ่งไว้ก่อนหน้า มาให้คะแนนโมเดลแต่ละโมเดล

4.5.8 ใช้คอมโพเนนท์ Evaluate Model เพื่อเปรียบเทียบค่าประสิทธิภาพระหว่างทั้งสองโมเดล



ภาพที่ 4-11 Data Pipeline ที่ออกแบบเพื่อนำชุดข้อมูลที่ได้มาเทรนโมเดล Machine Learning บน Azure ML Studio

ผลการทดสอบพบว่าค่าประสิทธิภาพของโมเดลทั้งสองแบบดีมาก สามารถนำมาจำแนกประเภทข้อความแชทเพื่อตรวจจับฟิชชิ่งได้อย่างแม่นยำในช่วง 97 – 99% โดยจากการทดสอบครั้งนี้

พบว่าโมเดลที่ใช้อัลกอริทึมแบบ Two-Class Boosted Decision Tree มีประสิทธิภาพโดยรวมเหนือกว่าโมเดล Two-Class Decision Forest เล็กน้อย แต่มีประสิทธิภาพการตรวจจับดีกว่าโมเดลที่ถูกเทรนโดยชุดข้อมูลตั้งต้นที่มีจำนวนจำกัดเพียงอย่างเดียว ดังแสดงในตารางที่ 4-2

ตารางที่ 4-2 ผลการประเมินประสิทธิภาพโมเดล Machine Learning ที่นำชุดข้อความแชทที่เพิ่มจำนวนได้มาเทรน

ค่าประสิทธิภาพ	ใช้ชุดข้อความแชทตั้งต้นจำนวนน้อย		ใช้ชุดข้อความแชทที่เพิ่มจำนวนมา	
	Two-Class Decision Forest	Two-Class Boosted Decision Tree	Two-Class Decision Forest	Two-Class Boosted Decision Tree
ค่าความแม่นยำ (Accuracy)	0.900	0.800	0.989	0.978
ค่าความเที่ยงตรง (Precision)	1.000	1.000	1.000	1.000
ค่า Recall	0.800	0.600	0.976	0.952
ค่าคะแนน F1	0.889	0.750	0.988	0.976
Area Under Curve (AUC)	0.960	0.800	1.000	1.000
Fault Positives	1 ข้อความ	2 ข้อความ	1 ข้อความ	2 ข้อความ

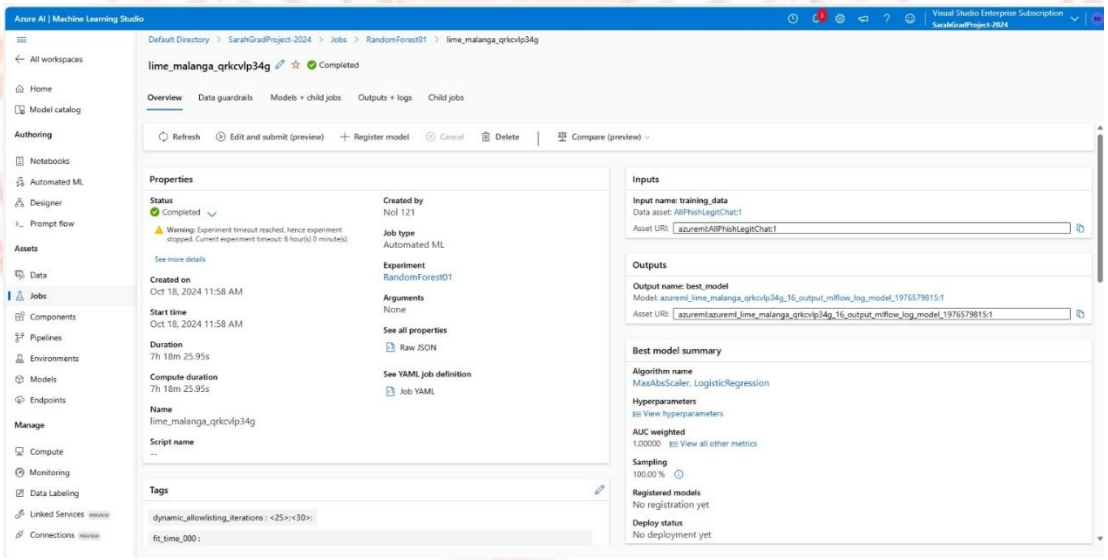
4.6 ผลการทดสอบประสิทธิภาพของชุดข้อมูลกับโมเดลที่คัดเลือกโดยการเทรนแบบอัตโนมัติ (AutoML)

เพื่อให้ไม่เกิดการลำเอียงจากการออกแบบ Data Pipeline ที่สร้างขึ้นเอง ผู้วิจัยได้ทดสอบประสิทธิภาพของการนำชุดข้อมูลที่สร้างขึ้นและผ่านการปรับปรุงแล้วมาผ่านกระบวนการเทรนโมเดล Machine Learning แบบอัตโนมัติ ด้วยฟีเจอร์ AutoML บนระบบ Azure ML Studio ซึ่งเป็นการรันเทียบหลายโมเดลตั้งต้นพร้อมกัน อันได้แก่ XGBoostClassifier, Extreme Random Trees, Random Forest, Extreme Random Trees, Logistic Regression, LightGBM, Stack Ensemble, และ Voting Ensemble เพื่อหาโมเดลที่ให้ประสิทธิภาพดีที่สุด

หลังจากรันการทำงานครบ 6 ชั่วโมง ได้โมเดลที่ทำงานดีที่สุดที่ใช้อัลกอริทึม MaxAbsScaler ในการแปลงข้อมูล (Data Transformation) ร่วมกับการใช้อัลกอริทึม Logistic Regression ในการเทรนโมเดล ซึ่งได้ค่าประสิทธิภาพโดยรวมดีกว่าโมเดลที่ออกแบบ Data Pipeline เอง โดยให้ค่าประสิทธิภาพแบบถ่วงน้ำหนัก (Weighted) ดังแสดงในตารางที่ 4-3 และภาพที่ 4-12

ตารางที่ 4-3 ผลการประเมินประสิทธิภาพโมเดลที่คัดเลือกโดยการเทรนแบบอัตโนมัติ (AutoML) ที่นำชุดข้อความแชทที่เพิ่มจำนวนได้มาเทรน

ค่าประสิทธิภาพ	ใช้ชุดข้อความแชทตั้งต้นจำนวนน้อย	ใช้ชุดข้อความแชทที่เพิ่มจำนวนมา
กลไกการสกัดฟีเจอร์ที่ดีที่สุด	StandardScalerWrapper	MaxAbsScaler
อัลกอริทึมเทรนโมเดลที่ได้ผลดีที่สุด	Logistic Regression	Logistic Regression
ค่าความแม่นยำ (Accuracy)	0.77000	0.99875
ค่าความเที่ยงตรง (Precision)	0.86778	0.99880
ค่า Recall	0.77000	0.99877
ค่าคะแนน F1	0.76025	0.99877
Area Under Curve (AUC)	1.00000	1.00000



ภาพที่ 4-12 การนำชุดข้อข้อมูลข้อความแชทมาใช้กับฟีเจอร์ Automated ML บน Azure ML Studio

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

จากผลการค้นคว้าอิสระเรื่อง การสร้างชุดข้อมูลข้อความแชทฟิชชิ่งด้วยโมเดลการเรียนรู้แบบ Few-Shot สามารถสรุป อภิปราย และได้ข้อเสนอแนะดังต่อไปนี้

5.1 สรุปผลการวิจัย

จากผลการเก็บรวบรวมข้อความแชทจริงสำหรับนำมาใช้ในการเพิ่มจำนวนข้อมูล แม้จำนวนที่วางแผนไว้ค่อนข้างน้อยแล้ว แต่ก็ไม่สามารถเก็บรวบรวมได้ครบตามจำนวนที่วางแผนไว้แต่แรกในแต่ละประเภท โดยเฉพาะข้อความแชทแบบฟิชชิ่ง เพราะแพลตฟอร์มโซเชียลมักมีการลบข้อความน่าสงสัยโดยอัตโนมัติ และมีการลบข้อความที่เก่าเกินไป (อายุตั้งแต่ 6 เดือนขึ้นไป) บางส่วนโดยอัตโนมัติเช่นกัน จึงต้องอาศัยการร้องขอข้อมูลเพิ่มจากบุคคลอื่น หรือที่เผยแพร่สาธารณะบนอินเทอร์เน็ตมาให้ครบตามจำนวนที่วางแผนไว้แต่ต้น ทั้งนี้สังเกตได้ว่าจำนวนข้อความแชทแบบฟิชชิ่งที่เก็บรวบรวมได้มากที่สุดคือ ประเภทที่ข่มขู่ให้หวาดกลัว ขู่จะปิดบัญชีถ้าไม่รีบทำตามที่บอก (Ban Scare)

จากผลการออกแบบ Prompt เพื่อเพิ่มจำนวนข้อมูลได้อย่างมีประสิทธิภาพ พบว่าต้องมีการแปลงข้อมูลให้อยู่ในรูปที่โมเดล LLM เข้าใจง่ายก่อน ซึ่งลดทอนลักษณะหลายอย่างของข้อความต้นฉบับไป และการให้โมเดลทำงานที่ซับซ้อนมากเท่าไร ก็ต้องใส่รายละเอียดใน Meta-Prompt ให้ครบองค์ประกอบมากเท่านั้น ดังจะเห็นได้ว่า Meta-Prompt ที่ใช้ในการเพิ่มจำนวนและตรวจคุณภาพข้อความแชทแบบถูกต้อง (Legitimate) ต้องเพิ่มการกำหนดบทบาทของโมเดลว่าเป็นผู้ช่วยในฐานะนักวิทยาศาสตร์ข้อมูลด้วย นอกจากบอกวัตถุประสงค์ อธิบายรายละเอียดสิ่งที่ต้องการ และเงื่อนไข เพราะข้อความแชทแบบถูกต้องมักมีขนาดสั้นกว่า และใช้ภาษาที่เป็นกันเองกว่าข้อความแชทแบบฟิชชิ่งที่ค่อนข้างมีลักษณะแน่นอน

จากผลการเพิ่มจำนวนข้อความแชท พบว่าโมเดล LLM มีความเที่ยงตรงต่ำ โดยเฉพาะเมื่อเราต้องการใช้ประโยชน์จากความคิดสร้างสรรค์จนต้องกำหนด Hyperparameter อย่างค่าความสุ่ม (Temperature) สูง ก็ยิ่งทำให้คำตอบบางครั้งผิดเพี้ยนไปจากที่ต้องการ เช่น กลายเป็นการตอบคำถามของ Prompt ข้อความต้นฉบับแทนการเพิ่มจำนวนให้ได้ข้อความตามเจตนาของต้นฉบับ ทำให้ต้องป้อน Prompt ที่เป็นการให้ความเห็นของคำตอบที่ออกมา พร้อมอธิบายเพิ่มเติมให้ชัดเจนขึ้นหลายครั้ง แตกต่างกันในแต่ละกรณี

จากผลการประเมินคุณภาพ ความสมจริงของข้อความแชทที่เพิ่มจำนวนขึ้นมา พบคะแนนอยู่

ในช่วง 10 – 6 ซึ่งจากการสังเกต และการให้เหตุผลของโมเดลประกอบคะแนน มักประเมินจากความยาว สัญลักษณ์พิเศษ ความสมจริงของ URL แต่องค์ประกอบเหล่านี้ต้องถูกลดทอนลงไปอีกเมื่อนำไปเทรนโมเดล ทั้งจากการแปลงเป็น CSV และการแปลงข้อมูลใน Data Pipeline ทั้งนี้พบว่าคะแนนความสมจริงเฉลี่ยของข้อความแชทแบบถูกต้องที่เพิ่มจำนวนมา (8.71) มากกว่าข้อความแชทแบบฟิชชิ่งทุกประเภท (อยู่ในช่วง 7.00 – 7.91) และพบว่าในกลุ่มข้อความแชทแบบฟิชชิ่งนั้น ข้อความที่ข่มขู่ให้หวาดกลัว (Ban Scare) ได้รับคะแนนความสมจริงมากที่สุด เฉลี่ยที่ 7.91

จากผลการเตรียมชุดข้อมูลให้พร้อมต่อการนำไปเทรนโมเดลเรียนรู้ด้วยเครื่อง (Machine Learning) พบว่าต้องใช้กระบวนการแปลงข้อมูลหลายขั้นตอนเพื่อให้ระบบปลายทางสามารถทำความเข้าใจชุดข้อมูลได้ ซึ่งในบางขั้นตอนจะมีการลดทอนลักษณะสำคัญของข้อความแชทด้วยเช่นกัน

จากผลการทดสอบคุณภาพของชุดข้อมูลข้อความแชทที่สร้างขึ้น ด้วยการนำไปเทรนโมเดลการเรียนรู้ด้วยเครื่อง (Machine Learning) แล้ววัดประสิทธิภาพ พบว่าได้โมเดลที่มีประสิทธิภาพสูงมาก ทั้งจากการเทรนด้วยการออกแบบ Data Pipeline ด้วยตนเอง และจากฟีเจอร์การเทรนหาโมเดลที่ดีที่สุดแบบอัตโนมัติ (Automated AI) จึงสรุปผลได้ว่า การสร้างชุดข้อมูลข้อความแชทแบบฟิชชิ่ง ด้วยเทคนิคการเรียนรู้จากตัวอย่างจำนวนน้อยมาก (Few-Shot) เพื่อให้โมเดล LLM เพิ่มจำนวนชุดข้อมูลตามขั้นตอนข้างต้นนี้ ทำให้ได้ชุดข้อมูลที่มีคุณภาพสูงเพียงพอ และสูงกว่าการใช้ชุดข้อมูลตั้งต้นที่มีจำนวนน้อย และสามารถนำไปใช้ประโยชน์ในการเทรนโมเดลต่อ หรือนำไปใช้ประโยชน์ในทางอื่น เช่น นำไปฝึกอบรมสร้างความตระหนัก (Awareness) ในองค์กรได้ เป็นต้น

มีข้อสังเกตว่า สำหรับโมเดลที่เทรนจาก Data Pipeline ที่เตรียมข้อมูลด้วยการวิเคราะห์ความถี่ของคำสำคัญ (N-gram) จะนำไปใช้งานจริงกับข้อความแชทใหม่ๆ ได้ยาก เนื่องจากผลลัพธ์ของ N-gram ที่สร้างคอลัมน์ตามจำนวนคำสำคัญจะแตกต่างกัน ทั้งชื่อคอลัมน์ และจำนวนคอลัมน์ ทำให้ไม่สามารถนำโมเดลเดิมมาใช้ได้

5.2 อภิปรายผลการวิจัย

งานวิจัยครั้งนี้เป็นการเรียนรู้อิสระเพื่อแก้ปัญหาธุรกิจออนไลน์ที่มีข้อจำกัดเรื่องการรวบรวมข้อมูลข้อความแชทเพื่อนำมาใช้ประโยชน์ในการเทรนโมเดลตรวจจับข้อความแชทฟิชชิ่งอัตโนมัติ และจนถึงปัจจุบันก็ยังไม่มีชุดข้อมูลข้อความแชทที่เผยแพร่สาธารณะในรูปแบบที่พร้อมนำมาใช้งานได้ ซึ่งผลที่ได้จากงานวิจัยนี้พิสูจน์ได้ว่าเราสามารถนำเทคโนโลยี AI แบบ LLM มาเพิ่มจำนวนชุดข้อมูลให้ได้คุณภาพมาใช้ประโยชน์ได้จริง

เพียงแต่ ขั้นตอนการรวบรวมและปรับแต่งข้อมูลกว่าจะนำมาเทรนโมเดลได้นั้น มีอุปสรรคทั้งจำนวนข้อความแชทเริ่มต้นที่รวบรวมได้ยาก และไม่สมดุลกันในแต่ละประเภทที่วางแผนไว้ และมีข้อจำกัดให้โดนลดทอนลักษณะสำคัญหลายประการที่ใช้คัดแยกข้อความแชทแบบฟิชชิ่งได้ ไม่ว่าการ

ขึ้นบรรทัด จัดรูปหน้า อีโมจิ คำอังกฤษที่ทับเสียงภาษาไทย ฯลฯ อีกทั้งความแม่นยำที่ต่ำของโมเดล LLM ทำให้คุณภาพของชุดข้อมูลที่ได้ไม่ได้สมบูรณ์แบบอย่างที่ตั้งใจไว้ แม้คุณภาพโดยรวมจะอยู่ในเกณฑ์ที่นำไปใช้ประโยชน์ได้ก็ตาม

5.3 ข้อเสนอแนะ

จากปัญหาหลายประการที่พบและต้องแก้ไขเฉพาะหน้าในงานวิจัยครั้งนี้ จนทำให้ยากที่จะทำให้กระบวนการสร้างชุดข้อมูลนี้เป็นระบบอัตโนมัติตั้งแต่ต้นทางออกมาเป็นชุดข้อมูลสุดท้ายที่ต้องการ และเป็นเหตุผลที่ผู้วิจัยเลือกกระบวนการเพิ่มจำนวนข้อมูลผ่านอินเทอร์เฟซหน้าเว็บด้วยตนเองแทนสคริปต์โปรแกรมมิ่งอัตโนมัติ (ผ่านฟีเจอร์ Playground ของ Azure AI Studio) เพื่อที่สามารถติดตามพฤติกรรมของโมเดล LLM ที่ค่อนข้างไม่แน่นอน ไม่แม่นยำ ให้ผู้วิจัยสามารถปรับแก้ได้อย่างละเอียด

ถ้าเราสามารถนำแต่ละปัญหาที่พบนี้มาออกแบบ อุดช่องโหว่ทั้งหมด พัฒนาเป็นระบบอัตโนมัติตั้งแต่ต้นจนจบ จะช่วยแบ่งเบาภาระในการพัฒนาโมเดลตรวจจับข้อความแชทแบบฟิชชิ่งที่เหมาะสมกับบริบทของแต่ละองค์กรได้

นอกจากนี้ ด้วยการพัฒนาของโมเดล LLM ที่มีความสามารถมากขึ้นอย่างต่อเนื่อง มีการเรียนรู้ชุดข้อมูลสาธารณะมากขึ้น ก็อาจปรับใช้โมเดล LLM มาตัดสินการเป็นข้อความแชทแบบฟิชชิ่งได้โดยตรงแบบไม่ต้องเรียนรู้จากตัวอย่าง (Zero-Shot Learning) ได้อย่างรวดเร็วและมีประสิทธิภาพ ซึ่งเราสามารถทำให้โมเดล LLM นั้นเข้าใจบริบทที่จำเพาะขององค์กรได้ด้วยการ Augment ข้อมูลองค์กรเข้าไปผ่านเทคโนโลยี RAG (Retrieval-Augmented Generation) เป็นต้น

บรรณานุกรม

Hijji, M., & Alam, G. (2021). A multivocal literature review on growing social engineering based cyber-attacks/threats during the COVID-19 pandemic: Challenges and prospective solutions. IEEE Access, 9, 7152–7169. https://doi.org/10.1109/ACCESS.2020.3048839

Leavitt, N. (2005). Instant messaging: A new target for hackers. Computer, *38*(4), 20–23. https://doi.org/10.1109/MC.2005.234

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy. IEEE Access, 11, 80218–80245. https://arxiv.org/abs/2307.00691

Roy, S., Naragam, K. V., & Nilizadeh, S. (2023). Generating phishing attacks using ChatGPT. arXiv preprint arXiv:2305.05133. https://arxiv.org/abs/2305.05133.

Bada, M., Sasse, A., & Nurse, J. R. C. (2014). Cyber security awareness campaigns: Why do they fail to change behaviour? arXiv preprint arXiv:1901.02672. https://arxiv.org/abs/1901.02672

Bower, J. (2023, May 3). Data augmentation using LLMs for better phishing datasets. AI Security. https://www.jamesbower.com/data-augmentation-using-llms-for-better-phishing-datasets/

Boumber, D., Tuck, B. R., Verma, R. M., & Qachfar, F. Z. (2024). LLMs for explainable few-shot deception detection. In Proceedings of the IWSPA ’24 (pp. 37–47). https://doi.org/10.1145/3643651.3659898

Koponen, J., & Rytsy, S. (2020). Social presence and e-commerce B2B chat functions. European Journal of Marketing, 54(6), 1205–1224. https://doi.org/10.1108/EJM-01-2019-0061

Mero, J. (2018). The effects of two-way communication and chat service usage on consumer attitudes in the e-commerce retailing sector. Electronic Markets, 28(2), 205–217. https://doi.org/10.1007/s12525-017-0281-2

Aleroud, A., & Zhou, L. (2017). Phishing environments, techniques, and countermeasures: A survey. Computers & Security, 68, 160–196. https://doi.org/10.1016/j.cose.2017.04.006.

Burtell, M., & Toner, H. (2024, March 8). The surprising power of next word prediction: Large language models explained, part 1. Center for Security and Emerging Technology (CSET). https://cset.georgetown.edu/article/the-surprising-power-of-next-word-prediction-large-language-models-explained-part-1/

Pachetti, E., & Colantonio, S. (2024). A systematic review of few-shot learning in medical imaging. Artificial Intelligence in Medicine, 156, 102949. https://doi.org/10.1016/j.artmed.2024.102949

Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H., & Wang, C. (2018). Machine learning and deep learning methods for cybersecurity. IEEE Access, 6, 35365–35381. https://doi.org/10.1109/ACCESS.2018.2836950

Oliveira, D., et al. (2017). Dissecting spear phishing emails for older vs. young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 6412–6424). ACM. https://doi.org/10.1145/3025453.3025831

Gualberto, E. S., de Sousa, R. D., Vieira, T. P. D. B., Da Costa, J. P. J. D., & Duque, C. (2020). The answer is in the text: Multi-stage methods for phishing detection based on feature engineering. IEEE Access, 8, 223529–223547. https://doi.org/10.1109/ACCESS.2020.3043396

Sonowal, G. (2020). Phishing email detection based on binary search feature selection. *SN Computer Science, 1*, Article 169. https://doi.org/10.1007/s42979-020-00194-z

Zabihimayvan, M., & Doran, D. (2019). Fuzzy rough set feature selection to enhance phishing attack detection. In 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1–6). https://arxiv.org/abs/1903.05675

Bhatnagar, P., & Degadwala, S. (2024). Efficient email spam classification with N-gram features and ensemble learning. International Journal of Emerging Trends in Engineering and Development. https://doi.org/10.32628/CSEIT2410220

Sapkota, R., et al. (2024). Multimodal large language models for image, text, and speech data augmentation: A survey. arXiv preprint arXiv:2501.18648. https://arxiv.org/abs/2501.18648

Hazell, J. (2023). Large language models can be used to effectively scale spear phishing campaigns. arXiv preprint arXiv:2305.06972. https://arxiv.org/abs/2305.06972

Loukili, S., Fennan, A., & Elaachak, L. (2024). Email subjects generation with large language models: GPT-3.5, PaLM 2, and BERT. International Journal of Emerging Technologies in Learning. http://doi.org/10.11591/ijece.v14i4.pp4655-4663

Desolda, G., Greco, F., & Viganò, L. (2024). APOLLO: A GPT-based tool to detect phishing emails and generate explanations that warn users. arXiv preprint arXiv:2410.07997. https://arxiv.org/pdf/2410.07997

Uplenchwar, S., Sawant, V., Surve, P., Deshpande, S., & Kelkar, S. (2022). Phishing attack detection on text messages using machine learning techniques. In 2022 IEEE Pune Section International Conference (PuneCon) (pp. 1–5). https://ieeexplore.ieee.org/document/10014876

Aljabri, M., & Mirza, S. (2022). Phishing attacks detection using machine learning and deep learning models. In 2022 7th International Conference on Data Science and Machine Learning Applications (CDMA) (pp. 175–180). https://ieeexplore.ieee.org/document/9736352

Wang, Y., Zhu, W., Xu, H., Qin, Z., Ren, K., & Ma, W. (2023). A large-scale pretrained deep model for phishing URL detection. In ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1–5). https://ieeexplore.ieee.org/abstract/document/10095719

Xu, P. (2021). A transformer-based model to detect phishing URLs. arXiv preprint arXiv:2109.02138. https://arxiv.org/abs/2109.02138

Ozcan, A., Catal, C., Donmez, E., & Ozcan, D. (2023). A hybrid DNN–LSTM model for detecting phishing URLs. Neural Computing and Applications, 35, 4957–4973. https://doi.org/10.1007/s00521-021-06401-z

Misra, K., & Rayz, J. (2022). LMs go phishing: Adapting pre-trained language models to detect phishing emails. In 2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT) (pp. 135–142). https://ieeexplore.ieee.org/document/10101955

Borges, P. V., Ladele, A. S., Cunha, Y. M. B. G., & Moraes, D. S. (2024). Multimodal prompt engineering for multimedia applications using the GPT model. ResearchGate. https://www.researchgate.net/publication/385073643

Wang, X., Li, Y., Chen, J., & Zhang, H. (2023). Promptagent: Strategic planning with language models enables expert-level prompt optimization. arXiv preprint arXiv:2310.16427. https://arxiv.org/pdf/2310.16427

Ye, Q., Axmed, M., Pryzant, R., & Khani, F. (2023). Prompt engineering a prompt engineer. arXiv preprint arXiv:2311.05661. https://arxiv.org/pdf/2311.05661

Keshmirian, A., Willig, M., & Hemmatian, B. (2024). Biased causal strength judgments in humans and large language models. ICLR 2024 Workshop Proceedings. https://openreview.net/pdf?id=544P6YidFk

Mozes, M., He, X., Kleinberg, B., & Griffin, L. D. (2023). Use of LLMs for illicit purposes: Threats, prevention measures, and vulnerabilities. arXiv preprint arXiv:2308.12833. https://arxiv.org/pdf/2308.12833

ภาคผนวก ก
ข้อความแชทตั้งต้น

ข้อความแชทตั้งต้นที่นำมาใช้ในงานวิจัยนี้ แบ่งตามแต่ละประเภท มีดังต่อไปนี้

ตารางที่ ก-1 ข้อความแชทจริง แบบฟิชชิ่งประเภทข่มขู่ให้หวาดกลัว

ข้อความที่	ข้อความ
1	⚠️ ATTENTION ⚠️ Dear User, You will not be allowed to use Meta Products to advertise anymore because your account didn't comply with our Advertising policies affecting business assets, such as having too many ads rejected, attempting to circumvent our ad review process, participating in fraudulent behavior, or associating with untrustworthy accounts. Our system will automatically disabled your ads account in 24 hours, unless you send us a disclaimer via: https://fyp.bio/info-confirm81226548963 Thank you for your help in improving our service. Sincerely, Business Support Team Noreply Facebook. Meta Platforms, Inc., Attention: Community Support, 1 Facebook Way, Menlo Park, CA 94012
2	• Last Warning Need to confirm that this page belongs to you and is being used normally. Confirm here: https://meta-manage-business.com After 24 hours of not receiving a response, we will temporarily lock this page. Meta Support ©2024 (Automatic notification _ Please contact now)
3	WARNING Dear admin page! Your page has infringed copyright information! Your account has been detected as violating our current copyright policy. For security reasons, verify your account by clicking the link below. Verify this related account : https://business-ads-support.com If you do not confirm, our system will automatically block your account. Thank you for helping us improve our account service. Copyright © 2024 FB Pages Support All rights reserved.

ข้อความที่	ข้อความ
4	⚠ Important Notification: Your Facebook page is scheduled for permanent deletion due to a post that has infringed upon our trademark rights. We have reached this decision after a thorough review and in accordance with our intellectual property protection policies. If you believe this to be a misunderstanding, we kindly request you to file a complaint seeking the reinstatement of your page prior to its removal from Facebook. Request for Review: https://diglink.id/Processing_system-AU We understand that this situation may impact your ongoing business operations. However, please be informed that if we do not receive a complaint from you, our decision will be final. Your cooperation and understanding are greatly appreciated. Should you have any inquiries or apprehensions, please feel free to reach out to us. Sincerely, Facebook Support Team
5	•Last Warning ⚠ Your fanpage has received a copyright warning, you need to contact and confirm your page. If you do not confirm, our system will automatically block your account. Contact us for support: https://ltx.bio/helpbusinesscase_1942 Your distribution will be limited if your Page is flagged repeatedly. If you do not contact us after 24 hours, the page will be unpublished to protect the community (Automatic notification _ Please contact now)
6	Important Notification: Your Facebook page is scheduled for permanent deletion due to a post that has infringed upon our trademark rights. We have reached this decision after a thorough review and in accordance with our intellectual property protection policies. If you believe this to be a misunderstanding, we kindly request you to file a complaint seeking the reinstatement of your page prior to its removal from Facebook. Request for Review: https://hyperfollow.com/helppageconsumers2fa We understand that this situation may impact your ongoing business operations. However, please be informed that if we do not receive a complaint

ข้อความที่	ข้อความ
	from you, our decision will be final. Thank you for your cooperation and understanding. If you have any questions or concerns, please do not hesitate to contact us. Sincerely, Facebook Support Team © Noreply Facebook. Meta Platforms, Inc., Attention: Community Support, 1 Facebook Way, Menlo Park, CA 94025
7	Your Facebook page is scheduled for permanent deletion due to a post that has infringed upon our trademark rights. We have reached this decision after a thorough review and in accordance with our intellectual property protection policies. If you believe this to be a misunderstanding, we kindly request you to file a complaint seeking the reinstatement of your page prior to its removal from Facebook. Request for Review: contact-mfb.com We understand that this situation may impact your ongoing business operations. However, please be informed that if we do not receive a complaint from you, our decision will be final. Thank you for your understanding and cooperation. Thank You Meta Support Team!
8	Dear Administrator, Your account will be deactivated. This is because your page, or the activity on it, does not comply with our Terms of Service. If you believe that the deactivation of this account was a mistake, we can guide you through some steps to request protection. You should complete these steps in just a few minutes to avoid your account being permanently deleted. Please confirm your account here ◉ https://contact.supportverify-fanpage.com/meta-community-standard/121179327930871 If you do not confirm, our system will automatically block your account and you will not be able to use it again. Thank you for helping us improve our service. Terms of Service Team © 2024 Inc
9	Dear administrator, We have received multiple reports that your business account has violated the terms of service and community guidelines. 1. Using false names/images of others. 2. Sharing content

ข้อความที่	ข้อความ
	that puts other users at risk. 3. Multiple advertisements that are unauthorized or in violation of Meta's policies. 4. Using tricks to bypass Meta's ad verification system. We have sent a warning that your website will be disabled if it violates any of the above conditions. As a result, your account is scheduled for review. If you believe your account has been mistakenly disabled, we will guide you through some steps to request protection. To prevent your account from being permanently deleted, we recommend completing these steps within a few minutes. Please review and submit your complaint here. ◉ pages-manage.info You have 24 hours to file an objection to our decision. If this deadline is missed, your account will be permanently disabled. Thank you for your cooperation in improving our services. Group Terms of Service © 2023 Inc"
10	⚠ We have restricted your ad account, Case #1000434829 Here's how you can resolve this issue seamlessly and keep your ads running: Your fanpage has received a copyright warning, you need to contact and confirm your page. If you do not confirm, our system will automatically block your account. Contact us for support: casepage3726.info Your distribution will be limited if your Page is flagged repeatedly. If you do not contact us after 24 hours, the page will be unpublished to protect the community (Automatic notification _ Please contact now)

ตารางที่ ก-2 ข้อความแชทจริง แบบฟิชชิ่งประเภทหลอกลวงว่าได้รางวัล หรือได้ของฟรี

ข้อความที่	ข้อความ
1	Free Accessories for you □ This is CREEPY ONES, a rising accessories brand for Alternative Style Enthusiasts. We've already sent 3-4 FREE pieces to: -Forbes Featured, Model. @vikkilenola (192k followers) -

ข้อความที่	ข้อความ
	Streamer/Cosplay Model. @Caitink_cosplay (1 1 1 k followers) -DJ. @Dysis_azzu (23.2k followers) ❤️‍🔥 We love your account's vibe. And we got 3-4 FREE accessories waiting for you to grab. Please drop a DM with the keyword “COLLB” to our Ambassadors Program page @Thecreepyones
2	Hello there 👋. We would love to invite you to join our world of Cats ambassadors, no matter how small or big your social following is! 😺 If you’re interested, we can send you 4 to 5 items of your choice for FREE. 🎁 Please send a DM «MODEL» to @Catsyou.official We’re excited for you to become a Catsyou Ambassador! 🌈 We look forward to hearing from you! 👉🏻 @Catsyou.official 👈🏻
3	Hello! This is Emma from @Thefuzzypet_Official Your profile caught our eye We are a world Famous brand Of toys and Clothing For Dogs/Cats Accessories We really like your pawtner And we believe that he's photogenic 📸 We Would Love To offer him 4 Free Gifts To represent our brand...All we ask for is to publish a post with a photo wearing the gifts so we can share them on our 400K+ Instagram Page Please drop us a message on our official Instagram page, @Thefuzzypet_Official. Let's embark on a journey of empowerment and style together, creating extraordinary creations that beautifully captivate. Cheers!
4	🕶️ Enjoy FREE Sunglasses on Us! Hello there! My name is Alessandra, and I’m reaching out on behalf of Graziaglow, a premier eyewear brand where luxury meets style. We believe you’d be the perfect ambassador for our brand! Graziaglow is excited to offer you 3 pairs of our luxurious sunglasses absolutely FREE! Each pair embodies our signature blend of elegance and modern chic. In return, we’d love to see your honest reviews and how you style our eyewear. Plus, there

ข้อความที่	ข้อความ
	are many more surprises and gifts in store for you! 🌟 To get started, simply DM "COLLAB" to @Graziaglow.ambassador and secure your exclusive spot before they're gone. (We're already expecting you!) 😃 We can't wait to see how you bring our collections to life. 💕 @Graziaglow.ambassador
5	Hello, Sarah! Our team is currently looking for Models/Ambassadors to wear our newest gold jewelry ⚡ By being an ambassador, you will get 3 FREE pieces ✨ huge exposure and plus lots of other amazing benefits! DM "Ambassador" to the main page @chrysusgems within 24 hours to claim your spot!
6	Hi there! Our agency is looking for paid advertisement on your Facebook page, if you are interested please get back to us, Thank you! Payments will be made before you post, the payment will be made on weekly basis via method of your choice and the payment per ad will be $5000 Thank you! Best regards
7	Hi, would you be interested in placing ads in your Instagram? We pay $1 0 0 per story You would be advertising our website https://instfollowers.net Please check it out and If interested DM me. Thank you.
8	Hello, sir / madam, we are collecting blue verification traffic for our site and we will only give you 1 article / video in a day that you can post on the page. I will pay you $2000 per week and my payment will be made in advance. Note: I never ask you for any administrator or editor access to your page. If you are interested so reply me this message or text me from your personal account! Thank you
9	Hey, My name is Sue! I specialize in achieving 95% and higher in billing and can save you money! What is best email to send you information?
10	Uncommon James: Thanks for introducing us to your friend! After they place their first order, we'll send you 25% off! While you wait, see all

ข้อความที่	ข้อความ
	the things you can use your discount on: http://ujsms.com/ankQOD-fKNd

ตารางที่ ก-3 ข้อความแชทจริง แบบฟิชชิ่งประเภทชักชวนให้ออกไปติดต่อช่องทางอื่น

ข้อความที่	ข้อความ
1	CONTACT US VIA DIRECT! ♦ COmments from real peOple ✍️ ♦ ViEws 👀 ♦ Likes 💙💜💛 ♦ SavEs 🔒 ♦ IncreAse yOur outreACh 💯 ⬆️ TOP YOur POsts will get tO the TOP COntACt us viA whatsAPp Or dirECt.
2	Hello, can you please send a message to @landypaw they want to collaborate with you ! We're looking for a few photogenic pets that can wear our bandanas and provide us with some photos. You will be getting free exclusive products as well as a featuring on their official page ! 💕 👉🏻 @landypaw 👈🏻
3	Hello! This is Eva from @Chicanva.official Your profile caught our eye We are a world Famous brand that can turn any pet picture into a beautiful art! we believe that your pet is a photogenic 📸 We Would Love To offer a good deal to represent our brand...All we ask for is to publish a post with a photo with our stuff so we can share repost it on our 70K+ Instagram Page Please drop us a message on our official Instagram page, @Chicanva.official. Let's embark on a journey of empowerment and style together, creating extraordinary creations that beautifully captivate. Cheers!
4	Woof woof pets Lovers 😻 We are Looking for pets from all over the world to represent our brand @ScoobyCats.official And become the new face for it. Get advantage of unlimited offers and 3 Free accessories for your Pawtner now ✨ Dm us now At @scoobycats.official to claim your spot 😻 👉🏻 @scoobycats.official

ข้อความที่	ข้อความ
	👈🏻
5	Hello, can you please send a message to @hppackofficial 🐶 they want to collaborate with you! We're looking for a few photogenic pets that can wear our products and provide us with some photos. You will be getting FREE exclusive products as well as a featuring on their official page! @hppackofficial 🐶 Just message us @hppackofficial now 😍 Can't wait to hear from you!! Tell them Jasmine invited you ❤️ @hppackofficial
6	Hi there 👋. We would love to invite you to join our world of Petsofworld ambassadors, no matter how small or big your social following is! If you're interested, we can send you 3 to 4 items of your choice for FREE. 🎁 Please send a DM "Model " to @Petsofworld.official We're excited for you to become a Petsofworld Ambassador! 🐶 We look forward to hearing from you! 😽 👉🏻 @Petsofworld.official 👈🏻
7	Hello! I work for @epnsupplements and we'd love to bring you on board as an official athlete. This would include free products, weekly pay, exposure through our Instagram account, etc. We reached out because we love the content you're putting out. We found your post on our explore page. Please DON'T answer here, click the link in @epnsupplements bio and fill out the Athlete Application.
8	Hello I have seen all your jewelry products. There are many beautiful jewelry on your website. I think you are not selling very well. I think you need a digital marketer. If you spend some money on your page and add an ad about your jewelry, then you can sell a lot. Many people will come to your website to buy jewelery from you. You can sell a lot of jewelry. I am a digital marketer, I can run you ads on

ข้อความที่	ข้อความ
	Facebook at a very low cost. If you spend some money, you can make your business known to everyone and you can sell your products in a good amount in the future. Contact Me- Whatsapp:01797814883
9	Hello Sarah, can you please send a message to @maisonpoirier 🦋 they want to collaborate with you ! You will be getting free exclusive products as well as a featuring on their official page ! 💕
10	We've got some thrilling news to share! 🚀 @dulifu_jewelry are searching for vibrant souls to join our exclusive Ambassador Team! 🌈 If you resonate with our vibe and want to be a part of an incredible journey, simply shoot us a DM with the magic word "Ambassador." to @dulifu_jewelry 📩✨ Let's create some unforgettable moments together! Excited to connect with each one of you.

ตารางที่ ก-4 ข้อความแชทจริง แบบฟิชชิ่งประเภทแนบไฟล์หรือลิงค์อันตราย

ข้อความที่	ข้อความ
1	Hi, I want to buy 6 Jewelry Sets as a gift. I have prepared a Jewelry order including: Model + Photo + Quantity + Delivery Address in the File below. Please check my order list and quote for me. Here is my order file: Order List Jewelry - Model - Picture TM-05.rar, Password - Open File : 123456. Open File with PC Computer or Laptop. Phone can't see.
2	Dear Customer, you have a message from CARS24. Kindly click the link below to see the message. http://m.cars.sh/OzGdQ9lxqvR
3	There is an update on your parcel. Item stopped due to unpaid custom fee. Follow the instructions here: http://aaa.bb.ccc
4	We had to limited some features, We require you to upload your ID and selfie via the link to continuw using your app: monzo-uk-help.com
5	Dear Customer, Thank you for the transaction done today at SBI 02264 branch. Plz click on the given link to share your experience. Https://crcf.sbi.co.in/ccf/home/GetFeedback?TxnDate=290323&TxnType=001060&Jno=16644791321. The feedback may be provided before 8 am tomorrow. No Personal Information would be captured. SBI
6	Your statement is ready for you to view online. Go to http://goo.gl/ak9t3k to view and manage your account.
7	We Were not able to complete your last payment for your Netflix Premium Membership. We Will Charge you again in the next couple of the days. If we are not able to complete the payment soon. We are to help if you need it. Visit the help center for more infor or contact us at https://www.netftfix.com/NMdL68l
8	👋 I've just opened membership to my exclusive traffic growth community. This is open for 24 hours ONLY. Click the link below for more information.
9	👋 I've just launched something new that you might be interested in. It's

ข้อความที่	ข้อความ
	a podcast completely focused on explaining the complex concepts behind blockchain technology in a language that anyone can understand. If you have an interest in cryptocurrencies, or are completely new to the space, I'd recommend giving the first 3 episodes a listen. Looking forward to hearing what you think 🙌
10	Hi Nique! See the attached invoice for your bathroom repair. Your payment is due 5/5. You can pay to the address on the invoice or online.

ตารางที่ ก-5 ข้อความแชทจริง แบบฟิชชิ่งประเภทโฆษณา และข้อความอื่นๆ ที่ไม่ต้องการ

ข้อความที่	ข้อความ
1	Are you interested in improving your IG page appearance by adding past 10K-100k+ of followers instantly? 🚀 🔥 HQ Followers (Lifetime Guaranteed): 1K - $13 10K - $95 20K - $165 50K - $300 100K - $520 ✅ Test Before Pay! DM "TRIAL" to our main page @avianoengage to test our services! 🤩
2	My name is Tabassum shaheen and My specialty is Etsy SEO. I've helped a lot of people rank on Etsy through my services. It is noted in your store's sales report that you sell 140 products with 188 total items, which could improve. You have many problems with your store; the title needs to be optimized, and you target competitive tags that affect your ranking. Let me assist you in optimizing
3	Dear Sir/Madam, Hope everything goes well with you. We have Natural Shazar Dendritic Agate Stone. Shazar Stone is obtained from Ken river which flows in the west of Banda district Bundelkhand region India. Shajar stones are used in the work of installing jewellery and decorative Home and gift items and other items like wall hangings. So if you are interested in buying Shazar Stone then please reply our email or also in messenger. Regard Anjali sarahstone@gmail.com https://www.sarahstone.com/

ข้อความที่	ข้อความ
4	My Name is Sarah I am city manager in Hong Kong and Bangkok for Beauty Pass App I am reaching out to as we would love to have your venue on Beauty Pass! Beauty Pass connects signed fashion models from the top agencies to exclusive venues like yours in exchange for social media posts. We are currently in Los Angeles, Miami, New York, Milan, Paris , London, Hong Kong , Istanbul , Bangkok, Singapore, Bali . Our service is FREE for you and FREE for the models. All you have to do is offer one of your services for free to the photoshoot models for your brand. They book a reservation through our app and then redeem & post about their experience. You choose when, how often and for how many. You are in control. My Whatsapp number +8 5 2 9 7 0 4 7 5 4 7 www.instagram.com/beautypass www.facebook.com/beautypassnet www.beautypass.net @beautypass Best regards Zlata Thank you!!
5	Hi I have an Hermes fake factory and a jewelry factory in China.We make best quality hand sewn Hermes bags and real gold and diamonds in the same luxury jewelry.If you are interested,please contact me on 📞 WhatsApp: +852 64396814 📷 Instagram: @hermes.posionq 🌐 Linktree: https://linktr.ee/birkin_angel
6	@dasalech6 1 I suggest we work together to make money.wa+85261489740.@dasalesch61
7	Hello, we have high-quality watches from various major brands,Follow our IG or whatsapp, product content will be updated regularly,Fast delivery,favorable price, no shipping fee, tax exemption,You can contact us directly via whatsapp for faster reply @vagabondeditorx ,whatsapp:+85291776418
8	Hello, we have watches of all major brands, we pay attention to product quality and service, only sell the highest quality, can be sent to the world and fast shipping, tax-free, if you are interested, please contact us @vagabondeditorx ,whatsapp: +85291776418
9	Hello, Hope you are in good health. We have scopus indexed journals

ข้อความที่	ข้อความ
	with low APC. If you have any unpublished articles then kindly let me know.
10	Hello, I'm Elijah verna from Oracle INDIA Communications, Can I have a minute of your time? This is a part time job that doesn't affect your other job, and you are not required to pay any joining fee. The job is very simple, All you have to do is like the Youtube videos and send screenshots, and you will be paid. We pay 50 rupees for each task and you can earn up to 2500 rupees a day by liking videos in your free time.

ตารางที่ ก-6 ข้อความแชทจริง ที่ถูกต้อง ไม่อันตราย (Legitimate)

ข้อความที่	ข้อความ
1	I love that Poseidon ring so much, but not available in my size. So I ordered one, and will happily wait two months for it 😊
2	Hi It's now just over two months since I ordered two rings, made to size. But I still haven't heard from your shop... 😕
3	Thank you so much -- I will collect it tomorrow (I live just one Skytrain station away). I didn't receive an email message, by the way, and also provided my mobile number to the sales lady. Anyway, I'm just happy that I can collect tomorrow. 😊 🙏
4	The email address I provided was sarah@yahoo.com -- the one used was not correct.
5	Hi I'm at your Siam Square shop and want to order a ring to size... But I don't want the option of additional interchangeable pieces...
6	Hello, can I get more info on this?
7	China pay can not be used, is there any other payment method
8	My ring just broke 😵 Is it fixable?
9	The gold part in the middle is a ring shape and its moveable, the that gold part broken in between
10	Okay And I send it to you by mail first?

ข้อความที่	ข้อความ
11	I am not in Thailand till the end of this year I probably send it by mail to you first
12	Thank you very much, I appreciate it 😊 😊 😊
13	Thanks 😊 Hope to know the expected time that need 🙏
14	Thanks for asking!
15	That's fantastic 🙏
16	Thanks for your help!
17	See you soon
18	I will drop by tomorrow 😊
19	Can I see more products?
20	Hello kaa How should I send you back the ring kaa
21	Do you mind if I get a smaller size So I can send you back the ring from Phuket as a parcel to your flagship store?
22	Can I please have your address kaa
23	Okay thank you kaa
24	That's fine kaa thank you
25	Do you want me to send the store bag as well kaa or just the box with the ring and the card ?
26	Should I send the receipt as well ?
27	Can I have in size 9 please
28	Have you received the ring yet kaa
29	How much is it and where can I buy it please?
30	Hello there Do you still carrying this one kaab? Im looking for red jewels
31	Can I buy it online?
32	Is this available in my size?
33	Hello Do u ship worldwide?
34	Is the gold bracelet available ? Bracket with all pendant same as the picture

ข้อความที่	ข้อความ
35	How much in total ?
36	Total 4 pendant and one bracelet correct?
37	Free shipping to Hong Kong ?
38	Any size for the bracelet ?
39	This one is nice
40	For the bracelet, do u have a brand new one ? not the display one
41	Can u try on it ?
42	Is the bracelet in rose golf color ? Rose gold
43	Do u think it is too big ?
44	Ok I want bracelet and all pendant in rose gold color,stopper in gold color
45	How to settle the payment ?
46	I am not able to open the link:(Or do you accept western union ?
47	Please send me the new piece of every item Thank u ❤
48	How much this one?
49	Can u send me more clear Pic of the beads? or can u design similar to the pic ?
50	I want this one To replace one of the middle one And can have more pink beads instead if the brown beads?

ภาคผนวก ข
ข้อมูลอื่นๆ

โค้ดสคริปต์ Visual Basic for Application (VBA) สำหรับลบการขึ้นบรรทัดใหม่ใน Excel เพื่อเตรียมข้อมูลในรูปไฟล์ CSV สำหรับนำไปใช้เทรนโมเดลต่อได้ เป็นดังต่อไปนี้:

```
Sub RemoveLineBreaks()
    Dim rng As Range
    Dim cell As Range
    On Error Resume Next
    Set rng = Selection
    For Each cell In rng
        If Not IsEmpty(cell) Then
            cell.Value = Replace(cell.Value, Chr(10), " ")
            cell.Value = Replace(cell.Value, Chr(13), " ")
        End If
    Next cell
End Sub
```

โค้ดสคริปต์ Visual Basic for Application (VBA) สำหรับการเปลี่ยนอักขระแบ่งฟิลด์ข้อมูล (Delimiter) จาก , เป็น | ใน Excel เพื่อเตรียมข้อมูลในรูปไฟล์ CSV สำหรับนำไปใช้เทรนโมเดลต่อได้ เป็นดังต่อไปนี้:

```
Sub ExportWithCustomDelimiter()
    Dim MyFileName As String
    Dim MyDelimiter As String
    Dim rng As Range
    Dim i As Long, j As Long
    Dim MyData As String

    MyFileName = Application.GetSaveAsFilename(InitialFileName:="",
```

```
FileFilter:="Text Files (*.txt), *.txt")
        If MyFileName = "False" Then Exit Sub 'User canceled

        MyDelimiter = "|" 'Set your custom delimiter here

        Set rng = ActiveSheet.UsedRange
        Open MyFileName For Output As #1
        For i = 1 To rng.Rows.Count
            MyData = ""
            For j = 1 To rng.Columns.Count
                MyData = MyData & rng.Cells(i, j).Value
                If j < rng.Columns.Count Then
                    MyData = MyData & MyDelimiter
                End If
            Next j
            Print #1, MyData
        Next i
        Close #1
        MsgBox "Data exported successfully!"
    End Sub
```

ประวัติผู้เขียน

ชื่อ นายศรัณย์ หงสกุล

ชื่อการค้นคว้าอิสระ การสร้างชุดข้อมูลข้อความแชทฟิชชิ่งด้วยโมเดลการเรียนรู้แบบฟิวช็อต

สาขาวิชา การบริหารเครือข่ายและความมั่นคงปลอดภัยสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ประวัติ เกิดเมื่อวันที่ 10 มิถุนายน 2528 ปัจจุบันอายุ 39 ปี โดยเกิดที่จังหวัดกรุงเทพมหานคร เข้าศึกษาระดับปริญญาตรี สาขาวิทยาศาสตร์และเทคโนโลยีการอาหาร คณะอุตสาหกรรมเกษตร ที่มหาวิทยาลัยเกษตรศาสตร์จนสำเร็จการศึกษาในปีการศึกษา 2550 และเข้าศึกษาต่อระดับปริญญาโท สาขาการบริหารเครือข่ายดิจิทัลและความมั่นคงปลอดภัยสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

ในด้านการทำงาน มีประสบการณ์ดังต่อไปนี้

- พ.ศ. 2550 – 2552 เจ้าหน้าที่วางแผนและประสานงานผลิต บริษัท พรีเซิร์ฟฟู้ด สเปเชียลตี้ จำกัด
- พ.ศ. 2552 – 2554 วิศวกรเครือข่าย และผู้สอนวิชาเครือข่าย CCNA/CCNP บริษัท ราเน็ท จำกัด
- พ.ศ. 2554 – 2555 ผู้จัดการด้านข้อมูลความสัมพันธ์ลูกค้า (CRM) ทางอิเล็กทรอนิกส์ บริษัท เรนโบว์ ซิลเวอร์ จำกัด
- พ.ศ. 2555 ถึงปัจจุบัน ผู้จัดการด้านไอที ผู้จัดการแบรนด์เครื่องประดับ AKE AKE บริษัท เรนโบว์ ซิลเวอร์ จำกัด