



สถาบันไอทีจีเนียส
IT GENIUS INSTITUTE

COURSE ID
PRC-37

PYTHON / NLP

COURSE OUTLINE · รายละเอียดหลักสูตรอบรม

NLP and Thai Text Analytics with LLMs

วิเคราะห์ข้อความภาษาไทยจากรีวิวและโซเชียลด้วย PyThaiNLP และ LLM

● 12 ชั่วโมง

● 2 วัน

● พื้นฐาน Python และ pandas

● ระดับ Intermediate



ระยะเวลา: 2 วัน

12 ชั่วโมง (2 วัน วันละ 6 ชม.) · 09:00-16:00 น.



รูปแบบ: Classroom / Hybrid

Onsite หรือ Online ผ่าน Microsoft Teams · Public / In-house



รายละเอียดหลักสูตร

www.itgenius.co.th

ข้อความจากลูกค้าเป็นข้อมูลที่มีค่ามากแต่ถูกใช้น้อยที่สุด ทั้งรีวิวสินค้า ความคิดเห็นบนโซเชียล แชตกับแอดมิน และแบบสอบถามปลายเปิด ข้อความภาษาไทยวิเคราะห์ยากกว่าภาษาอังกฤษ เพราะไม่มีการเว้นวรรคระหว่างคำ มีคำสแลง คำสะกดผิด อีโมจิ และการพิมพ์ซ้ำอย่าง 555 หลายองค์กรจึงยังอ่านที่ละข้อความ หรือนับแค่จำนวนดาว ทำให้มองไม่เห็นว่าคุณค่าพอใจหรือไม่พอใจเรื่องอะไรกันแน่

หลักสูตรนี้พาวิเคราะห์ข้อความภาษาไทยด้วย Python ตั้งแต่พื้นฐานจนถึงการใช้ LLM เริ่มจากตัดคำและทำความสะอาดข้อความด้วย PyThaiNLP สรุปคำสำคัญด้วย TF-IDF สร้างโมเดล Sentiment แบบดั้งเดิมด้วย scikit-learn และใช้ Embedding กับโมเดลภาษาไทยบน Hugging Face จากนั้นค้นหาคำหัวข้อที่ลูกค้าพูดถึงด้วย Topic Modeling ใช้ LLM จัดหมวด วิเคราะห์ความรู้สึกรายประเด็น และดึงข้อมูลออกมาเป็น JSON พร้อมวัดความแม่นยำและคุมต้นทุน ก่อนสรุปเป็นรายงาน Insight และปิดท้ายด้วยโปรเจกต์วิเคราะห์รีวิวตั้งแต่ต้นจนจบ (อบรม 2 วัน วันละ 6 ชั่วโมง รวม 12 ชั่วโมง ระดับ Intermediate)

เครื่องมือ

Py Python

TN PyThaiNLP

HF Hugging Face

GPT LLM

Co Google Colab

ราคาค่าอบรม · PUBLIC / IN-HOUSE

7,500 บาท / ท่าน

ราคายังไม่รวมภาษีมูลค่าเพิ่ม (VAT 7%) · รวมเอกสารประกอบการอบรม Lab Guide Notebook ควบทุก Lab ชุดข้อความตัวอย่าง และ Certificate · รับจัด In-house ขอใบเสนอราคาได้

สมัครผ่านเว็บไซต์ ลด 10%



เพิ่มเพื่อน
LINE

☎ 02-570-8449

☎ 088-807-9770

LINE @itgenius

🌐 www.itgenius.co.th

01 วัตถุประสงค์ Objectives

- เข้าใจความท้าทายของ NLP ภาษาไทย และเลือกวิธีวิเคราะห์ให้เหมาะกับโจทย์
- ตัดคำ ทำความสะอาด และ Normalize ข้อความภาษาไทยด้วย PyThaiNLP
- สรุปคำสำคัญและเปรียบเทียบกลุ่มข้อความด้วย TF-IDF
- สร้างโมเดลจำแนก Sentiment ด้วย scikit-learn และโมเดลภาษาไทยบน Hugging Face
- ค้นหาหัวข้อที่ลูกค้าพูดถึงด้วย Embedding และ Topic Modeling
- ใช้ LLM จัดหมวด วิเคราะห์ความรู้สึกรายประเด็น และดึงข้อมูลเป็น JSON
- วัดความแม่นยำและเปรียบเทียบต้นทุนระหว่างโมเดลดั้งเดิมกับ LLM
- สรุปผลการวิเคราะห์เป็นรายงาน Insight ที่ฝ่ายธุรกิจนำไปใช้ได้

02 กลุ่มเป้าหมาย Target Audience

- Data Analyst และ Data Scientist ที่ต้องวิเคราะห์ข้อความภาษาไทย
- ทีม Marketing และ Social Listening ที่ต้องติดตามความคิดเห็นของลูกค้า
- ทีม Customer Experience และ Call Center ที่มีข้อความร้องเรียนหรือแชตจำนวนมาก
- Python Developer ที่ต้องการสร้างระบบจัดหมวดข้อความอัตโนมัติ
- นักวิจัยและนักศึกษาที่ทำงานด้านข้อความภาษาไทย

03 จุดเด่นของหลักสูตร Highlights

- **เน้นภาษาไทยจริง** สแลง อีโมจิ และคำสละสลวย
- **ดั้งเดิมคู่กับ LLM** PyThaiNLP, scikit-learn และ LLM
- **วิเคราะห์รายประเด็น** รู้ว่าลูกค้าพอใจหรือไม่พอใจเรื่องอะไร
- **วัดผลและคุมต้นทุน** เลือกวิธีที่คุ้มค่างาน
- **Lab บน Google Colab** ใช้ชุดข้อความตัวอย่างได้ทันที

04 ความรู้พื้นฐาน Prerequisites

- เขียน Python ขึ้นพื้นฐานได้ และใช้ pandas จัดการตารางข้อมูลเบื้องต้นได้
- เข้าใจแนวคิด Machine Learning เบื้องต้นจะช่วยให้เรียนได้เร็วขึ้น แต่ไม่บังคับ
- มีบัญชี Google สำหรับ Google Colab
- มี API Key ของ LLM อย่างน้อยหนึ่งราย เช่น Gemini, OpenAI หรือ Claude

05 สิ่งที่ต้องเตรียมและเครื่องมือ Setup & Tools

รายการ	รายละเอียด
อุปกรณ์	Notebook เชื่อมต่ออินเทอร์เน็ต พร้อม Web Browser รุ่นล่าสุด
สภาพแวดล้อม Lab	Google Colab ด้วยบัญชี Google ของผู้เรียน
Library	PyThaiNLP, pandas, scikit-learn, Hugging Face Transformers และ BERTopic
LLM	Gemini API, OpenAI API หรือ Claude API ด้วย API Key ของผู้เรียนเอง
ข้อมูลตัวอย่าง	ชุดรีวิวและความคิดเห็นภาษาไทยแบบเปิด และชุดข้อความของบริษัทสมมุติ

หมายเหตุ: Lab ทำบน Google Colab ด้วยบัญชีของผู้เรียน ใช้ชุดข้อความแบบเปิดและข้อมูลของบริษัทสมมุติ ไม่ใช่ข้อมูลลูกค้าจริง การเรียก LLM API อาจมีค่าใช้จ่ายตามการใช้งานจริง (บางรายมีโควตาใช้ฟรีให้ทดลอง) ผู้เรียนใช้ API Key ของตนเองหรือของบริษัท หากนำข้อความของลูกค้าจริงไปวิเคราะห์ ควรตรวจเงื่อนไขของแพลตฟอร์มต้นทางและ PDPA ก่อนเสมอ ชื่อโมเดลและความสามารถของ LLM เปลี่ยนเร็ว เนื้อหาจะอัปเดตตามเวอร์ชันล่าสุด ณ วันอบรม

06 ภาพรวมการเรียนรู้ 2 วัน Big Picture

วัน	หัวข้อหลัก
Day 1 Section 1-6	เข้าใจความยากของข้อความภาษาไทย, ตัดคำและทำความสะอาดข้อความด้วย PyThaiNLP, สรุปคำสำคัญด้วย TF-IDF, สร้างโมเดล Sentiment แบบดั้งเดิม และใช้ Embedding กับโมเดลภาษาไทยบน Hugging Face
Day 2 Section 7-12	ค้นหาหัวข้อด้วย Topic Modeling, ใช้ LLM จัดหมวดหมู่และวิเคราะห์ความรู้สึกรายประเด็น, ดึงข้อมูลเป็น JSON, วัดผลและคุมต้นทุน, ทำรายงาน Insight และปิดท้ายด้วยโปรเจกต์วิเคราะห์รีวิว

DAY
1

พื้นฐาน NLP ภาษาไทยด้วย PyThaiNLP

เนื้อหาการอบรม · Section 1-6

รวม
6 ชั่วโมง

1 ภาพรวม Text Analytics ภาษาไทย

- งานวิเคราะห์ข้อความในธุรกิจ: รีวิว โซเชียล แชต และแบบสอบถามปลายเปิด
- ทำไมภาษาไทยยาก: ไม่เว้นวรรค สแลง คำสะกดผิด และอีโมจิ
- เลือกวิธีให้เหมาะ: กฎ, Machine Learning, โมเดลภาษา หรือ LLM
- Lab: เตรียม Google Colab และโหลดชุดรีวิวภาษาไทยตัวอย่าง

2 Lab: ตัดคำและประมวลผลด้วย PyThaiNLP

- ตัดคำด้วย Engine ต่าง ๆ และเปรียบเทียบผลลัพธ์
- ตัดประโยค ระบุชนิดของคำ และ Normalize ข้อความ
- เพิ่มคำศัพท์เฉพาะธุรกิจ เช่น ชื่อสินค้าและชื่อสาขา ลงพจนานุกรม
- Lab: ตัดคำรีวิวทั้งหมดแล้วนับคำที่พบบ่อย

3 ทำความสะอาดข้อความจากโซเชียล

- จัดการ URL, Hashtag, Mention และอีโมจิ
- ตัดตัวอักษรซ้ำ เช่น มากกกก และแปลง 555 ให้มีความหมาย
- Stopword ภาษาไทยและเมื่อไรไม่ควรตัดทิ้ง
- Lab: สร้างฟังก์ชันทำความสะอาดข้อความที่ใช้ซ้ำได้

4 Lab: คำสำคัญและการเปรียบเทียบด้วย TF-IDF

- Bag-of-Words, N-gram และ TF-IDF กับข้อความภาษาไทย
- หาคำที่โดดเด่นของรีวิวเชิงบวกเทียบกับเชิงลบ
- Word Cloud ภาษาไทยที่แสดงพ้องคำได้ถูกต้อง
- Lab: เปรียบเทียบคำสำคัญของแต่ละสาขาหรือแต่ละสินค้า

5 Lab: โมเดล Sentiment แบบดั้งเดิม

- เตรียม Label และแบ่งข้อมูล Train กับ Test
- TF-IDF ร่วมกับ Logistic Regression ด้วย scikit-learn
- วัดผลด้วย Precision, Recall, F1 และ Confusion Matrix
- ดูข้อความที่โมเดลทายผิดเพื่อหาสาเหตุ
- Lab: สร้างโมเดลจำแนกรีวิวบวก กลาง และลบ

6 Embedding และโมเดลภาษาไทยบน Hugging Face

- แนวคิด Embedding และ Transformer แบบเข้าใจง่าย
- โมเดลภาษาไทยอย่าง WangchanBERTa และโมเดล Embedding หลายภาษา
- ใช้ Pipeline ของ Hugging Face จำแนกข้อความโดยไม่ต้อง Train เอง
- Lab: ค้นหารีวิวที่ความหมายใกล้กันด้วย Embedding



DAY
2

วิเคราะห์ข้อความเชิงลึกด้วย LLM

เนื้อหาการอบรม · Section 7-12

รวม
6 ชั่วโมง

7 Lab: ค้นหาหัวข้อด้วย Topic Modeling

- จัดกลุ่มข้อความด้วย Embedding และ Clustering
- ใช้ BERTopic ร่วมกับตัวตัดคำภาษาไทย
- ให้ LLM ตั้งชื่อและสรุปแต่ละหัวข้อเป็นภาษาไทย
- Lab: หาว่าลูกค้าพูดถึงเรื่องอะไรมากที่สุดในแต่ละเดือน

8 LLM สำหรับจัดหมวดและ Sentiment

- Zero-shot และ Few-shot Classification ด้วย Prompt
- Aspect-based Sentiment: แยกความรู้สึกต่อราคา บริการ และสินค้า
- จัดการข้อความประชิดประชันและข้อความที่มีหลายประเด็น
- Lab: จัดหมวดเรื่องร้องเรียนตามหมวดที่ธุรกิจกำหนด

9 Lab: ดึงข้อมูลเป็น JSON ด้วย LLM

- Structured Output ตาม Schema ที่กำหนด
- ดึงชื่อสินค้า สาขา ปัญหา และความเร่งด่วนจากข้อความ
- ตรวจสอบและปิดบังข้อมูลส่วนบุคคลก่อนส่งให้ LLM
- ประมวลผลข้อความจำนวนมากด้วย Batch และจัดการ Error
- Lab: แปลงแชตลูกค้าเป็นตารางพร้อมวิเคราะห์

10 วัดผลและคุมต้นทุน

- สร้างชุดข้อความที่คน Label ไว้เป็นมาตรฐานอ้างอิง
- เปรียบเทียบโมเดลดั้งเดิม โมเดลภาษาไทย และ LLM ด้วยตัวเลขเดียวกัน
- คำนวณต้นทุนต่อพันข้อความ และลดต้นทุนด้วยการคัดกรองก่อน
- Lab: ตารางเปรียบเทียบความแม่นยำ ความเร็ว และต้นทุน

11 จาก Insight สู่รายงานธุรกิจ

- ติดตามแนวโน้มความรู้สึกและหัวข้อตามช่วงเวลา
- ให้ LLM สรุปประเด็นสำคัญพร้อมยกตัวอย่างข้อความจริง
- ส่งออกผลเป็น CSV หรือ Google Sheets เพื่อทำ Dashboard ต่อ
- Lab: รายงาน Voice of Customer หนึ่งหน้าสำหรับผู้บริหาร

12 Workshop: Capstone วิเคราะห์รีวิวครบวงจร

- เลือกชุดข้อมูลตัวอย่างหรือข้อความของตนเองที่ไม่มีข้อมูลส่วนบุคคล
- ทำครบตั้งแต่ทำความสะอาด จัดหมวด Sentiment จนถึงหัวข้อ
- วัดความแม่นยำและสรุปต้นทุนของวิธีที่เลือก
- นำเสนอ Insight และวิธีร่วมกัน

07 เรียนจบแล้วได้อะไร Outcomes

- ประมวลผลภาษาไทยได้ ตัดคำและทำความสะอาดด้วย PyThaiNLP
- สร้างโมเดล **Sentiment** scikit-learn และโมเดลภาษาไทย
- หาหัวข้อจากข้อความ Embedding และ Topic Modeling
- ใช้ LLM วิเคราะห์ลึก จัดหมวด รายประเด็น และดึง JSON
- เลือกวิธีอย่างมีหลัก วัดความแม่นยำและต้นทุน
- ส่งต่อ Insight ได้ รายงาน Voice of Customer

08 สิ่งที่คุณเรียนจะได้รับ Deliverables

- เอกสารประกอบการอบรมฉบับสมบูรณ์ครบทั้ง 2 วัน และ Lab Guide ที่ละเอียด
- Google Colab Notebook ครบทุก Lab พร้อมฟังก์ชันทำความสะอาดข้อความภาษาไทยที่ใช้ซ้ำได้
- ชุดข้อความตัวอย่าง Template ของ Prompt สำหรับจัดหมวดและดึงข้อมูล และ Template รายงาน Voice of Customer
- ประกาศนียบัตรจบหลักสูตร (Certificate of Completion) และช่องทางสอบถามวิทยากรหลังจบหลักสูตร

09 ข้อมูลผู้สอน Instructor

อ.สาธิต ไทยม ผู้สอนที่ทำงานวิเคราะห์ข้อความภาษาไทยจากรีวิว โซเชียล และงานบริการลูกค้าในงานจริง จุดเด่นคือการสอนให้เห็นทั้งเครื่องมือแบบดั้งเดิมและ LLM แล้วเลือกใช้ตามความแม่นยำและต้นทุนของงาน ใช้ข้อความภาษาไทยที่มีสแลงและคำสแลงที่แบบที่เจอจริง และพาเปลี่ยนผลลัพธ์ทางเทคนิคให้เป็น Insight ที่ฝ่ายธุรกิจนำไปตัดสินใจได้

ราคาค่าอบรม 7,500 บาท/ท่าน · สนใจ IN-HOUSE ขอใบเสนอราคา

7,500 บาท / ท่าน (ยังไม่รวม VAT 7%)

สมัครผ่านเว็บไซต์ ลด 10%

☎ 02-570-8449

☎ 088-807-9770

📞 @itgenius

✉ contact@itgenius.co.th



เพิ่มเพื่อน
LINE