

2023

เทคนิคการเรียกคืนระดับตัวอักษรของศิลาจารึกสุโขทัยด้วยโมเดลภาษาแบบจุดเปิด

สุจิตรา ทองขำ  
คณะวิศวกรรมศาสตร์

Follow this and additional works at: <https://digital.car.chula.ac.th/chulaetd>



Part of the [Computer Sciences Commons](#)

---

### Recommended Citation

ทองขำ, สุจิตรา, "เทคนิคการเรียกคืนระดับตัวอักษรของศิลาจารึกสุโขทัยด้วยโมเดลภาษาแบบจุดเปิด" (2023). *Chulalongkorn University Theses and Dissertations (Chula ETD)*. 10281.  
<https://digital.car.chula.ac.th/chulaetd/10281>

This Thesis is brought to you for free and open access by Chula Digital Collections. It has been accepted for inclusion in Chulalongkorn University Theses and Dissertations (Chula ETD) by an authorized administrator of Chula Digital Collections. For more information, please contact [ChulaDC@car.chula.ac.th](mailto:ChulaDC@car.chula.ac.th).

เทคนิคการเรียกคืนระดับตัวอักษรของศิลาจารึกสุโขทัยด้วยโมเดลภาษาแบบถูกปิด



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2566

A CHARACTER-LEVEL RESTORATION OF SUKHOTHAI INSCRIPTIONS USING THE MASKED  
LANGUAGE MODEL



A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Science in Computer Science  
Department of Computer Engineering  
Faculty Of Engineering  
Chulalongkorn University  
Academic Year 2023

หัวข้อวิทยานิพนธ์	เทคนิคการเรียกคืนระดับตัวอักษรของศิลาจารึกสุโขทัยด้วยโมเดลภาษาแบบถูกปิด
โดย	น.ส.สุจิตรา ทองขำ
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.สุกรี สิ้นธุฎิณโณ

---

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

.....	คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)	
คณะกรรมการสอบวิทยานิพนธ์	
.....	ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.นัทธี นิภานันท์)	
.....	อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สิ้นธุฎิณโณ)	
.....	กรรมการ
(รองศาสตราจารย์ ดร.ณัฐพงศ์ ชินธเนศ)	
.....	กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.เดวิด ประดับสุวรรณ)	

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

สุจิตรา ทองขำ : เทคนิคการเรียกคืนระดับตัวอักษรของศิลาจารึกสุโขทัยด้วยโมเดลภาษาแบบถูกปิด. ( A CHARACTER-LEVEL RESTORATION OF SUKHOTHAI INSCRIPTIONS USING THE MASKED LANGUAGE MODEL) อ.ที่ปรึกษาหลัก : ผศ. ดร.สุกรี สิ้นธุภิณฺญ

ศิลาจารึกจัดเป็นหนึ่งในวรรณกรรมชนิดลายลักษณ์อักษรที่บันทึกเรื่องราวทางประวัติศาสตร์และแสดงเอกลักษณ์ของวัฒนธรรมในช่วงเวลานั้น ๆ ผ่านวิธีการสลักตัวอักษรลงบนแผ่นหินด้วยวัสดุโลหะแหลมคมที่ละตัวอักษรจนกลายเป็นประโยคที่สามารถสื่อใจความให้ผู้อ่านสามารถเข้าใจความหมายดังนั้นความสมบูรณ์ของประโยคนั้นมีความสำคัญเป็นอย่างมากในงานด้านการประมวลผลภาษาธรรมชาติ โดยเฉพาะอย่างยิ่งเมื่อมีการถอดความศิลาจารึกและพบว่าศิลาจารึกบางส่วนไม่สามารถถอดความได้อันเป็นเหตุเนื่องมาจากช่วงของระยะเวลาผ่านไปจารึกเหล่านั้นอาจเกิดการเสื่อมสภาพจากหลายสาเหตุจนเป็นรอยขีดข่วนทับข้อความหรือตัวอักษรจนเลือนลาง, ถูกทำลายจากภัยธรรมชาติจนไม่สามารถวิเคราะห์ได้ว่าตัวอักษรที่ชำรุดคือตัวอักษรใด ๆ เพื่อเพิ่มความสมบูรณ์ให้กับประโยคที่ขาดหายไปงานวิจัยนี้จึงใช้วิธีการสร้างโมเดลทำนายอักขระของตัวอักษรที่ขาดหายไปจากข้อความโดยใช้เทคนิคโมเดลภาษาแบบมาส์กเข้ามาช่วยในการประมวลผลการทดลองโดยใช้เป็นพรีเทรนโมเดลหลากหลายภาษา (multilingual pre-train model) 3 ประเภทดังต่อไปนี้ (1) โมเดลเอ็กซ์แอลเอ็ม-โรเบอตา (2) โมเดลเบิร์ต-เบส-หลายภาษา-เคสดี (3) โมเดลดีสทิลเบิร์ต-เบส-หลายภาษา-เคสดี และการใช้โมเดลในกลุ่มแบบจำลองโครงข่ายประสาทเทียมแบบลึกแบบเรียนรู้สองทิศทางได้แก่ (4) โมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง โดยในแต่ละรอบของการเทรนจะทำการสุ่มปิดอักขระด้วยมาส์กโทเคน “ ” หรือ “[MASK]” ให้โมเดลทำนายค่าที่ขาดหายไปตรงจุดที่ถูกมาส์กโทเคนและจากผลการทดลองพบว่าความถูกต้องของการทำนายจากพรีเทรนโมเดลทั้ง 4 ประเภท คือ (1) 42 % (2) 53 % (3) 50 % และ (4) 36 % ตามลำดับในงานวิจัยฉบับนี้ได้นำเสนอวิธีการปรับปรุงความสามารถในการทำนายของโมเดลโดยใช้คลังคำศัพท์เฉพาะทำให้โมเดลมีประสิทธิภาพในการทำนายความถูกต้องได้แม่นยำมากขึ้น

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์  
ปีการศึกษา 2566

ลายมือชื่อนิสิต .....  
ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

# # 6472086421 : MAJOR COMPUTER SCIENCE

KEYWORD: Natural language processing, Bidirectional encoder representations from transformers (BERT), Masked language model

Sujitra Tongkhum : A CHARACTER-LEVEL RESTORATION OF SUKHOTHAI INSCRIPTIONS USING THE MASKED LANGUAGE MODEL. Advisor: Asst. Prof. DR. SUKREE SINTHUPINYO

The stone inscription is one type of written literature that recorded the history story and the manifestation of cultural identity in that era through a character engraving method on the stone with sharp metal material for each character until a sentence formed. To convey the message for the readers to understand the meaning. Therefore, the completeness of that sentence is of great importance natural language processing tasks. In particular, when transcription stone inscriptions, it is found that inscriptions' parts cannot interpret. As a result of the period that elapsed, those inscriptions may have suffered deterioration from various causes, resulting in scratches over the text or faded markings, destroyed from natural disasters that making it impossible to analyze which specific characters were damaged. To address enhance the completeness of the missing sentence, this research employs a method of generating predictive models for the missing characters from the text. It utilizes the technique of incorporating a masked language model to assist in processing the experimental data, utilizing 4 types of multilingual pre-trained models as following models are used: (1) XLM-RoBERTa, (2) Bert-base-multilingual-cased, (3) DistilBERT-base-multilingual-cased, and the use of models in the category of a bidirectional long short-term memory deep neural network, including (4) Bi-LSTM. In each training round, random characters are masked using the token "" or "[MASK]" to prompt the model to predict the missing words at the masked positions. From the experimental results, it was found that the accuracy of prediction from the three types of pre-trained models is as follows: (1) 42 %, (2) 53 %, (3) 50 % and 36 % respectively. In this research paper, a method for improving the predictive capabilities of the model is presented by utilizing a domain-specific dictionary. The models are more effective in making accurate predictions.

Field of Study: Computer Science

Student's Signature .....

Academic Year: 2023

Advisor's Signature .....

## กิตติกรรมประกาศ

การศึกษานี้สามารถสำเร็จลุล่วงได้ด้วยดีด้วยความช่วยเหลือจาก อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก ผู้ช่วยศาสตราจารย์ ดร. สุกรี สินธุภิญโญ ผู้มีส่วนเกี่ยวข้องในการชี้แนะแนวทางให้คำแนะนำทั้งด้านเทคนิคและแนวทางการศึกษาหาความรู้ ดั่งนั้นข้าพเจ้าขอขอบพระคุณที่ได้ให้ความช่วยเหลือในทุก ๆ ด้านตลอดระยะเวลาของการศึกษางานวิจัยฉบับนี้ ขอขอบคุณหน่วยงานฐานข้อมูลจารึกในประเทศไทยโดยศูนย์มานุษยวิทยาสิรินธรที่สร้างฐานข้อมูลเผยแพร่สู่สาธารณะ และขอขอบคุณภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่มอบความรู้ทางด้านวิทยาศาสตร์คอมพิวเตอร์เพื่อผลิตผลงานวิจัยและสร้างองค์ความรู้ใหม่ ๆ ตลอดระยะเวลาของการศึกษา

สุจิตรา ทองคำ





## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ค
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ฉ
สารบัญรูปภาพ.....	ฉ
บทที่ 1 บทนำ .....	1
1.1 ที่มาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	1
1.3 ขอบเขตการวิจัย .....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.5 แผนการดำเนินงาน.....	2
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง .....	3
ทฤษฎีที่เกี่ยวข้อง.....	3
2.1 จารีก.....	3
2.1.1 จารีกวัดป่ามะม่วง .....	3
2.1.2 จารีกนครชุม.....	4
2.1.3 จารีกพ่อขุนรามคำแหง .....	4
2.1.4 จารีกวัดศรีชุม.....	4
2.1.5 จารีกปู่ขุนจิตขุนจอต .....	4
2.2 การประมวลผลภาษาทางธรรมชาติ (Natural language processing, NLP) .....	4

2.3 โมเดลภาษาแบบถูกปิด (Masked Language Model, MLM) ..... 5

2.4 การทรานฟอร์มเมอร์ข้อมูลแบบสองทิศทางเชิงลึก (Deep Bidirectional Transformers)  
..... 5

    2.4.1 สถาปัตยกรรมโมเดล (Model Architecture) [10]..... 6

2.5 การปรับละเอียด (Fine tune)..... 9

2.6 การเรียนรู้เชิงลึก (Deep Learning) [12]..... 9

    2.6.1 ทฤษฎีความน่าจะเป็น (Probability Theory)..... 10

        2.6.1.1 การใช้กฎความน่าจะเป็น (Laws of Probability)..... 10

        2.6.1.2 การใช้ความน่าจะเป็นและสถิติวิเคราะห์พฤติกรรมของระบบที่ AI  
        นำเสนอผลงาน ..... 10

    2.6.2 ตัวแปรสุ่ม (A random variable) ..... 10

    2.6.3 การแจกแจงความน่าจะเป็น (Probability Distributions)..... 10

        2.6.3.1 ตัวแปรไม่ต่อเนื่องและฟังก์ชันมวลของความน่าจะเป็น (Discrete  
        Variables and Probability Mass Functions, PMF) ..... 10

        2.6.3.2 ตัวแปรต่อเนื่องและฟังก์ชันความหนาแน่นของความน่าจะเป็น  
        (Continuous Variables and Probability Density Functions,  
        PDF) ..... 11

        2.6.3.3 การแจกแจงแบบเอกซ์โปเนนเชียลและลาปลาซ (Exponential and  
        Laplace Distributions)..... 11

    2.6.4 กฎของเบย์ (Bayes' Rule) ..... 12

2.7 พื้นฐานการเรียนรู้ด้วยเครื่อง (Machine Learning Basics) [12] ..... 12

    2.7.1 การวัดประสิทธิภาพของโมเดลด้วยวิธีการตรวจสอบแบบข้าม (Measuring  
    model performance by cross validation)..... 12

    2.7.2 การประเมินภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation, MLE)  
    ..... 13

2.8 หน่วยความจำระยะสั้นระยะยาว (Long Short-Term Memory, LSTM)..... 13

งานวิจัยเกี่ยวข้อง .....	14
2.9 เบิร์ต: การเรียนรู้เชิงลึกล่วงหน้าของโมเดลทรานฟอร์เมอร์สำหรับเข้าใจภาษาในแง่ของ การเข้าใจความหมายทั้งแบบแยกตามทิศทางและแบบไม่แยกตามทิศทาง .....	14
2.10 ตัวแทนการเรียนรู้ข้ามภาษาแบบไม่มีการสอนขนาดใหญ่ (Unsupervised Cross- lingual Representation Learning at Scale) [18].....	15
2.11 วิธีการปรับละเอียด เบิร์ตสำหรับการจำแนกข้อความ (How to Fine-Tune BERT for Text Classification?) [11].....	16
2.12 หัวข้อในรูปแบบเบิร์ต : โมเดลภาษาประสาทที่ปรับปรุงตามหัวข้อซึ่งถูกปรับละเอียด สำหรับการจำแนกประเภทความรู้สึก (TopicBERT: A Topic-Enhanced Neural Language Model Fine-Tuned for Sentiment Classification) [19] .....	18
2.13 การศึกษาเปรียบเทียบการใช้การเรียนรู้เชิงลึกในการทำนายคำถัดไป (Next word prediction using deep learning: A comparative study)[15] .....	18
บทที่ 3 แนวคิดและวิธีการวิจัย.....	20
3.1 แนวทางการวิจัย .....	20
3.2 วิธีการวิจัย .....	21
3.2.1 การลงพื้นที่สำรวจหลักศิลาจารึก .....	21
3.2.2 ชุดข้อมูลภาษาโบราณในฐานข้อมูลจารึกแห่งประเทศไทยศูนย์มานุษยวิทยาสิรินธร... 21	
3.2.2.1 จารึกวัดป่ามะม่วง .....	21
3.2.2.2 จารึกนครชุม .....	21
3.2.2.3 จารึกพ่อขุนรามคำแหง.....	21
3.2.2.4 จารึกวัดศรีชุม.....	21
3.2.2.5 จารึกปู่ขุนจิตขุนจอต .....	21
3.2.3 กระบวนการเตรียมข้อมูลจากศิลาจารึก .....	21
3.2.3.1 คัดลอกข้อความจากต้นฉบับ .....	22
3.2.3.2 ทำการแบ่งช่องว่างของคำในประโยค .....	22

3.2.3.3 การเก็บรวบรวมคำศัพท์และรูปแบบของกลุ่มคำจากในประโยคเพื่อนำไปทำ custom dictionary ซึ่งคำศัพท์เหล่านี้จะไม่มีปรากฏในคลังคำศัพท์ (corpus) ภาษาไทยจึงจำเป็นที่จะต้องรวบรวมคำศัพท์ขึ้นมาใหม่ .....	22
3.2.4 กระบวนการทราานฟอร์เมอร์ข้อมูล.....	22
3.2.4.1 ไลบารีที่จำเป็นต้องติดตั้งสำหรับการทำโมเดลมีดังต่อไปนี้ .....	22
3.2.4.2 กระบวนการทำเตรียมข้อมูล (Preparing the dataset).....	23
3.2.4.3 การกำหนดขนาดของขนาดบล็อก (Determining the block size).....	23
3.2.4.4 การกำหนดโมเดลเรียนรู้ล่วงหน้า [20].....	23
3.2.4.5 การกำหนดเครื่องมือเก็บข้อมูล (Determining data collector) [21].....	23
3.2.4.6 กระบวนการแบ่งข้อมูลด้วยการตรวจสอบแบบข้ามและสอนโมเดล .....	24
3.2.5 กระบวนการทำนายคำถัดไปด้วยหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง ...	24
3.2.5.1 ไลบารีที่จำเป็นต้องติดตั้งสำหรับการทำโมเดลมีดังต่อไปนี้ .....	24
3.2.5.2 กระบวนการทำเตรียมข้อมูล (Preparing the dataset).....	24
3.2.5.3 กระบวนการแบ่งข้อมูลด้วยการตรวจสอบแบบข้ามและสอนโมเดล .....	24
3.2.6 ระเบียบและวิธีในการทดสอบโมเดล [22].....	25
3.2.6.1 การประเมินผลลัพธ์ความถูกต้องของโมเดลการจำแนกประเภทแบบ หลากหลาย.....	28
3.2.7 กระบวนการปรับปรุงประสิทธิภาพในการประเมินผลลัพธ์ .....	29
3.2.7.1 การวิเคราะห์ผลลัพธ์ของแต่ละโมเดล .....	29
3.2.7.2 การเพิ่มคำศัพท์เฉพาะเข้าไปในตัวแบ่งคำ.....	29
3.2.7.3 การกรองและคัดแยกคำศัพท์ในตัวแบ่งคำ .....	29
3.2.7.4 การประยุกต์ใช้คลังคำศัพท์เฉพาะในการทำนาย.....	29
3.2.8 การเปรียบเทียบสมรรถนะภาพของโมเดล (model's performance benchmarking) .....	30
บทที่ 4 การทดลองและอภิปรายผลการทดลอง .....	31

4.1 ผลการการวัดประสิทธิภาพความถูกต้องของโมเดล.....	31
4.1.1 รายงานผลการทดลอง.....	31
4.1.1.1 รายงานผลการทดลองกลุ่มโมเดลเรียนรู้ล่วงหน้า.....	31
4.1.1.2 รายงานผลการทดลองโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง.....	35
4.1.2 ผลการเปรียบเทียบสมรรถนะของโมเดลในกลุ่มโมเดลเรียนรู้ล่วงหน้าทีและโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง.....	36
4.2 ผลลัพธ์ความถูกต้องในการทำนายผลลัพธ์จากโมเดลทั้งก่อนและหลังปรับปรุงกระบวนการทำนาย.....	37
4.2.1 การวิเคราะห์ผลลัพธ์ของแต่ละโมเดลก่อนการปรับปรุง.....	38
4.2.2 การวิเคราะห์ผลลัพธ์ของแต่ละโมเดลหลังการปรับปรุง.....	39
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....	44
5.1 สรุปผลการวิจัย.....	44
5.2 ข้อเสนอแนะ.....	47
บรรณานุกรม.....	48
ประวัติผู้เขียน.....	50

## สารบัญตาราง

	หน้า
ตารางที่ 1 ความถี่ที่ถูกรับว่ามีตำแหน่งขาดหายไปจากประโยค .....	29
ตารางที่ 2 ผลลัพธ์ค่าความถูกต้องของโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา ทั้ง 10 พับ.....	31
ตารางที่ 3 ผลลัพธ์ค่าความถูกต้องของโมเดลเบิร์ต-เบส-หลายภาษา-เคสดี ทั้ง 10 พับ.....	33
ตารางที่ 4 ผลลัพธ์ค่าความถูกต้องของโมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสดี ทั้ง 10 พับ.....	34
ตารางที่ 5 ผลลัพธ์ค่าความถูกต้องของโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง ทั้ง 10 พับ .....	36
ตารางที่ 6 การเปรียบเทียบสมรรถนะของโมเดลทั้ง 4 โมเดล.....	36
ตารางที่ 7 รายงานผลลัพธ์ความถูกต้องของการทำนายอักขระในประโยคของโมเดลทั้ง 4 โมเดลก่อน .....	37
ตารางที่ 8 รายงานผลลัพธ์ความถูกต้องของการทำนายอักขระในประโยคของโมเดลทั้ง 4 โมเดลหลังปรับปรุง.....	40
ตารางที่ 9 ผลลัพธ์ความถูกต้องโมเดลในแต่ละตำแหน่งของวิธีการปิด 2 ตำแหน่งติดกัน.....	41
ตารางที่ 10 ผลลัพธ์ความถูกต้องโมเดลในแต่ละตำแหน่งของวิธีการปิด 5 ตำแหน่งติดกัน.....	42
ตารางที่ 11 ผลลัพธ์ความถูกต้องโมเดลในแต่ละตำแหน่งของวิธีการปิด 2 ตำแหน่งที่ไม่ต่อเนื่อง .....	43

## สารบัญรูปภาพ

	หน้า
ภาพที่ 1 สถาปัตยกรรมโมเดลของทรานฟอเมอร์ [10] .....	6
ภาพที่ 2 การทำงานของลำดับอินพุต (input sequence) ในแต่ละการฝัง.....	6
ภาพที่ 3 ชุดข้อมูลที่ถูกแยกออกเป็น 10 ส่วน ( $k=10$ ) โดยชุดข้อมูลที่นำไปเทรนเท่ากับ $k-1$ .....	13
ภาพที่ 4 ผลความแม่นยำโดยเฉลี่ยของการจำแนกข้ามภาษา (Average accuracy results on cross-lingual classification) [18].....	16
ภาพที่ 5 ขั้นตอนโดยสังเขปของกระบวนการดำเนินงาน.....	20
ภาพที่ 6 หลักศิลาจารึกพ่อขุนรามคำแหง.....	21
ภาพที่ 7 เมทริกซ์คอนฟิวชันแบบการจำแนกประเภทหลายคลาสในรูปแบบอาร์เรย์ .....	25
ภาพที่ 8 เมทริกซ์คอนฟิวชันแบบการจำแนกประเภทหลายคลาส .....	25
ภาพที่ 9 จำแนกค่าความถูกต้องจากเมทริกซ์คอนฟิวชัน .....	26
ภาพที่ 10 ตัวอย่างเมทริกซ์คอนฟิวชันแบบการจำแนกประเภทหลายคลาส .....	27
ภาพที่ 11 รายงานการจำแนกประเภทแบบมาโคร ของ A, B และ C .....	28
ภาพที่ 12 แผนภูมิแท่งแสดงค่าความถูกต้องของโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา.....	32
ภาพที่ 13 เมทริกซ์คอนฟิวชันแบบการจำแนกประเภท 139 คลาส ของโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา .....	32
ภาพที่ 14 แผนภูมิแท่งแสดงค่าความถูกต้องของโมเดลเบิร์ต-เบส-หลายภาษา-เคสต์ .....	33
ภาพที่ 15 เมทริกซ์คอนฟิวชันแบบการจำแนกประเภท 94 คลาส ของโมเดลเบิร์ต-เบส-หลายภาษา-เคสต์ .....	33
ภาพที่ 16 แผนภูมิแท่งแสดงค่าความถูกต้องของโมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสต์.....	34
ภาพที่ 17 เมทริกซ์คอนฟิวชันแบบการจำแนกประเภท 88 คลาส ของโมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสต์.....	35

ภาพที่ 18 แผนภูมิแท่งแสดงค่าความถูกต้องของโมเดลหน่วยความจำระยะสั้นระยะยาวแบบ สองทิศทาง.....	36
ภาพที่ 19 ผลลัพธ์ของตำแหน่งสุ่มปิด 1 อักขระในคำที่โมเดลทำนายผิด .....	38
ภาพที่ 20 ผลลัพธ์ของตำแหน่งสุ่มปิด 2 อักขระในคำที่โมเดลทำนายผิด .....	38
ภาพที่ 21 ผลลัพธ์ของตำแหน่งสุ่มปิด 5 อักขระในคำที่โมเดลทำนายผิด .....	39
ภาพที่ 22 ผลลัพธ์ของตำแหน่งสุ่มปิด 2 อักขระแบบไม่ติดกันในโมเดลทำนายผิด .....	39
ภาพที่ 23 ผลลัพธ์ค่าความถูกต้องของตำแหน่งสุ่มปิด 1 ตำแหน่ง .....	40
ภาพที่ 24 ผลลัพธ์ความถูกต้องของการทำนายอักขระในประโยคของตำแหน่งสุ่มปิด 2 ตำแหน่ง....	40
ภาพที่ 25 ผลลัพธ์ความถูกต้องของการทำนายอักขระในประโยคของตำแหน่งสุ่มปิด 5 ตำแหน่ง....	41
ภาพที่ 26 ผลลัพธ์ความถูกต้องของการทำนายอักขระในประโยคของตำแหน่งสุ่มปิด 2 ตำแหน่งที่ไม่ ต่อเนื่อง.....	42
ภาพที่ 27 ค่าความถูกต้องของโมเดลทั้งสี่ประเภทจากเมตริกซ์คอนฟิวชัน .....	45
ภาพที่ 28 ค่าความถูกต้องของโมเดลที่ตำแหน่งของ <mask> หนึ่งตำแหน่ง.....	45
ภาพที่ 29 ค่าความถูกต้องของโมเดลที่ตำแหน่งของ <mask> จะต้องติดกันทั้งสองตำแหน่ง .....	45
ภาพที่ 30 ค่าความถูกต้องของโมเดลที่ตำแหน่งของ <mask> จะต้องติดกันทั้งห้าตำแหน่ง .....	46
ภาพที่ 31 ค่าความถูกต้องของโมเดลที่ตำแหน่งของ <mask> จะต้องไม่ติดกันทั้งสองตำแหน่ง .....	46





จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

# บทที่ 1

## บทนำ

### 1.1 ที่มาและความสำคัญของปัญหา

ศิลาจารึกเป็นเอกลักษณ์ทางประวัติศาสตร์และแสดงวัฒนธรรมของช่วงเวลานั้น ๆ ในประวัติศาสตร์ที่ช่วยสะท้อนความเป็นมาของประวัติศาสตร์ให้สังคมในยุคปัจจุบันได้เข้าใจสถานการณ์และวัฒนธรรมในยุคสมัยนั้นด้วยวิธีการสลักตัวอักษร รูปภาพหรือสัญลักษณ์ลงบนศิลาตามสถานที่หรือสิ่งที่ถูกสร้างมาเช่น ข้อความที่ถูกสลักบนเสา แผ่นหินหรือผนังถ้ำ เป็นต้น เพื่อประกาศเรื่องราวของเหตุการณ์ที่เกิดขึ้นในขณะนั้น และเมื่อช่วงของระยะเวลาผ่านไปจารึกเหล่านั้นอาจเกิดการเสื่อมสภาพจากหลายสาเหตุจนเป็นรอยขีดข่วนทับข้อความ ตัวอักษรจนเลือนลางหรือถูกทำลายจากภัยธรรมชาติ จนกระทั่งทำให้ไม่สามารถถอดความจากตัวอักษรหรือถอดความมาแล้วแต่ข้อความนั้นไม่สมบูรณ์เพียงพอที่จะตีความหมายให้คนในสังคมในยุคปัจจุบันได้เข้าใจจึงส่งผลให้ยากต่อการที่จะศึกษาประวัติศาสตร์จากสิ่งบอกเล่าที่ถูกจารึกไว้

งานวิจัยฉบับนี้ได้นำชุดข้อมูลในช่วงยุคสุโขทัยที่ได้รับการถอดความจากภาษาในยุคนั้นเป็นภาษาไทยจากฐานข้อมูลจารึกแห่งประเทศไทย ศูนย์มานุษยวิทยาสิรินธร ได้แก่ จารึกวัดป่ามะม่วง จารึกนครชุม จารึกพ่อขุนรามคำแหง จารึกวัดศรีชุม และ จารึกปู่ขุนจิตขุนจอต มีทั้งข้อความที่สมบูรณ์และข้อความที่มีตัวอักษรขาดหายไปตั้งแต่ 1 ถึง 8 อักขระ จึงเสนอวิธีการสร้างโมเดลทำนายอักขระของตัวอักษรที่ขาดหายไปจากข้อความโดยการใช้เทคนิคโมเดลภาษาแบบถูกปิด (Masked language model, mlm) เข้ามาช่วยในการสอนโมเดล (train model) และใช้โมเดลเอ็กซ์แอลเอ็ม-โรเบอตา (XLM-Roberta , XLM), โมเดลเบิร์ต-เบส-หลายภาษา-เคสดี (Bert-base-multilingual-cased), โมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสดี (DistilBERT-base-multilingual-cased) เป็นโมเดลเรียนรู้ล่วงหน้าแบบหลากหลายภาษา และโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง โดยในแต่ละรอบของการเทรนจะทำการสุ่มปิดอักขระหรือข้อความประมาณ 15% และแทนที่ข้อความด้วยมาส์กโทเคน (mask token) ที่ความถี่ 1-3 ที่ตัวอักขระหายไปมากที่สุดและให้โมเดลทำนายคำที่ขาดหายไปตรงจุดที่ถูกมาส์กโทเคน

### 1.2 วัตถุประสงค์ของงานวิจัย

การสร้างโมเดลทำนายอักขระของตัวอักษรที่ขาดหายไปจากข้อความโดยการใช้เทคนิคโมเดลภาษาแบบถูกปิด

### 1.3 ขอบเขตการวิจัย

1. ชุดข้อมูลที่ใช้ในการทำวิจัยเป็นชุดข้อมูลในช่วงยุคสุโขทัยที่ได้รับการถอดความจากภาษาในยุคนั้นเป็นภาษาไทยจากฐานข้อมูลจารึกแห่งประเทศไทย ศูนย์มานุษยวิทยาสิรินธร ได้แก่

- 1.1 จารึกวัดป่ามะม่วง 110 บรรทัด
- 1.2 จารึกนครชุม 273 บรรทัด
- 1.3 จารึกพ่อขุนรามคำแหง 248 บรรทัด
- 1.4 จารึกวัดศรีชุม 406 บรรทัด
- 1.5 จารึกปู่ขุนจิตขุนจอต 97 บรรทัด

2. โมเดลและเทคนิคที่นำมาใช้ในการดำเนินงานวิจัยได้แก่ ทราานพอร์เมอร์และโมเดลภาษาแบบถูกปิด

### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

โมเดลมีความสามารถในการเติมอักขระของตัวอักษรที่ขาดหายไปจากข้อความโดยการใช้โมเดลภาษาแบบถูกปิด

### 1.5 แผนการดำเนินงาน

1. ศึกษาและทำความเข้าใจทฤษฎีและงานวิจัยที่เกี่ยวข้อง
2. สํารวจสภาพของศิลาจารึกและแหล่งสืบค้นข้อมูล
3. ศึกษาแนวทางการประมวลผลและการทำงานของ MLM
4. การดำเนินงานในขั้นตอนของการประมวลผลโมเดล
5. ประเมินผลงานวิจัย
6. สรุปลงานวิจัย
7. ตีพิมพ์ผลงานทางวิชาการ
8. จัดทำวิทยานิพนธ์

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

#### ทฤษฎีที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้องในการดำเนินงานวิจัยนี้ ประกอบด้วยชุดฐานข้อมูลที่ถอดความจากภาษาโบราณเป็นภาษาไทยจากโครงการฐานข้อมูลจารึกในประเทศไทยโดยศูนย์มานุษยวิทยาสิรินธรและการประมวลผลทางภาษาธรรมชาติ (Natural language processing, NLP)

#### 2.1 จารึก

หลักฐานทางโบราณคดีประเภทจารึกเปรียบเสมือนบันทึกบอกเล่าความเป็นมาของประวัติศาสตร์ของมนุษยชาติที่บันทึกเรื่องราวเป็นตัวอักษรหรือรูปภาพลงบนเนื้อวัตถุต่าง ๆ เช่น แผ่นศิลา, แผ่นไม้, แผ่นทอง เป็นต้น ด้วยกรรมวิธีจารึกลงบนเนื้อวัตถุด้วยวัตถุที่แหลมคมที่ถูกเรียกว่าเหล็กจาร เพื่อประกาศเรื่องราวของเหตุการณ์ที่เกิดขึ้นในขณะนั้น จารึกจึงเป็นเอกสารสำคัญที่สะท้อนความเป็นมาของประวัติศาสตร์ให้สังคมในยุคปัจจุบันได้เข้าใจ [1]

ในประเทศไทยมีการค้นพบศิลาจารึกมากกว่า 1200 หลัก มีทั้งศิลาจารึกชิ้นเล็ก ศิลาจารึกที่อยู่บนผนังถ้ำ หรือศิลาจารึกที่มีขนาดใหญ่ไม่สามารถเคลื่อนย้ายได้จำเป็นต้องมีการทำสำเนาคัดลอกศิลาจารึกเพื่อเก็บไว้ศึกษาอยู่ 2 วิธี คือ

1. คัดลอกโดยการทำสำเนา ด้วยการใช้กระดาษสาชุบน้ำแนบกับร่องอักษรในตัวแผ่นศิลาและใช้ลูกประคบขูดหมึกเงินประคบบลงกระดาษ
2. คัดลอกโดยการถ่ายภาพ ใช้แปรงเม็ดเล็กทาหน้าในสมัยโบราณทาลงบนตัวแผ่นศิลาให้ทั่วและใช้ผ้าสำลีชุบน้ำหมาด ๆ เช็ดลงบนแผ่นศิลาเหลือเพียงแปรงที่ตรึงเป็นสีขาวชัดเจนเหมาะแก่การถ่ายภาพ [1]

ปัจจุบันหน่วยงานที่เก็บรักษาหลักศิลาจารึกบางหน่วยงานไม่อนุญาตให้ใช้กรรมวิธีที่ได้อธิบายไปในข้างต้นในการทำสำเนาจารึกเนื่องจากกลัวว่าโบราณวัตถุจะได้รับความเสียหาย นอกจากนี้ยังมีปัจจัยอื่น ๆ ที่ทำให้สำเนาหรือจารึกเสียหายได้อีก เช่น น้ำซึม น้ำรั่ว ขาด หรือชำรุดจากการยืมไปใช้ [2]

ชุดฐานข้อมูลจารึกที่ใช้ในการศึกษาสำหรับงานวิจัยมีดังต่อไปนี้

##### 2.1.1 จารึกวัดป่ามะม่วง

ถูกพบที่ วัดใหม่ (ปราสาททอง) อำเภอนครหลวง จังหวัดพระนครศรีอยุธยาถูกสร้างในช่วงพุทธศักราช 1904 ส่วนอักษรที่จารึกตัวอักษรไทยโบราณ ไทยสุโขทัย ในรัชสมัยที่สุโขทัยเป็นราชธานี มีความเชื่อว่าจารึกหลักนี้จารึกพร้อมกับหลักศิลาที่จารึกเป็นภาษาเขมร โดยสมเด็จพระเจ้าบรมวงศ์

เธอกรมพระยาดำรงราชานุภาพ เสด็จขึ้นไปทอดพระเนตร ทรงอ่านหนังสือซึ่งอยู่บนหลักศิลาจารึกได้ ความเป็นศิลาจารึกของพระธรรมราชาไทย คู่กับหลักภาษาเขมร ซึ่งอยู่ในวัดพระศรีรัตนศาสดาราม คือ จารึกความอย่างเดียวกัน เป็นภาษาเขมรหลัก 1 เป็นภาษาไทยหลัก 1 [3]

### 2.1.2 จารึกนครชุม

ถูกพบที่ วัดพระบรมธาตุ ตำบลนครชุม อำเภอเมือง จังหวัดกำแพงเพชร ถูกสร้างในช่วงพุทธศักราช 1900 ส่วนอักษรที่จารึกตัวอักษรไทยโบราณ ไทยสุโขทัย ศิลาจารึกหลักนี้ตั้งอยู่ที่วัดพระบรมธาตุ ริมแม่น้ำปิง ช่างทึศได้ หน้าเมืองกำแพงเพชร ในรัชสมัยพระธรรมราชาที่ 1 เดิมวัดนี้มีพระเจดีย์ 3 องค์ องค์กลางเป็นรูปพระเจดีย์ ปัจจุบันเป็นรูปพระเจดีย์พม่าเนื่องด้วยขณะนั้นหน่วยงานรัฐบาลได้อนุญาตให้ชาวพม่าซื้อ พญาตะกา ซ่อมแซมพระเจดีย์นี้เมื่อประมาณ 30 ปีที่ ซึ่งในศิลาจารึกหลักที่ 8 เรียกว่า นครพระชุม จึงสันนิษฐานว่าเดิมคงอยู่ริมแม่น้ำปิงทิศตะวันตก [4]

### 2.1.3 จารึกพ่อขุนรามคำแหง

ถูกพบที่ เนินปราสาทเมืองเก่าสุโขทัย ตำบลเมืองเก่า อำเภอเมือง จังหวัดสุโขทัย ถูกสร้างในช่วงพุทธศักราช 1835 ส่วนอักษรที่จารึกตัวอักษรไทยโบราณ ไทยสุโขทัย เป็นเสาศิลาที่มาจากแต่เมืองสุโขทัย พบในรัชสมัยพระบาทสมเด็จพระนั่งเกล้าเจ้าอยู่หัว รัชกาลที่ 3 แห่งกรุงรัตนโกสินทร์ ปัจจุบันถูกจัดเก็บที่ พิพิธภัณฑสถานแห่งชาติ พระนคร ถนนหน้าพระธาตุ แขวงพระบรมมหาราชวัง เขตพระนคร กรุงเทพมหานคร [5]

### 2.1.4 จารึกวัดศรีชุม

ถูกพบที่ ในอุโมงค์วัดศรีชุม ตำบลเมืองเก่า อำเภอเมือง จังหวัดสุโขทัย ถูกสร้างในช่วงพุทธศตวรรษ 19-20 ส่วนอักษรที่จารึกตัวอักษรไทยโบราณ ไทยสุโขทัย ได้ถูกตรวจค้นพบศิลาจารึกเมืองสุโขทัยเมื่อ พ.ศ. 2430 โดย พลโท พระยาสมุหเสนาสงคราม ขณะดำรงตำแหน่งหลวงสมุหเสนาสงคราม ในอุโมงค์วัดศรีชุม ต่อมาใน พ.ศ. 2451 ได้ถูกส่งมาที่พิพิธภัณฑสถาน กรุงเทพฯ และถูกส่งมอบอีกครั้งให้จัดเก็บรักษาไว้ที่หอพระสมุดสำหรับพระนคร [6]

### 2.1.5 จารึกปู่ขุนจิตขุนจืด

ถูกพบที่ ริมเสาเบ็องขวาน้ำวิหารหลวง ด้านหลังวิหารสูง วัดมหาธาตุ สุโขทัย อุทยานประวัติศาสตร์สุโขทัย ตำบลเมืองเก่า อำเภอเมืองสุโขทัย จังหวัดสุโขทัย สุโขทัย ถูกสร้างในช่วงพุทธศักราช 1935 ส่วนอักษรที่จารึกตัวอักษรไทยโบราณ ไทยสุโขทัย ไม่ปรากฏหลักฐานว่าผู้ใดเป็นคนสร้าง ถูกค้นพบโดยหน่วยชุดแต่งและบูรณะเมืองสุโขทัย กองโบราณคดี กรมศิลปากร พ.ศ. 2499 [7]

## 2.2 การประมวลผลภาษาทางธรรมชาติ (Natural language processing, NLP)

การประมวลผลภาษาทางธรรมชาติ คือหนึ่งในสาขาวิชาปัญญาประดิษฐ์และวิทยาศาสตร์คอมพิวเตอร์ซึ่งกล่าวถึงปฏิสัมพันธ์ระหว่างภาษาศาสตร์ของมนุษย์และคอมพิวเตอร์เป็นการศึกษา

โมเดลการคำนวณและคณิตศาสตร์ในแง่มุมมองต่าง ๆ ของภาษาจนกระทั่งรวมไปถึงระบบภาษาพูด การประมวลผลโมเดลภาษาทางธรรมชาติเป็นการรวบรวมเทคนิคเทคนิคในการแยกโครงสร้างของไวยากรณ์และความหมายจากข้อความที่ถูกป้อนเข้ามาและประมวลผลผลลัพธ์ [8]

### 2.3 โมเดลภาษาแบบถูกปิด (Masked Language Model, MLM)

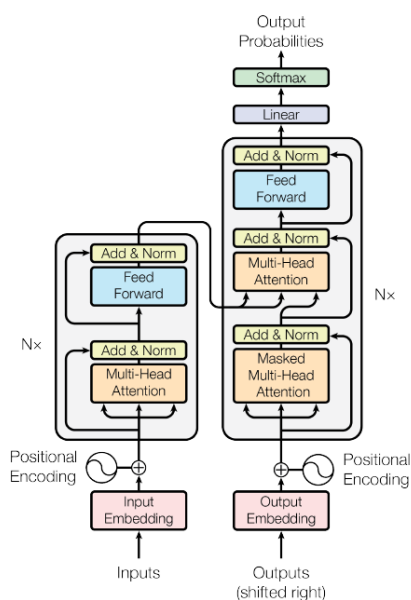
โมเดลภาษาแบบถูกปิด คือ ประเภทของโมเดลภาษาที่ถูกใช้ในโมเดลการเรียนรู้การจดจำเชิงลึกล่วงหน้า (pre-training deep neural language models) โดยในแต่ละรอบของการทำงานของโมเดลภาษาแบบถูกปิดจะทำการสุ่มใส่มาสก์ลงในข้อความอินพุตที่ถูกป้อนเข้ามาและตัวโมเดลจะถูกฝึกสอนให้เข้าใจบริบทและความสัมพันธ์ของบทความเพื่อทำนายอักขระหรือคำศัพท์ที่เป็นคำดั้งเดิมที่ถูกปิดไว้ด้วยมาสก์โทเคน “<mask>” หรือ “[MASK]” [9] โมเดลถูกออกแบบมาจากแนวคิดของแบบฝึกหัดในการเรียนรู้ภาษาที่ทำการลบคำบางส่วนออกจากข้อความในประโยคและให้ผู้เรียนได้ทำการเติมเต็มคำที่ขาดหายไปจากบทความ (Cloze task) ดังนั้นการทำงานของโมเดลคือจะทำการรับ v อินพุตเข้ามาทำการสุ่มปิดคำจากนั้นจะทำการแปลงข้อมูลที่เป็นตัวอักษรไปเป็นเวกเตอร์ในรูปแบบของตัวเลขเพิ่มข้อมูลเพิ่มเติมลงในการฝัง (embedding) ของลำดับ (sequence) ที่ไม่ได้มีลำดับให้มีป้ายกำกับที่บ่งชี้ลำดับเพื่อช่วยให้โมเดลสามารถรับรู้ลำดับของคำที่ถูกประมวลผลต่อมาโมเดลก็จะทำการแปลคำแต่ละคำโดยเฉพาะที่โมเดลสนใจเรียกว่า ความสนใจ (attention) โดยการใช้ควีรี (query) คีย์ (key) และ มูลค่า (value) ตามลำดับในการค้นหาและจับคู่คำพร้อม ๆ กันหลายความสนใจ เพื่อทำการคาดคะเนคำที่ใกล้เคียงมากที่สุดออกมาซึ่งเทคนิคที่นำมาปรับใช้กับโมเดลภาษาถูกปิดเรียกว่าทรานฟอร์เมอร์และตัวโมเดลเป็นหนึ่งในการเรียนรู้แบบล่วงหน้าแบบสองทิศทางอ้างอิงคำอธิบายเพิ่มเติมในบทที่ 2.4 และ 2.8

### 2.4 การทรานฟอร์เมอร์ข้อมูลแบบสองทิศทางเชิงลึก (Deep Bidirectional Transformers)

การทรานฟอร์เมอร์ข้อมูลแบบสองทิศทางเชิงลึก คือโมเดลการเรียนรู้การจดจำเชิงลึกที่ถูกเกิด (Google) ได้ทำการพัฒนาขึ้น และนำมาใช้ในการประมวลผลโมเดลภาษาทางธรรมชาติ โดยที่แบบสองทิศทาง (Bidirectional) คือการที่โมเดลสามารถเข้าใจและประมวลผลข้อมูลที่ถูกป้อนเข้ามาจากทั้งด้านหน้าและด้านหลังของประโยค ทำให้โมเดลมีความเข้าใจได้ดียิ่งขึ้นในการเก็บรักษาารูปแบบของภาษา และทรานฟอร์เมอร์ หมายถึงโครงสร้างพื้นฐานของโมเดลที่มีความสำคัญสำหรับการประมวลผลข้อมูลของการประมวลผลโมเดลภาษาทางธรรมชาติ มีความสามารถในการจัดการปัญหาการเรียนรู้แบบ ลำดับ-ถึง-ลำดับ (Sequence-to-Sequence) [10] และ เบิร์ต (BERT) เป็นหนึ่งในโมเดลที่ใช้การทรานฟอร์เมอร์ข้อมูลแบบสองทิศทางเชิงลึก ที่ได้รับความนิยมเนื่องจากมีประสิทธิภาพสูงในการทำความเข้าใจภาษาทางธรรมชาติ (Natural Language Understanding, NLU) และ การอนุมานภาษาธรรมชาติ (Natural Language Inference, NLI) ในภาษาอังกฤษและภาษาอื่น ๆ อีก

ทั้งนี้ยังกลายเป็นพื้นฐานในการพัฒนาเทคนิคหรือโมเดลต่าง ๆ ในปัจจุบัน เช่น โมเดลเอกซ์แอลเอ็ม-โรเบอร์ตาเบส, โมเดลติสทิลเบิร์ต เป็นต้น

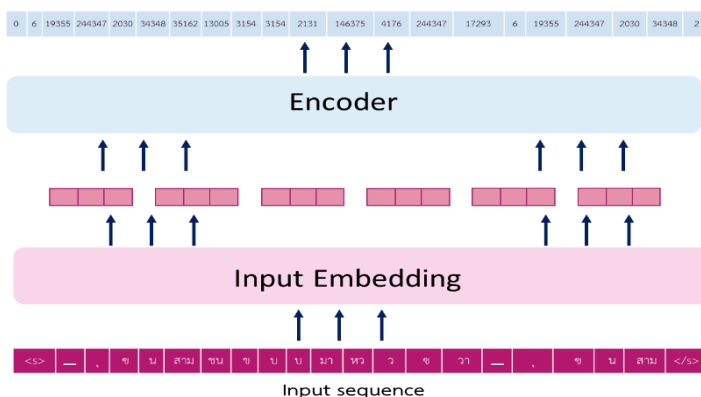
#### 2.4.1 สถาปัตยกรรมโมเดล (Model Architecture) [10]



ภาพที่ 1 สถาปัตยกรรมโมเดลของทรานฟอร์มเมอร์ [10]

1. การฝังอินพุต (Input Embedding) เป็นกระบวนการแรกของโมเดลทรานฟอร์มเมอร์ที่แปลงข้อมูลที่เป็นตัวอักษรไปเป็นเวกเตอร์ในรูปแบบของตัวเลขก่อนเข้าสู่โมเดลตามรูปภาพที่ 2 เนื่องจากโครงข่ายประสาทสามารถประมวลได้กับชุดข้อมูลที่ตัวเลขอีกทั้งเพื่อให้โมเดลสามารถประมวลผลได้ง่ายยิ่งขึ้น

ตัวอย่างของข้อความ “จนสามชนขบมาทวชว จนสาม”



ภาพที่ 2 การทำงานของลำดับอินพุต (input sequence) ในแต่ละการฝัง

2. การเข้ารหัสตำแหน่ง (Positional Encoding) เป็นการเพิ่มข้อมูลเพิ่มเติมลงในการฝังของลำดับที่ไม่ได้มีลำดับในตัวข้อมูลเองเนื่องจากในขณะที่ไม่เดลกำลังประมวลผลคำใดคำหนึ่งอยู่โมเดลไม่มีสิ่งที่บ่งบอกตำแหน่งของคำปัจจุบันเพื่อช่วยให้โมเดลสามารถรับรู้ลำดับของคำที่ถูกประมวลผลซึ่งการใช้การเข้ารหัสตำแหน่งจะใช้ฟังก์ชันไซน์ (sine) และ โคไซน์ (cosine) ในการแทนตำแหน่งต่าง ๆ ด้วยสมการที่ความแตกต่าง ๆ ตามสมการที่ 2.1 และ 2.2

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000 \frac{2i}{d_{model}}}\right) \quad (2.1)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000 \frac{2i}{d_{model}}}\right) \quad (2.2)$$

ซึ่งแต่ละตำแหน่งในประโยคจะถูกแทนที่ด้วยเวกเตอร์อันหนึ่งและหลังจากโมเดลได้รับการฝังคำ (word embedding) และการเข้ารหัสตำแหน่งของแต่ละคำมาแล้วจะนำมาบวกกัน แล้วส่งต่อไปยังชั้น (layer) ที่อยู่ถัด ๆ ไป

3. การให้ความสนใจ คือฟังก์ชันในการประมวลผลจับคู่คิวิรีเข้ากับคีย์แล้วนำผลลัพธ์ที่ใกล้เคียงความต้องการมากที่สุดออกมาโดยที่ค่าของการจับคู่ คิวิรี คีย์ มูลค่า ได้ผลลัพธ์เป็นค่าเวกเตอร์ทั้งหมดโดยผลลัพธ์ที่ได้จะถูกคำนวณเป็นผลรวมถ่วงน้ำหนักของค่าและค่านำหนักที่นำมาคำนวณจะใช้ฟังก์ชันที่มีความเข้ากันได้ของ คิวิรีและคีย์ ที่มีความเกี่ยวข้องกัน ซึ่งสิ่งที่โมเดลให้ความสนใจ คือการเรียนรู้เฉพาะข้อมูลของคำที่มีความเกี่ยวข้องมีแนวโน้มใกล้เคียงกัน

### ตัวอย่างเช่น

การพิมพ์คีย์เวิร์ดในเครื่องมือค้นหากูเกิล (google search engine) เปรียบเสมือนคิวิรีจากนั้นตัวระบบจะค้นหาคีย์เวิร์ดที่ตรงกันหรือมีความใกล้เคียงกันมากที่สุดโดยเอาคิวิรีไปเปรียบเทียบกับคีย์ ซึ่งอาจจะเปรียบเทียบกับอักขระ ข้อความ หัวข้อหรืออื่น ๆ เป็นต้น จากนั้นระบบก็จะส่งสิ่งที่ต้องการออกมาจากคำที่ถูกพิมพ์คีย์เวิร์ดเข้าไปในระบบก่อนหน้า

จากที่ได้กล่าวมาในข้างต้นในการคำนวณความสัมพันธ์ของคำในประโยคที่กำลังถูกแปลงเป็นโดเมนเวกเตอร์จะต้องใช้ "การให้ความสนใจโดยการคูณแบบจุด-ผลรวมที่ถูกปรับขนาด (Scaled Dot-Product Attention)" ในการเรียนรู้ความสัมพันธ์ระหว่างคำที่แตกต่างกันได้และจะต้องประกอบไปด้วย 3 องค์ประกอบที่ได้กล่าวมาข้างต้น คือ คิวิรี คีย์ และมูลค่า โดยขั้นตอนในการคำนวณจะประกอบไปด้วยการนำอินพุต คิวิรี คีย์ ของมิติ  $d_k$  และมูลค่า ของมิติ  $d_v$  มาทำการคำนวณด้วยการนำคิวิรีมาคูณกับคีย์แล้วทำการหาผลรวม (sum) ของผลคูณนั้น จากนั้นนำผลรวมที่ได้มาหารด้วยค่า  $\sqrt{d_k}$  และใช้ฟังก์ชัน ซอฟต์แมกซ์ (softmax) เพื่อรับค่าน้ำหนักต่างๆ ตามสมการต่อไปนี้

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2.3)$$



4. การให้ความสนใจแบบหลายหัว (Multi-Head Attention) เป็นการนำเอาการให้ความสนใจตนเอง (self-attention) ที่ได้กล่าวมาในข้างต้นมาเพิ่มความสามารถให้กับโมเดลทรานฟอร์มเมอร์โดยการให้ความสนใจพร้อมกันหลายๆ ชุด แทนที่จะใช้คิวรี คีย์ และมูลค่าที่อยู่บนมิติเดียวกันของโมเดลเพียงชุดเดียวโดยวิธีการนี้ช่วยเพิ่มความสามารถในการเรียนรู้และเข้าใจความสัมพันธ์ในข้อมูลได้หลากหลายมุมมองได้มากขึ้นและพบประโยชน์จากการประมาณการเชิงเส้นคิวรี คีย์ และมูลค่าออกมาเป็น  $h$  ครั้งด้วยการเรียนรู้การประมาณการเชิงเส้นที่มีความต่างกันของ  $d_k$  และของ  $d_v$  ตามลำดับ ส่งผลให้การให้ความสนใจแบบหลายหัวสามารถพิจารณาข้อมูลจากพื้นที่ย่อย ๆ (subspaces) ต่าง ๆ ที่แตกต่างกันรวมกันได้และในแต่ละตำแหน่งสามารถพิจารณาข้อมูลได้หลายๆ แบบพร้อมกันด้วยการใช้การเรียนรู้การประมาณการเชิงเส้นที่แตกต่างกันจากนั้นจะทำการประมวลผลการให้ความสนใจบนแต่ละหัวการให้ความสนใจ (attention head) และทำการเชื่อมต่อหรือรวมผลลัพธ์ทั้งหมดเข้าด้วยกันตามสมการต่อไปนี้

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1 + \dots + \text{head}_n) \quad (2.4)$$

$$\text{where } \text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (2.5)$$

ประเภทของการให้ความสนใจแบ่งออกเป็นสามประเภทดังต่อไปนี้

1. ชั้น "การให้ความสนใจระหว่างตัวเข้ารหัสและตัวถอดรหัส (encoder-decoder attention)" คิวรีข้อมูลมาจากชั้นก่อนหน้าของตัวถอดรหัส (decoder) และหน่วยความจำของคีย์ (memory key) และมูลค่าที่ได้จากผลลัพธ์ของการเข้ารหัสโดยที่โมเดลสามารถดูทุก ๆ ตำแหน่งในประโยคของข้อมูลเข้าสู่โมเดลได้ในตัวเข้ารหัสประกอบด้วยชั้นการให้ความสนใจตนเอง (self-attention layer) และในชั้นการให้ความสนใจตนเองแต่ละตำแหน่งจะมีการคำนวณค่า คีย์ มูลค่า และ คิวรี จากตำแหน่งก่อนหน้านั้นใช้เป็นอินพุตให้กับชั้นการให้ความสนใจตนเองที่ตำแหน่งนั้นๆ อีกทั้งแต่ละตำแหน่งใน ตัวเข้ารหัสสามารถมองหาคำความสัมพันธ์ระหว่างตนเองกับตำแหน่งทั้งหมดในชั้นก่อนหน้าของตัวถอดรหัสได้ทุกตำแหน่ง

2. ชั้นการให้ความสนใจตนเองในตัวถอดรหัส อนุญาตให้แต่ละตำแหน่งในตัวถอดรหัสสามารถเข้าร่วมกับตำแหน่งอื่น ๆ ในตัวถอดรหัสได้โดยเฉพาะกับตำแหน่งก่อนหน้าเพื่อรักษาคุณสมบัติของโมเดลให้เป็นอัลกอริทึมออโต้เรกเรสซีฟ (auto regressive) ป้องกันการไหลของข้อมูลไปในทางซ้ายของตัวถอดรหัสโดยการปรับแต่งฟังก์ชันการให้ความสนใจโดยการคูณแบบจุด-ผลรวมที่ถูกปรับขนาดโดยทำการปิดบัง (mask out) ค่าที่ไม่เกี่ยวข้องของกับตำแหน่งที่ไม่ได้รับอนุญาต ในข้อมูลนำเข้าของ ซอฟต์แม็กซ์ที่จะนำไปใช้ในการคำนวณคะแนนซอฟต์แม็กซ์

3. เครือข่ายกำหนดข้อมูลแยกตามตำแหน่งของข้อมูล (Position-wise Feed-Forward Networks, FFN) คือส่วนหนึ่งของโมเดลทรานส์ฟอร์มเมอร์ในการเรียนรู้เชิงลึกที่มีไว้สำหรับประมวลผลแบบต่อเนื่องบนแต่ละตำแหน่งของข้อมูล โดยสามารถสร้างความซับซ้อนและความสามารถในการเรียนรู้และแยกแยะความหมายของข้อมูลได้มากขึ้นซึ่ง เครือข่ายกำหนดข้อมูลแยกตามตำแหน่งของข้อมูลได้รับการออกแบบมาให้ทำงานเลียนแบบเซลล์ประสาทของมนุษย์โดยที่แต่ละชั้นมีการใช้ฟังก์ชันเชิงเส้นที่มีน้ำหนักแตกต่างกัน โดยชั้นที่ 1 จะมีจำนวนของอินพุตเท่ากับมิติของคำศัพท์ (dmodel) และจำนวนผลลัพธ์เท่ากับมิติของชั้นส่งข้อมูลไปข้างหน้า (dimension feed-forward layer, dff) ส่วนชั้นที่ 2 จะมีจำนวนอินพุตเท่ากับมิติของชั้นส่งข้อมูลไปข้างหน้าและจำนวนผลลัพธ์เท่ากับมิติของคำศัพท์นั่นคือจำนวนผลลัพธ์จะเท่ากับจำนวนอินพุตในชั้นก่อนหน้านี้ ดังนั้นชื่อ ตามตำแหน่งของข้อมูล (Position-wise) มาจากการที่แต่ละตำแหน่งของข้อมูลจะได้รับการประมวลผลเป็นอิสระต่อกัน

## 2.5 การปรับละเอียด (Fine tune)

หมายถึงเทคนิคที่ใช้ในการปรับแต่งโมเดลเชิงลึก (Deep Learning Model) ที่ได้ทำการสร้างและถูกสอนมาก่อนหน้านี้เพื่อปรับให้โมเดลสามารถทำงานด้วยข้อมูลที่เฉพาะเจาะจงโดยการปรับละเอียดและการปรับละเอียด คือการนำโมเดลที่เคยถูกสอนไว้กับชุดข้อมูลอื่น ๆ มาปรับเข้าใช้กับชุดข้อมูลใหม่สำหรับงานอื่น ๆ ด้วยการเรียนรู้ความสัมพันธ์ของชุดข้อมูลเก่าและชุดข้อมูลใหม่ [11] โดยเทคนิคนี้เป็นที่นิยมศาสตร์วิชาการเรียนรู้ของเครื่อง (machine learning) การเรียนรู้เชิงลึก (deep learning) และการประมวลผลภาษาธรรมชาติของโมเดลเบิร์ต โมเดลติสทิลเบิร์ตและโมเดลเอกซ์แอลเอ็ม-โรเบอร์ตาเบสด้วยเทคนิคการปรับละเอียดซึ่งการปรับละเอียดจำเป็นที่จะต้องเลือกโมเดลที่เหมาะสมกับข้อมูลและงาน การปรับค่าพารามิเตอร์และการเทรนโมเดลใหม่ด้วยโมเดลเรียนรู้ล่วงหน้าที่ถูกเลือกและชุดข้อมูลของงานใหม่เพื่อให้โมเดลใหม่ที่ถูกเทรนมีประสิทธิภาพและวิเคราะห์ข้อมูลได้ดีขึ้น

## 2.6 การเรียนรู้เชิงลึก (Deep Learning) [12]

การเรียนรู้เชิงลึก คือหนึ่งศาสตร์ของการเรียนรู้ด้วยเครื่องโดยใช้โครงข่ายประสาทเทียม (neural network) ที่มีหลายชั้น (Deep) ซึ่งวิธีการเรียนรู้เชิงลึกได้ถูกพัฒนามาจากโครงข่ายประสาทเทียม (artificial neural networks, ANNs) [13] โดยมีวัตถุประสงค์เพื่อให้โมเดลสามารถเรียนรู้, วิเคราะห์และประมวลผลข้อมูลได้อย่างเหมาะสม

### 2.6.1 ทฤษฎีความน่าจะเป็น (Probability Theory)

ทฤษฎีความน่าจะเป็น คือ หนึ่งในสาขาของคณิตศาสตร์ที่ถูกนำมาใช้เป็นตัวแทนที่ไม่แน่นอนในการวิเคราะห์และอธิบายความน่าจะเป็นของเหตุการณ์นั้น ๆ ทฤษฎีความน่าจะเป็นถูกนำมาใช้ในทางปัญญาประดิษฐ์ (Artificial Intelligence, AI) อยู่ 2 แนวทางหลัก ๆ ดังต่อไปนี้

#### 2.6.1.1. การใช้กฎความน่าจะเป็น (Laws of Probability)

ที่สามารถช่วยปัญญาประดิษฐ์ (Artificial intelligence, AI) สามารถชี้แนะและให้เหตุผลได้ดั่งนั้น มนุษย์จึงออกแบบอัลกอริทึมเพื่อให้คอมพิวเตอร์สามารถประมวลผลและประมาณโดยการใช้ทฤษฎีความน่าจะเป็น

#### 2.6.1.2. การใช้ความน่าจะเป็นและสถิติวิเคราะห์พฤติกรรมของระบบที่ AI นำเสนอผลงาน

ในขณะที่ทฤษฎีความน่าจะเป็นช่วยให้เราสามารถให้เหตุผลและวิเคราะห์ข้อมูลได้จากความไม่แน่นอน แต่ยังคงมีทฤษฎีสารสนเทศที่ช่วยให้เราสามารถหาปริมาณความไม่แน่นอนได้จากการแจกแจงความน่าจะเป็น

### 2.6.2 ตัวแปรสุ่ม (A random variable)

ตัวแปรสุ่ม คือ ตัวแปรที่สามารถรับค่าต่าง ๆ แบบสุ่มได้โดยทั่วไปด้วยการระบุตัวอักษรพิมพ์เล็กเช่น  $x_1, x_2, \dots, x_n$  ให้กับตัวแปรเหล่านั้น สำหรับตัวแปรที่เป็นเวกเตอร์ (vector) อาจจะมีการกำหนดอักษรหรือสัญลักษณ์ให้เป็นตัวแทนของค่าเหล่านั้น ๆ ได้ เช่น  $x$  เป็นต้น หรือการกำหนด  $x$  ให้กับค่าใดค่าหนึ่งในตัวแปร ดังนั้นตัวแปรสุ่มสามารถอธิบายสถานะความเป็นไปได้กับการแจกแจงความน่าจะเป็นควบคู่ไปด้วยกัน

### 2.6.3 การแจกแจงความน่าจะเป็น (Probability Distributions)

การแจกแจงความน่าจะเป็น คือการแจกแจงความเป็นไปได้ของตัวแปรสุ่มในรูปแบบของการกระจายความเป็นไปได้ของค่าต่างๆ ทุกค่า ซึ่งสามารถใช้เพื่อคำนวณความน่าจะเป็นของเหตุการณ์ต่างๆ ที่เกี่ยวข้องกันกับตัวแปรสุ่มนั้นๆ

#### 2.6.3.1 ตัวแปรไม่ต่อเนื่องและฟังก์ชันมวลของความน่าจะเป็น (Discrete Variables and Probability Mass Functions, PMF)

การแจกแจงความน่าจะเป็นบนตัวแปรที่ไม่ต่อเนื่องสามารถอธิบายได้โดยใช้ฟังก์ชันมวลของความน่าจะเป็นและใช้  $P$  เป็นสัญลักษณ์ตัวแทนของฟังก์ชันนี้ โดยมักจะถูกเขียนแทนด้วยฟังก์ชัน  $P(x)$  โดยที่  $x$  คือค่าของตัวแปรสุ่มและ  $P(x)$  คือความน่าจะเป็นที่เหตุการณ์  $x$  จะเกิดขึ้น หรือ  $P(y)$  โดยที่  $y$  คือค่าของตัวแปรสุ่มและ  $P(y)$  คือความน่าจะเป็นที่เหตุการณ์  $y$  จะเกิดขึ้นโดยที่  $P(x)$  และ  $P(y)$  เป็นเหตุการณ์หรืออื่น ๆ ที่ไม่เหมือนกัน โดยที่ความน่าจะเป็นที่ฟังก์ชัน  $P(x)$  มีค่า ความน่าจะเป็นเท่ากับ 1 แสดงว่า  $x = x$  อย่างแน่นอน และหากความน่าจะเป็นเท่ากับ 0 แสดงว่าไม่มีความเป็นไปได้ที่  $x = x$  PMF สามารถระบุตัวแปรหลายตัวได้ในเวลาเดียวกันหรือระบุความน่าจะเป็นของ

เหตุการณ์สองเหตุการณ์หรือมากกว่า ที่เกิดขึ้นพร้อมกัน (joint probability distribution) โดยระบุความน่าจะเป็นของทุกๆค่า โดยมีค่าความน่าจะเป็นรวมเท่ากับ 1 ในทุกๆสถานะเป็นหลัก จะใช้สัญลักษณ์  $P(x=x, y=y)$  หรือ  $P(x,y)$  ในการแสดงความน่าจะเป็นของเหตุการณ์สองเหตุการณ์หรือมากกว่าที่เกิดขึ้นพร้อมกันโดย  $x$  และ  $y$  เป็นตัวแปรสุ่มที่สนใจการทำให้เป็นฟังก์ชันมวลของความน่าจะเป็นบนตัวแปรสุ่ม  $x$  จะต้องเป็นไปตามคุณสมบัติดังต่อไปนี้

โดเมนของ  $P$  ต้องเป็นเซตที่เป็นไปได้ทั้งหมดของ  $x$

$$\forall x \in x, 0 \leq P(x) \leq 1 \quad (2.6)$$

เหตุการณ์ที่มีความน่าจะเป็นที่จะเกิดขึ้นหรือไม่มีโอกาสที่จะเกิดเลยจะมีค่าเท่ากับ 0 และไม่สามารถมีค่าน้อยกว่า 0 ได้ และในทางกลับกันหากเหตุการณ์นั้นมีความน่าจะเป็นที่จะเกิดขึ้นอย่างแน่นอนจะมีค่าเท่ากับ 1 และไม่สามารถมีค่ามากกว่า 1 ได้ คุณสมบัตินี้ถูกทำให้เป็นมาตรฐานหากปราศจากคุณสมบัตินี้สามารถหาความน่าจะเป็นจากการคำนวณความน่าจะเป็นของหนึ่งในหลาย ๆ เหตุการณ์ที่เกิดขึ้น

2.6.3.2 ตัวแปรต่อเนื่องและฟังก์ชันความหนาแน่นของความน่าจะเป็น (Continuous Variables and Probability Density Functions, PDF)

เมื่อโมเดลมีการใช้งานกับตัวแปรสุ่มแบบต่อเนื่องสามารถอธิบายการแจกแจงความน่าจะเป็นด้วยฟังก์ชัน PDF ด้วยการคำนวณพื้นที่ใต้เส้นกราฟของฟังก์ชันนั้นๆ โดยเป็นค่าบวกและมีค่ารวมทั้งหมดเท่ากับ 1 ในแต่ละช่วงของตัวแปรสุ่ม การทำให้เป็นฟังก์ชันความหนาแน่นของความน่าจะเป็นบนตัวแปรสุ่ม  $x$  จะต้องเป็นไปตามคุณสมบัติดังต่อไปนี้

โดเมนของ  $P$  ต้องเป็นเซตที่เป็นไปได้ทั้งหมดของ  $x$

$$\forall x \in x, P(x) \geq 0 \text{ (ไม่จำเป็นที่ } P(x) \leq 1) \quad (2.7)$$

$$\int p(x)dx = 1 \quad (2.8)$$

2.6.3.3 การแจกแจงแบบเอกซ์โปเนนเชียลและลาปลาซ (Exponential and Laplace Distributions)

ในบริบทของการเรียนรู้เชิงลึก นักวิจัยมักพบสถานการณ์ที่เราต้องการให้มีการแจกแจงความน่าจะเป็นที่มีจุดหักเห  $x = 0$  การใช้การแจกแจงความน่าจะเป็นด้วยเอกซ์โปเนนเชียลสามารถถูกนำมาใช้ในการแก้ปัญหานี้ได้โดยการแจกแจงแบบเอกซ์โปเนนเชียลเป็นการแจกแจงความน่าจะเป็น

แบบต่อเนื่องที่อธิบายเวลาระหว่างเหตุการณ์ในกระบวนการปัวซอง (Poisson process) ซึ่งเหตุการณ์ต่างๆ เกิดขึ้นอย่างต่อเนื่องและเป็นอิสระจากกันในอัตราคงที่ด้วยฟังก์ชันต่อไปนี้

$$p(x; \lambda) = \lambda 1_{x \geq 0} \exp(-\lambda x) \quad (2.9)$$

การแจกแจงแบบเอกซโพเนนเชียลใช้ตัวบ่งชี้ฟังก์ชัน  $1_{x \geq 0}$  เพื่อกำหนดค่าความน่าจะเป็นเท่ากับ 0 ให้กับค่าลบทั้งหมด

การแจกแจงความน่าจะเป็นที่เกี่ยวข้องกันอย่างใกล้ชิดที่ช่วยให้สามารถหาจุดตัดสูงสุด (sharp peak) ของความน่าจะเป็น ณ จุดใดจุดหนึ่ง  $\mu$  คือการแจกแจงแบบลาปลาซ

$$Laplace(x; \mu, \gamma) = \frac{1}{2\gamma} \exp\left(-\frac{|x-\mu|}{\gamma}\right) \quad (2.10)$$

## 2.6.4 กฎของเบย์ (Bayes' Rule)

กฎของเบย์เป็นกฎทางคณิตศาสตร์ที่ใช้ในการคำนวณความน่าจะเป็นของสิ่งต่างๆ โดยพิจารณาจากข้อมูลหลายๆ อย่างพร้อมกัน

$$P(x|y) = \frac{P(x)P(y|x)}{P(y)} \quad (2.11)$$

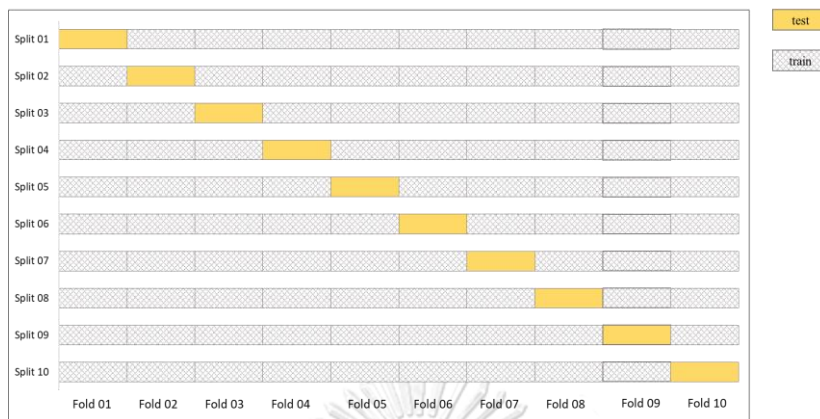
โดยกฎของเบย์กำหนดว่า ความน่าจะเป็นของเหตุการณ์  $x$  เมื่อมีข้อมูลของเหตุการณ์  $y$  เกิดขึ้น จะเท่ากับ ความน่าจะเป็นของเหตุการณ์  $y$  เมื่อมีข้อมูลของเหตุการณ์  $x$  เกิดขึ้น คูณด้วยความน่าจะเป็นของเหตุการณ์  $x$  แล้วหารด้วยความน่าจะเป็นของเหตุการณ์  $y$  โดยไม่มีเงื่อนไขเบ็ดเสร็จใดๆ

## 2.7 พื้นฐานการเรียนรู้ด้วยเครื่อง (Machine Learning Basics) [12]

### 2.7.1 การวัดประสิทธิภาพของโมเดลด้วยวิธีการตรวจสอบแบบข้าม (Measuring model performance by cross validation)

การตรวจสอบแบบข้าม คือ วิธีการประเมินประสิทธิภาพการทำงานในทางสถิติที่มีความเสถียรในการแยกส่วนของข้อมูลเพื่อนำไปเทรนและทดสอบซึ่งการวัดประสิทธิภาพด้วยวิธีการแบ่งข้อมูลเป็นที่นิยมใช้กันมากที่สุดคือ การวัดประสิทธิภาพโมเดลด้วยการแบ่งข้อมูลเป็น  $K$  ส่วน (k-fold cross validation) โดยที่  $K$  คือ ตัวเลขคงที่ที่ผู้ใช้งานระบุ ปกติแล้วมันจะกำหนดกันที่ 5 หรือ 10 ตัวอย่างเช่น  $K = 10$  เมื่อมีการแบ่งชุดข้อมูล 10-fold cross-validation ข้อมูลจะถูกแบ่งออกเป็น 10 ส่วนขนาดเท่า ๆ กัน ในส่วนนี้จะถูกเรียกว่า พับ (folds) ลำดับต่อมาจะเป็นการสอนโมเดล โมเดลแรกที่ถูกสอนจะใช้พับแรกเป็นชุดทดสอบและพับถัดไปเป็นชุดสอน โดยใช้ข้อมูลในชุดสอนสร้าง

โมเดลขึ้นมาและใช้พบแรกมาประเมินความแม่นยำของโมเดลและทำอย่างนี้ซ้ำจนครบจำนวนของ K



ดั่งภาพที่ 3

ภาพที่ 3 ชุดข้อมูลที่ถูกแยกออกเป็น 10 ส่วน (k=10) โดยชุดข้อมูลที่น่าไปเทรนเท่ากับ k-1

## 2.7.2 การประเมินภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation, MLE)

การประเมินภาวะน่าจะเป็นสูงสุด เป็นเทคนิคในการคำนวณค่าพารามิเตอร์ของโมเดลทางสถิติ โดยที่ค่าพารามิเตอร์จะถูกกำหนดให้มีค่าที่เหมาะสมที่สุด เพื่อให้โมเดลทำนายผลลัพธ์ได้ตรงกับข้อมูลจริงมากที่สุด วิธีการนี้จะใช้ข้อมูลที่เป็นตัวอย่าง (sample data) เพื่อคำนวณค่าความน่าจะเป็นที่สูงที่สุด (maximum likelihood) ของค่าพารามิเตอร์ที่เหมาะสมพิจารณาชุดตัวอย่างขนาด  $m$  ที่มีค่า  $x(1), \dots, x(m)$  ถูกสุ่มมาจากการกระจายของข้อมูลจริงที่ไม่รู้จัก  $P_{model}(x; \theta)$  ให้  $\theta$  เป็นตัวแทนของพารามิเตอร์การแจกแจงความน่าจะเป็นของ  $P_{model}$  ที่อยู่ในพื้นที่ที่ตัดกันเดียวกัน  $P_{model}(x; \theta)$  จะจับคู่กำหนดค่า  $x$  ใด ๆ ไปเป็นตัวเลขจำนวนจริงหนึ่ง ซึ่งเป็นการประมาณความน่าจะเป็นของการเกิดข้อมูล  $x$  จากการแจกแจงนั้น ๆ ซึ่งจะอ้างอิงไปยังพารามิเตอร์  $\theta$  ที่กำหนดโดย  $P_{model}$  โดยจะใช้พารามิเตอร์ที่แตกต่างกันเพื่อปรับปรุงความแม่นยำของการประมาณ การประเมินภาวะน่าจะเป็นสูงสุด สำหรับ  $\theta$  มีนิยามดังต่อไปนี้

$$\theta_{ML} = \arg_0 \max P_{model}(x; \theta) \quad (2.10)$$

$$\theta_{ML} = \arg_0 \max \prod_{i=1}^m P_{model}(x^i; \theta) \quad (2.11)$$

## 2.8 หน่วยความจำระยะสั้นระยะยาว (Long Short-Term Memory, LSTM)

หน่วยความจำระยะสั้นระยะยาว [14] เป็นประเภทหนึ่งของโครงข่ายประสาทเทียมแบบโครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network, RNN) ที่ถูกออกแบบมาเพื่อแก้ปัญหาของการเรียนรู้และจำข้อมูลลำดับ (Sequential Learning and Memory) โดยเฉพาะใน

กรณีข้อมูลลำดับที่มีความยาวระดับหนึ่ง (Long-Term Dependencies) หรือความสัมพันธ์ระหว่างข้อมูลที่อยู่ห่างกันในลำดับ (Distant Dependencies) แต่โครงข่ายประสาทเทียมแบบแบบวนซ้ำที่ใช้อยู่ทั่วไปมักพบปัญหาการสูญเสียลำดับของข้อมูลในข้อความที่ขนาดยาวส่งผลให้โมเดลมีความยากในการจดจำและเรียนรู้ความสัมพันธ์ระหว่างข้อมูลในระยะยาว ดังนั้นหน่วยความจำระยะสั้นระยะยาวจึงถูกออกแบบมาให้มีโครงสร้างพิเศษที่ช่วยในการจดจำข้อมูลระยะยาวได้ดีกว่าโครงข่ายประสาทเทียมแบบวนซ้ำแบบธรรมดาโดยประกอบด้วยเซลล์หลัก ๆ ที่มีการควบคุมการไหลของข้อมูล (Cell State) และเกตเวย์ต่าง ๆ เพื่อควบคุมกระบวนการในการอัปเดตข้อมูลในเซลล์ ต่อมาได้มีการพัฒนาหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง (Bidirectional Long Short-Term Memory, BiLSTM) [15] จากโครงข่ายประสาทเทียมแบบวนซ้ำเพื่อใช้ในการประมวลผลข้อมูลลำดับ (Sequential data) เช่น ข้อความ หรือเสียง สามารถค้นหาความสัมพันธ์ที่ซับซ้อน โดยหลักการของ BiLSTM คือการใช้โครงสร้างโครงข่ายประสาทเทียมแบบวนซ้ำสองชั้นที่มีทิศทางการผ่านข้อมูลตรงข้ามกัน ซึ่งช่วยให้โมเดลสามารถดูข้อมูลในลำดับทั้งด้านซ้ายและด้านขวาของตำแหน่งปัจจุบันได้ เช่น เมื่อประมวลผลข้อความ โมเดลจะมีการพิจารณาคำที่อยู่ก่อนและคำที่อยู่หลังของคำนั้น ซึ่งช่วยให้โมเดลเก็บข้อมูลความสัมพันธ์ที่เป็นประโยชน์ได้มากขึ้น

### งานวิจัยเกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องสำหรับการประมวลผลภาษาธรรมชาติด้วยวิธีการต่าง ๆ มีดังนี้

#### 2.9 เบิร์ต: การเรียนรู้เชิงลึกล่วงหน้าของโมเดลทรานฟอร์มเมอร์สำหรับเข้าใจภาษาในแง่ของการเข้าใจความหมายทั้งแบบแยกตามทิศทางและแบบไม่แยกตามทิศทาง

(BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding) [16]

งานวิจัยฉบับนี้ได้นำเสนอโมเดลภาษาตัวใหม่ถูกเรียกว่า เบิร์ต มีชื่อย่อมาจาก Bidirectional Encoder Representations from Transformers เบิร์ตถูกออกแบบมาเพื่อทำการเรียนรู้ล่วงหน้าในการประมวลผลของข้อความที่ไม่ได้ถูกกำกับโดยปรับเงื่อนไขข้อความจากทั้งด้านซ้ายและด้านขวาของประโยคเพื่อปรับปรุงพื้นฐานการปรับละเอียดด้วยเบิร์ตในการคลี่คลายปัญหาของการจำกัดขอบเขตโมเดลภาษามาตรฐาน (standard language models) ที่มีข้อจำกัดในการเชื่อมโยงข้อมูลทำให้ไม่สามารถใช้กับงานที่โครงสร้างซับซ้อนได้ เช่น โมเดลแหล่งกำเนิดการเรียนรู้ล่วงหน้าทรานฟอร์มเมอร์ ที่ถูกพัฒนาโดย OpenAI (Generative Pre-trained Transformer , OpenAI GPT) เป็นโมเดลประมวลผลทิศทางเดียวของข้อความทำให้ไม่สามารถนำไปใช้งานที่จำเป็นต้องเรียนรู้บริบททั้งสองทิศทางเช่น งานตอบคำถาม (Question Answering, QA) หรือ งานเติมคำที่ขาดหาย (Fill Mask) เป็นต้น ผู้วิจัยจึงเสนอการทำโมเดลภาษาแบบถูกปิดที่เรียนรู้ล่วงหน้าโดยได้รับแรงบันดาลใจ

มาจากแบบฝึกหัดในการเรียนรู้ภาษาที่ทำการลบคำบางส่วนออกจากข้อความในประโยคและให้ผู้เรียนได้ทำการเติมเต็มคำที่ขาดหายไปจากบทความ[17] แบบฝึกหัดในการเรียนรู้ภาษา ที่มีการลบ บางส่วนของข้อความออกไปแล้วให้ผู้เรียนเติมคำลงไปในช่วงว่างในข้อความและคำที่ต้องเติมจะต้อง สอดคล้องกับบริบทของประโยคนั้น ดังนั้นโมเดลภาษาแบบถูกปิดจะทำการสุ่มปิดคำในข้อความ ด้วย มาสก์โทเคน (masked token) ที่ถูกป้อนเข้าไปเพื่อทำนายไอดี (id) ของอักขระหรือคำศัพท์ดั้งเดิม เฉพาะบทความนั้นเท่านั้นโดยที่การทำงานจะไม่เหมือนกับโมเดลที่เรียนรู้ข้อความจากซ้ายไปขวา (left-to right language model) แบบเก่าเนื่องจากโมเดลภาษาแบบถูกปิดถูกออกแบบมาเพื่อให้ สามารถรวมการเรียนรู้จากซ้ายไปขวาและขวาไปซ้ายเข้าไว้ด้วยกันซึ่งวิธีการแบบนี้ทำให้สามารถใช้กา รทรานเฟอร์เมอร์ข้อมูลแบบสองทิศทางเชิงลึกและยังสามารถคาดคะเนประโยคถัดไป (next sentence prediction)

## 2.10 ตัวแทนการเรียนรู้ข้ามภาษาแบบไม่มีการสอนขนาดใหญ่ (Unsupervised Cross-lingual Representation Learning at Scale) [18]

งานวิจัยนี้ได้นำเสนอการศึกษาวิจัยด้านโมเดลภาษาธรรมชาติที่ช่วยในการโอนย้าย (Transfer) ข้อมูลของภาษาต่าง ๆ ผู้วิจัยทำการสอนทรานเฟอร์เมอร์โมเดลภาษาแบบถูกปิดด้วยภาษาต่าง ๆ ถึง 100 ภาษา โดยใช้ข้อมูลที่ CommonCrawl ที่ผ่านการกรองมาแล้วมากกว่า 2 เทระไบต์ และมีชื่อ เรียกว่า เอกซ์แอลเอ็ม-อาร์ (XLM-R) โมเดลนี้ได้รับการวัดประสิทธิภาพว่าสามารถทำงานดีกว่าโมเดล เบิร์ทหลายภาษา (mBERT) ในเกณฑ์มาตรฐานของงานโอนย้ายข้ามภาษา (cross-lingua) ซึ่ง ประกอบไปด้วยค่าเฉลี่ยของความแม่นยำที่เพิ่มขึ้น 14.6 เปอร์เซ็นต์ บนการอนุมานภาษาธรรมชาติ ข้ามภาษา (Cross-lingual Natural Language Inference, XNLI), ค่าเฉลี่ยของคะแนน F1 (F1 score) เพิ่มขึ้น 13 เปอร์เซ็นต์ บน MLQA และเฉลี่ยของคะแนน F1 ที่เพิ่มขึ้น 2.4 เปอร์เซ็นต์ บน NER และการทำงานโมเดลนี้ยังรองรับการทำงานกับภาษาที่มีข้อมูลน้อย ซึ่งมีการปรับปรุงความแม่นยำของ XNLI เพิ่มขึ้น 15.7 เปอร์เซ็นต์ในภาษาสวาฮีลี (Swahili) และ 11.4 เปอร์เซ็นต์ ในภาษาอูรดู (Urdu) ปัจจัยหลักที่ช่วยให้โมเดลเอกซ์แอลเอ็ม-โรเบอตา ประสบความสำเร็จ คือ

1. การแลกเปลี่ยนระหว่างการควบคุมความสามารถในการแบ่งแยก (capacity dilution) และการถ่ายโยงการเรียนรู้ทางบวกในระดับต่าง ๆ

เป็นกระบวนการที่ช่วยให้โมเดลเอกซ์แอลเอ็ม-โรเบอตา สามารถสร้างแบบจำลองที่มีความสามารถทั้งในการจัดการกับข้อมูลที่ซับซ้อน การเข้าใจและวิเคราะห์ข้อความได้อย่างถูกต้อง แต่ก็มีข้อจำกัดบางอย่างในเรื่องของความจุของโมเดล ที่ต้องพิจารณาในกระบวนการพัฒนาโมเดลเช่น ความจำและความเร็วของการฝึกสอนและการใช้งานขณะทำนาย ดังนั้น (1) การควบคุมความสามารถ ในกระบวนการที่โมเดลต้องการความสามารถในการจดจำและเรียนรู้ข้อมูลที่มีขนาดใหญ่หรือซับซ้อน โดยการควบคุมความสามารถนี้ช่วยให้โมเดลสามารถจัดการกับข้อมูลที่ซับซ้อนได้อย่างมีประสิทธิภาพ



ไม่จำเป็นต้องใช้ทรัพยากรมากมายในการฝึกสอนหรือการทำงานจริง (2) การถ่ายโยงการเรียนรู้ทางบวกในระดับต่าง ๆ เป็นกระบวนการที่โมเดลใช้เพื่อรับรู้แนวความคิดหรือความหมายของข้อความที่ถูกส่งเข้ามา โดยการถ่ายโยงนี้ช่วยให้โมเดลสามารถเข้าใจและวิเคราะห์ข้อความได้อย่างถูกต้อง และมีความสอดคล้องกับบริบทที่กำหนดไว้ ดังนั้นเมื่อนำ (1) (2) มารวมกันจึงเป็นกระบวนการที่ช่วยทำให้โมเดลเอกซ์แอลเอ็ม-โรเบอร์ตาประสบความสำเร็จสามารถสร้างแบบจำลองที่มีความสามารถในการอนุมานภาษาธรรมชาติข้ามภาษา

## 2. ประสิทธิภาพของทรัพยากรภาษาที่สูงและต่ำในระดับต่าง ๆ

ผู้วิจัยได้ทำการวัดประสิทธิภาพของทรัพยากรภาษาที่สูงและต่ำในระดับต่าง ๆ และผลกระทบจากจำนวนตัวอย่างของภาษาและขนาดของคำศัพท์จากการทดลองพบว่าเมื่อมีการกำหนดขนาดความสามารถในการรองรับของโมเดลแบบคงที่ถึงจำนวนของภาษาที่มากขึ้นยิ่งทำให้ประสิทธิภาพของโมเดลดีขึ้นกว่าภาษาที่มีทรัพยากรต่ำจนถึงจุดหนึ่งอ้างอิงประสิทธิภาพตามภาพที่ 4

Model	D	#M	#g	en	fr	es	de	el	bg	ru	tr	ar	vi	th	zh	hi	sw	ur	Avg
<i>Fine-tune multilingual model on English training set (Cross-lingual Transfer)</i>																			
Lample and Conneau (2019)	Wiki+MT	N	15	85.0	78.7	78.9	77.8	76.6	77.4	75.3	72.5	73.1	76.1	73.2	76.5	69.6	68.4	67.3	75.1
Huang et al. (2019)	Wiki+MT	N	15	85.1	79.0	79.4	77.8	77.2	77.2	76.3	72.8	73.5	76.4	73.6	76.2	69.4	69.7	66.7	75.4
Devlin et al. (2018)	Wiki	N	102	82.1	73.8	74.3	71.1	66.4	68.9	69.0	61.6	64.9	69.5	55.8	69.3	60.0	50.4	58.0	66.3
Lample and Conneau (2019)	Wiki	N	100	83.7	76.2	76.6	73.7	72.4	73.0	72.1	68.1	68.4	72.0	68.2	71.5	64.5	58.0	62.4	71.3
Lample and Conneau (2019)	Wiki	1	100	83.2	76.7	77.7	74.0	72.7	74.1	72.7	68.7	68.6	72.9	68.9	72.5	65.6	58.2	62.4	70.7
XLM-R <sub>Base</sub>	CC	1	100	85.8	79.7	80.7	78.7	77.5	79.6	78.1	74.2	73.8	76.5	74.6	76.7	72.4	66.5	68.3	76.2
XLM-R	CC	1	100	<b>89.1</b>	<b>84.1</b>	<b>85.1</b>	<b>83.9</b>	<b>82.9</b>	<b>84.0</b>	<b>81.2</b>	<b>79.6</b>	<b>79.8</b>	<b>80.8</b>	<b>78.1</b>	<b>80.2</b>	<b>76.9</b>	<b>73.9</b>	<b>73.8</b>	<b>80.9</b>
<i>Translate everything to English and use English-only model (TRANSLATE-TEST)</i>																			
BERT-en	Wiki	1	1	88.8	81.4	82.3	80.1	80.3	80.9	76.2	76.0	75.4	72.0	71.9	75.6	70.0	65.8	65.8	76.2
RoBERTa	Wiki+CC	1	1	<b>91.3</b>	82.9	84.3	81.2	81.7	83.1	78.3	76.8	76.6	74.2	74.1	77.5	70.9	66.7	66.8	77.8
<i>Fine-tune multilingual model on each training set (TRANSLATE-TRAIN)</i>																			
Lample and Conneau (2019)	Wiki	N	100	82.9	77.6	77.9	77.9	77.1	75.7	75.5	72.6	71.2	75.8	73.1	76.2	70.4	66.5	62.4	74.2
<i>Fine-tune multilingual model on all training sets (TRANSLATE-TRAIN-ALL)</i>																			
Lample and Conneau (2019) <sup>†</sup>	Wiki+MT	1	15	85.0	80.8	81.3	80.3	79.1	80.9	78.3	75.6	77.6	78.5	76.0	79.5	72.9	72.8	68.5	77.8
Huang et al. (2019)	Wiki+MT	1	15	85.6	81.1	82.3	80.9	79.5	81.4	79.7	76.8	78.2	77.9	77.1	80.5	73.4	73.8	69.6	78.5
Lample and Conneau (2019)	Wiki	1	100	84.5	80.1	81.3	79.3	78.6	79.4	77.5	75.2	75.6	78.3	75.7	78.3	72.1	69.2	67.7	76.9
XLM-R <sub>Base</sub>	CC	1	100	85.4	81.4	82.2	80.3	80.4	81.3	79.7	78.6	77.3	79.7	77.9	80.2	76.1	73.1	73.0	79.1
XLM-R	CC	1	100	<b>89.1</b>	<b>85.1</b>	<b>86.6</b>	<b>85.7</b>	<b>85.3</b>	<b>85.9</b>	<b>83.5</b>	<b>83.2</b>	<b>83.1</b>	<b>83.7</b>	<b>81.5</b>	<b>83.7</b>	<b>81.6</b>	<b>78.0</b>	<b>78.1</b>	<b>83.6</b>

ภาพที่ 4 ผลความแม่นยำโดยเฉลี่ยของการจำแนกข้ามภาษา (Average accuracy results on cross-lingual classification) [18]

### 2.11 วิธีการปรับละเอียด เบิร์ตสำหรับการจำแนกข้อความ (How to Fine-Tune BERT for Text Classification?) [11]

งานวิจัยนี้นำเสนอการทดลองด้วยวิธีการที่ละเอียดถี่ถ้วนเพื่อตรวจสอบวิธีการปรับละเอียดเบิร์ต แบบต่าง ๆ ในงานประเภทของข้อความในบริบทต่าง ๆ และนำเสนอแนวทางวิธีการแก้ไขปรับละเอียด โมเดลเบิร์ต เพื่อให้เกิดประโยชน์สูงสุดสำหรับการใช้งานโดยประกอบไปด้วยสามขั้นตอนในการตรวจสอบวิธีการปรับแต่งดังต่อไปนี้ 1. ทำการเรียนรู้ล่วงหน้าของโมเดลเบิร์ตโดยใช้ข้อมูลที่เกี่ยวข้องกับการทรนงานหรือข้อมูลในโดเมน 2. ตัวเลือกในการปรับละเอียดโมเดลเบิร์ตกับงานที่หลากหลายที่มีความไม่เกี่ยวข้องกัน 3. การปรับละเอียดโมเดลเบิร์ตสำหรับงานหลักประกอบไปด้วยกระบวนการเตรียมข้อมูลของบทความที่มีความยาว การเลือกชั้น (layer selection) อัตราการเรียนรู้

ของแต่ละชั้น (layer wise learning rate), ปัญหาการเรียนรู้ของโมเดลและปรับปรุงความเข้าใจเกี่ยวกับชุดข้อมูลนั้นแต่ไม่สามารถจดจำสิ่งที่เรียนไปแล้ว (catastrophic forgetting) และปัญหาการเรียนรู้ของโมเดลด้วยทรัพยากรข้อมูลที่น้อย (low-shot learning problems.)

ผู้ศึกษาวิจัยได้ทำการวางกลยุทธ์สำหรับการปรับละเอียด 3 แนวทางดังต่อไปนี้

1. กลยุทธ์การปรับละเอียด (Fine Tuning Strategies) ในระดับชั้นที่ต่างกันของโครงข่ายประสาท (neural network) สามารถจับข้อมูลระดับวากยสัมพันธ์และความหมายที่แตกต่างกันได้ในการปรับเบิร์ตให้เข้ากับงานเป้าหมายจะต้องพิจารณาปัจจัยเหล่านี้

1.1 กระบวนการเตรียมขนาดความยาวของข้อความตั้งแต่ความยาวลำดับสูงสุดของเบิร์ตซึ่งก็คือ 512

1.2 การเลือกชั้น เนื่องจากพื้นฐานของโมเดลเบิร์ต ประกอบด้วยชั้นฝังข้อมูลชั้นเข้ารหัส 12 ชั้น (12-layer encoder) และชั้นพูล (pooling layer) ที่มีประสิทธิภาพสูงสุดสำหรับการจำแนกประเภทข้อความ

ปัญหาการเรียนรู้เกินขนาด (overfitting) เป็นปัญหาที่พบได้ในการฝึกโมเดล จึงต้องหาค่าที่เหมาะสมที่ดีกว่าพร้อมกับอัตราการเรียนรู้ที่เหมาะสมในการแก้ไขปัญหานี้

2. การเรียนรู้ล่วงหน้าเพิ่มเติม (Further Pre-training) โมเดลเบิร์ตเป็นโมเดลที่ถูกนำมาเทรนคลังข้อมูลภายในล่วงหน้าสำหรับในงานการจำแนกประเภทข้อความเฉพาะโดเมนนั้น ๆ เช่น บทวิจารณ์ภาพยนตร์ ซึ่งการกระจายกันของข้อมูลมีส่วนแตกต่างกันจากเบิร์ตแต่มันสามารถถูกนำมาสอนข้อมูลเพิ่มเติมกับโมเดลภาษาแบบถูกปิดและใช้ในงานการทำนายคำถัดไปจากลำดับของข้อความได้

3. การปรับละเอียดกับงานหลาย ๆ งานพร้อมกัน (Multi-Task Fine Tuning) เป็นการใช้ข้อมูลการสอนของงานแต่ละงาน แต่ใช้โมเดลเดียวกันในการทำงานปรับละเอียดงานหลาย ๆ งานพร้อมกันเป็นแนวทางที่มีประสิทธิภาพในการแบ่งปันความรู้ที่ได้รับจากงานภายใต้การดำเนินงานที่เกี่ยวข้อง

ผลการดำเนินงานวิจัยนี้ได้ทำการทดลองอย่างครอบคลุมเพื่อตรวจสอบแนวทางในการปรับละเอียดและพบว่าชั้นบนสุดของเบิร์ต มีประโยชน์มากสำหรับงานการจำแนกประเภทข้อความด้วยอัตราการลดลงของอัตราการเรียนรู้ของแต่ละชั้น เบิร์ตสามารถข้ามปัญหาการไม่สามารถจดจำสิ่งที่เรียนไปแล้วได้ การเรียนรู้ล่วงหน้าเพิ่มเติมสามารถเพิ่มประสิทธิภาพของผลลัพธ์ได้เป็นอย่างมาก การปรับละเอียด กับงานหลาย ๆ งานพร้อมกันก่อนการปรับละเอียดกับงานงานเดียวสามารถช่วยเพิ่มประสิทธิภาพในการปรับปรุงโมเดลได้และสุดท้ายโมเดลเบิร์ต สามารถปรับปรุงและพัฒนาางานที่มีทรัพยากรข้อมูลที่น้อยๆ ได้

## 2.12 หัวข้อในรูปแบบเบิร์ต : โมเดลภาษาประสาทที่ปรับปรุงตามหัวข้อซึ่งถูกปรับละเอียดสำหรับการจำแนกประเภทความรู้สึก (TopicBERT: A Topic-Enhanced Neural Language Model Fine-Tuned for Sentiment Classification) [19]

งานวิจัยนี้นำเสนอวิธีการพัฒนาหัวข้อในรูปแบบเบิร์ต โดยการปรับปรุงโมเดลเบิร์ต ด้วยการเพิ่มฟีเจอร์ (Feature) ในการจำแนกความรู้สึกของข้อความโดยใช้ข้อมูลเกี่ยวกับหัวข้อ (Topic) ที่เป็นคำสำคัญในบริบทของการเพิ่มประสิทธิภาพโมเดลในการจำแนกความรู้สึก ซึ่งผู้วิจัยได้ทำการพัฒนาโมเดลหัวข้อในรูปแบบเบิร์ตออกมาในสองรูปแบบคือ หัวข้อในรูปแบบเบิร์ต-เอทีพี (TopicBERT-ATP) และ หัวข้อในรูปแบบเบิร์ต-ทีเอ (TopicBERT-TA) หัวข้อในรูปแบบเบิร์ต-เอทีพี (aspect topic prediction, ATP) คือ การทำนายด้วยหัวข้อโดยใช้ข้อมูลที่ถูกเทรนจากนั้นนำวิธีการแอลดีเอ (Latent Dirichlet Allocation, LDA) เทคนิคในการแยกประเภทหัวข้อจากเอกสารหรือบริบทและการสุ่มตัวอย่างแบบกิบส์ (Gibbs sampling) หัวข้อในรูปแบบเบิร์ต-ทีเอ (topic augmentation, TA) คือ โมเดลภาษาที่ถูกเสริมด้วยหัวข้อ (Topic-Enhanced Neural Language Model) ที่ถูกนำมาใช้สำหรับงานการจัดการแบ่งแยกแ่งมุมของข้อความ (sentiment classification) โดยหัวข้อในรูปแบบเบิร์ต-ทีเอ นั้นมีการนำเอาโดเมนหรือแ่งมุมของข้อความมาใช้เพื่อเสริมความแม่นยำในการจัดแบ่งแ่งมุมด้วยการนำข้อมูลหัวข้อของโดเมนโดยตรงป้อนเข้าไปในชั้นของโมเดล การเพิ่มหัวข้อ (Topic Augmentation) ผู้วิจัยได้ทำการศึกษาพัฒนาโมเดลในการเพิ่มขีดความสามารถของโมเดลเพื่อทำความเข้าใจความหมายผ่านกระบวนการปรับละเอียดแทนที่จะเพิ่มความรู้อให้กับโมเดลและรูปแบบของการปรับละเอียดหัวข้อในรูปแบบเบิร์ตทั้งสองรูปแบบด้วยวิธีต่างๆ ส่งผลให้โมเดลหัวข้อในรูปแบบเบิร์ตมีข้อได้เปรียบที่แตกต่าง ๆ จากโมเดลเบิร์ต รุ่นอื่น ๆ ก่อนหน้านี้ทำให้โมเดลสามารถแยกแยะและเข้าใจความหมายของหัวข้อต่าง ๆ ได้ยิ่งขึ้นโดยตรงด้วยการป้อนเข้าไปในชั้นของโมเดลการเพิ่มหัวข้อ และเลือกใช้ชุดข้อมูลจาก SemEval 2014 Task 4 ชุดข้อมูลแรกเกี่ยวกับแล็ปท็อปและชุดข้อมูลเกี่ยวกับร้านอาหารลบป้ายกำกับ (Labels) ที่มีความขัดแย้งออกและระบุว่า เป็นบวก, ลบ และเป็นกลางตามบริบทเป็น 10 เปอร์เซนต์ของชุดข้อมูลเทรน

## 2.13 การศึกษาเปรียบเทียบการใช้การเรียนรู้เชิงลึกในการทำนายคำถัดไป (Next word prediction using deep learning: A comparative study)[15]

งานวิจัยฉบับนี้ได้มุ่งเน้นไปที่การศึกษาคำเปรียบเทียบการใช้การเรียนรู้เชิงลึกด้านการประมวลผลภาษาธรรมชาติโดยใช้เทคนิคต่าง ๆ เช่น หน่วยความจำระยะสั้นระยะยาว และหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง (BiLSTM) ในการสร้างแบบจำลองและวัดประสิทธิภาพของโมเดลด้วยค่าความถูกต้อง (accuracy) โดยใช้ชุดข้อมูลที่มีข้อความถึง 6508 บทความและเป็นชุดข้อมูลสาธารณะสามารถเข้าถึงได้จาก Kaggle จากนั้นใช้โมเดลหน่วยความจำระยะสั้นระยะยาว และ หน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง เพื่อการทำนายคำถัดไป

ในประโยค โดยการนำคำถัดไปเป็นการทำนายต่อเนื่องจากคำที่เคยบ่อนเข้าไปในโมเดล โดยหน่วยความจำระยะสั้นระยะยาวเป็นโมเดลประเภทโครงข่ายประสาทเทียมแบบวนซ้ำที่ออกแบบมาเพื่อแก้ปัญหาประมวลผลลำดับข้อมูลเฝ้าสังเกตการณ์ทำงานในแต่ละชั้นของหน่วยความจำระยะสั้นระยะยาว และหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง เป็นการรวม หน่วยความจำระยะสั้นระยะยาวที่มีการทำงานทั้งแบบไปข้างหน้าและข้างหลังเพื่อให้โมเดลมีความสามารถในการเข้าถึงข้อมูลทั้งจากหน้าไปหลังและจากหลังไปหน้า โดยการใช้หน่วยความจำระยะสั้นระยะยาวและหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง ในการเปรียบเทียบและวัดประสิทธิภาพของโมเดล



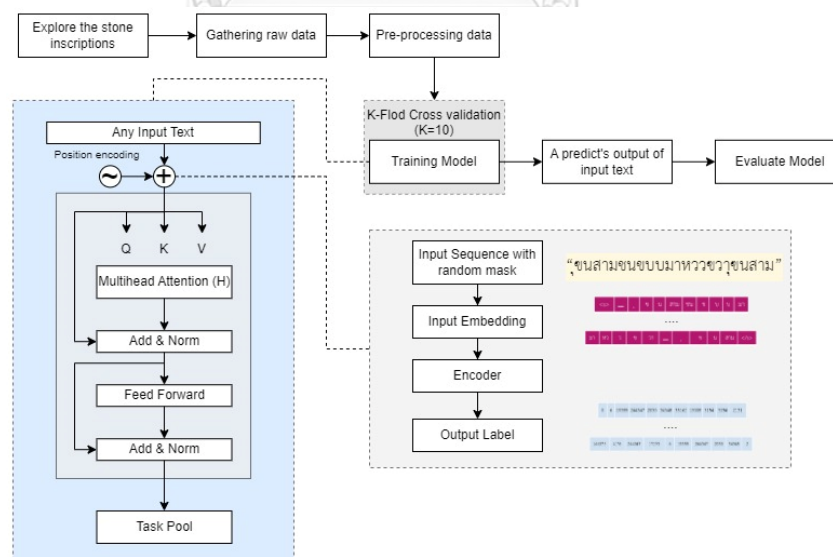
## บทที่ 3

### แนวคิดและวิธีการวิจัย

งานวิจัยฉบับนี้นำเสนอวิธีการปรับปรุงปัญหาข้อความที่ไม่สมบูรณ์หลังการถอดความภาษาโบราณจากศิลาจารึกเป็นภาษาไทยด้วยชุดฐานข้อมูลจารึกแห่งประเทศไทย ศูนย์มานุษยวิทยาสิรินธร โดยการนำเทคนิคการเรียนรู้เชิงลึกเข้ามาเป็นส่วนหนึ่งของการดำเนินงานวิจัย

#### 3.1 แนวทางการวิจัย

การดำเนินงานวิจัยภาษาไทยยุคสุโขทัยโบราณมีแนวคิดหลักคือการนำโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา โมเดลเบิร์ต-เบส-หลายภาษา-เคสต์ และ โมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสต์ ซึ่งเป็นโมเดลการเรียนรู้เชิงลึกที่ถูกพัฒนาต้นแบบมาจากโมเดลเบิร์ต มาใช้ในการประมวลผลภาษาสุโขทัยโบราณ เพื่อให้สามารถแยกแยะอักขระหรือคำศัพท์ และสร้างความเข้าใจในเนื้อหาของภาษาโบราณได้ดียิ่งขึ้น โดยที่ไม่จำเป็นต้องมีการสร้างโมเดลการเรียนรู้ล่วงหน้าตัวใหม่หรือการปรับแต่งโมเดลให้เหมาะสมกับภาษาโบราณ เนื่องจากพื้นฐานของโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา โมเดลเบิร์ต-เบส-หลายภาษา-เคสต์ และ โมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสต์ มีความสามารถในด้านการเรียนรู้และการถ่ายโอนข้อมูลของข้ามภาษาต่าง ๆ โดยในจะมีขั้นตอนในการดำเนินงานวิจัยดังภาพที่ 5



ภาพที่ 5 ขั้นตอนโดยสังเขปของกระบวนการดำเนินงาน

## 3.2 วิธีการวิจัย

### 3.2.1 การลงพื้นที่สำรวจหลักศิลาจารึก

ในการลงพื้นที่สำหรับการตรวจสอบและศึกษาลักษณะศิลาจารึกที่พิพิธภัณฑ์พระนคร จังหวัดกรุงเทพมหานคร จากรูปภาพที่ 6 ศิลาจารึกที่มีปัญหารอยขีดข่วนที่อาจส่งผลกระทบต่อความของนักโบราณคดี



ภาพที่ 6 หลักศิลาจารึกพ่อขุนรามคำแหง

### 3.2.2 ชุดข้อมูลภาษาโบราณในฐานข้อมูลจารึกแห่งประเทศไทยศูนย์มานุษยวิทยาสิรินธร

ชุดข้อมูลที่ถูกจัดเตรียมสำหรับศึกษางานวิจัยจะเป็นข้อความที่ถูกถอดความออกจากศิลาจารึกในช่วงของพุทธศักราชที่ 18 – 20 มีรายการดังต่อไปนี้

- 3.2.2.1 จารึกวัดป่ามะม่วง
- 3.2.2.2 จารึกนครชุม
- 3.2.2.3 จารึกพ่อขุนรามคำแหง
- 3.2.2.4 จารึกวัดศรีชุม
- 3.2.2.5 จารึกปู่ขุนจิตขุนจืด

### 3.2.3 กระบวนการเตรียมข้อมูลจากศิลาจารึก

กระบวนการนี้ผู้วิจัยจะต้องทำการคัดลอกข้อความจากไฟล์ PDF ต้นฉบับก่อนทำการเตรียมตัดคำ

### 3.2.3.1 คัดลอกข้อความจากต้นฉบับ

### 3.2.3.2 ทำการแบ่งช่องว่างของคำในประโยค

ตัวอย่างเช่น

ประโยคดั้งเดิม          บ้านฉอดอยู่อควรนอกทอนทอย

ประโยคก่อนแบ่งคำใหม่      พุกี ซี สรืน ทราทิ ทอย

ประโยคที่ควรจะเป็น          พุกี ซี สรืน ทราทิ ทอย

จากประโยคข้างต้น “ประโยคก่อนแบ่งคำใหม่” หากนำคำเหล่านี้เข้าไปฝึกโมเดลหรือตัดคำศัพท์จะทำให้ตัวโมเดลเรียนรู้คำศัพท์ที่ไม่มีความหมายตรงกับสิ่งที่ควรจะเป็น เช่น หากต้องการให้โมเดลแยกคำศัพท์หรือเรียนรู้ได้ถูกต้องจึงจำเป็นต้องเว้นช่องว่างระหว่างคำก่อนในข้อมูลไปทำในกระบวนการถัดไป

3.2.3.3 การเก็บรวบรวมคำศัพท์และรูปแบบของกลุ่มคำจากในประโยคเพื่อนำไปทำ custom dictionary ซึ่งคำศัพท์เหล่านี้จะไม่มีปรากฏในคลังคำศัพท์ (corpus) ภาษาไทยจึงจำเป็นต้องรวบรวมคำศัพท์ขึ้นมาใหม่

ตัวอย่างรูปแบบคำศัพท์เบื้องต้น

กนน = กัน

หหว = หัว

กท่า = กระท่า

ดอกไม้ต้นใหญ่ = ดอกไม้ต้นใหญ่

โดยคำเหล่านี้มักสังเกตได้ง่ายจากรูปแบบของคำศัพท์ดังเช่นตัวอย่างหากมีพยัญชนะขึ้นต้นและมีพยัญชนะตามหลังติดกันเหมือนกันทั้งสองตัวอักษรสามารถสันนิษฐานได้ว่า พยัญชนะตามหลังติดกันเหมือนกันทั้งสองตัวอักษร อ่านออกเสียงเป็นสระอะ ซึ่งเป็นคำประสมสระ

## 3.2.4 กระบวนการทรานฟอร์มเมอร์ข้อมูล

งานวิจัยนี้ได้นำเสนอโครงสร้างพื้นฐานของโมเดลที่มีความสำคัญสำหรับการประมวลผลข้อมูลของ NLP มีความสามารถในการจัดการปัญหาการเรียนรู้แบบ ลำดับ-ถึง-ลำดับด้วยการใช้ไลบรารี ทรานฟอร์มเมอร์ของ Hugging Face และ Pytorch ในการประมวลผลโมเดล

### 3.2.4.1 ไลบรารีที่จำเป็นต้องติดตั้งสำหรับการทำโมเดลมีดังต่อไปนี้

1. ตัวแบ่งคำ (tokenizers)
2. ชุดข้อมูล (datasets)
3. ทรานฟอร์มเมอร์
4. ทอร์ช (torch)

## 5. git-lfs

### 3.2.4.2 กระบวนการทำเตรียมข้อมูล (Preparing the dataset)

สำหรับกระบวนการนี้จำเป็นที่จะต้องใช้ชุดข้อมูลผ่านกระบวนการ 3.1.3 เพื่อทำการอัปโหลดข้อมูลเก็บไว้ที่ Hugging Face และเรียกใช้ชุดข้อมูลผ่านไลบรารี datasets ด้วยโค้ด 1 บรรทัดที่มีความรวดเร็วและประสิทธิภาพในการดำเนินงานสูง

### 3.2.4.3 การกำหนดขนาดของขนาดบล็อก (Determining the block size)

ขนาดบล็อก หมายถึงจำนวนของโทเคนที่ถูกนำไปสอนในแต่ละรุ่น (batch) หรือช่วงข้อมูล โดยปกติแล้วค่าของขนาดบล็อกจะถูกกำหนดเป็นค่าคงที่ที่กำหนดไว้ก่อนการสอนและตัวขนาดบล็อกจะเป็นตัวกำหนดขนาดของลำดับของที่สามารถรับอินพุตเข้าไปในโมเดลได้ในแต่ละครั้งจำนวนโทเคนที่มีในแต่ละภาษาอาจจะมีค่าแตกต่างกันไป ดังนั้น ขนาดบล็อกจึงถูกกำหนดเป็นค่าคงที่เท่ากับ 512 โดยไม่รวมเลขโทเคนของภาษาใด ๆ เพิ่มเติมเข้าไปในลำดับของโมเดลแต่จะต้องทำการตัดแบ่ง 8e (segment) แต่ละภาษาอย่างแยกกันในการสร้างรุ่นของข้อมูล

สำหรับงานวิจัยนี้ค่าของขนาดบล็อกจะถูกกำหนดอยู่ที่ 20 เนื่องจากความยาวของข้อความในบรรทัดมีจำนวนที่น้อยจึงจำเป็นที่จะต้องลดขนาดค่าคงที่ให้เล็กลงเพื่อให้ผลที่ได้จากการทำนายจะมีขนาดใกล้เคียงกับค่าที่ถูกสุ่มปิด

### 3.2.4.4 การกำหนดโมเดลเรียนรู้ล่วงหน้า [20]

การเลือกประเภทของโมเดลเรียนรู้ล่วงหน้ามีดังต่อไปนี้

1. การเลือกโมเดล งานวิจัยนี้ได้ใช้วิธีในการเลือกโมเดลจากความต้องการในการทำนายค่าที่ขาดหายไปจากข้อความซึ่งลักษณะของโมเดลที่เหมาะสมคือ โมเดลภาษาแบบถูกปิด
2. การเลือกโมเดลเรียนรู้ล่วงหน้า จำเป็นที่จะต้องรองรับภาษาไทยและตัวโมเดลที่เรียนรู้ล่วงหน้าควรจัดอยู่ในกลุ่มงานของการสุ่มปิดค่าเหมือนกัน โดยงานวิจัยนี้ได้ทำการเลือกโมเดลมาทั้งหมด 3 โมเดล ดังต่อไปนี้

2.1 เอ็กซ์แอลเอ็ม-โรเบอตา

2.2 เบิร์ต-เบส-หลายภาษา-เคสต์

2.3 ดิสทิลเบิร์ต-เบส-หลายภาษา-เคสต์

โดยทั้ง 3 โมเดล เป็นโมเดลที่รองรับการทำงานบนภาษาไทยและจัดอยู่ในงานประเภทของโมเดลภาษาแบบถูกปิด

### 3.2.4.5 การกำหนดเครื่องมือเก็บข้อมูล (Determining data collector) [21]

เครื่องมือเก็บข้อมูล คือ วัตถุ (objects) หรือ โมดูล (module) ที่ใช้ในการจัดการข้อมูลที่ได้จากการประมวลผลเพื่อให้สามารถนำข้อมูลเข้าโมเดลได้อย่างถูกต้องและเพื่อให้สามารถสร้าง



ขนาดรุ่น (batch size) ตามที่กำหนดได้ เครื่องมือเก็บข้อมูลจะทำการประมวลผลในการเติมช่องว่าง (padding) สุ่มปิดคำด้วยมาสกโทเคนตามอัตราส่วนที่กำหนดและทำการตัดแบ่งคำ

#### 3.2.4.6 กระบวนการแบ่งข้อมูลด้วยการตรวจสอบแบบข้ามและสอนโมเดล

งานวิจัยฉบับนี้นำเสนอวิธีการประเมินประสิทธิภาพการทำงานในทางสถิติที่มีความสเถียรในการแยกส่วนของข้อมูลเพื่อนำไปสอนและทดสอบด้วยการแบ่งชุดข้อมูลออกเป็นพับที่  $k$  เพื่อทำการสอนและทดสอบแบบข้ามดังต่อไปนี้ ทำการแบ่งชุดข้อมูลออกเป็น สอน:ทดสอบ ที่อัตราส่วนของ 90:10 โดยกำหนดค่าที่  $k = 10$  จากนั้นเทรนโมเดล โดยมีขั้นตอนต่อไปนี้

1. กำหนดค่าและพารามิเตอร์ (parameter) ต่าง ๆ โดยเรียกใช้ไลบรารี training\_args [18] จาก Hugging Face Transformers จากนั้นจึงกำหนดค่าพารามิเตอร์ต่าง ๆ เช่น อัตราการเรียนรู้ (learning rate) หรือ รอบจำนวนการฝึก (epoch) เป็นต้น
2. เขียนฟังก์ชันสำหรับใช้ในการเทรนโมเดลด้วยชุดข้อมูลแบบสำหรับสอนและทดสอบ หลังการสอนโมเดลจบในแต่ละรอบโมเดลจะถูกจัดเก็บบนฮับ (hub) ของ Hugging Face เสมอ เพื่อให้ง่ายต่อการเรียกโมเดลมาใช้
3. เขียนฟังก์ชันสำหรับใช้ในการแบ่งชุดข้อมูลออกเป็น  $k$  ส่วนมีค่าเท่ากับ 10 เพื่อเรียกใช้ฟังก์ชันการเทรนโมเดลจากข้อที่ 2.
4. เรียกใช้ฟังก์ชันในข้อที่ 3 เพื่อสอนโมเดล

#### 3.2.5 กระบวนการทำนายคำถัดไปด้วยหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง

งานวิจัยนี้ได้นำเสนอโครงสร้างพื้นฐานของโมเดลที่มีความสำคัญสำหรับการประมวลผลข้อมูลของการประมวลผลภาษาธรรมชาติ มีความสามารถในการจัดการปัญหาการเรียนรู้แบบ ลำดับ-ถึง-ลำดับ แบบสองทิศทางด้วยการใช้ไลบรารี KERAS และ TensorFlow ในการประมวลผลโมเดล

##### 3.2.5.1 ไลบรารีที่จำเป็นต้องติดตั้งสำหรับการทำโมเดลมีดังต่อไปนี้

1. Tensorflow
2. keras

##### 3.2.5.2 กระบวนการทำเตรียมข้อมูล (Preparing the dataset)

กระบวนการดำเนินงานอ้างอิงข้อ 3.2.4.2

##### 3.2.5.3 กระบวนการแบ่งข้อมูลด้วยการตรวจสอบแบบข้ามและสอนโมเดล

งานวิจัยฉบับนี้นำเสนอวิธีการประเมินประสิทธิภาพการทำงานในทางสถิติที่มีความสเถียรในการแยกส่วนของข้อมูลเพื่อนำไปสอนและทดสอบด้วยการแบ่งชุดข้อมูลออกเป็นพับที่  $k$  เพื่อทำการสอนและทดสอบแบบข้ามดังต่อไปนี้ ทำการแบ่งชุดข้อมูลออกเป็น สอน:ทดสอบ ที่อัตราส่วนของ 90:10 โดยกำหนดค่าที่  $k = 10$  จากนั้นเทรนโมเดล โดยมีขั้นตอนต่อไปนี้

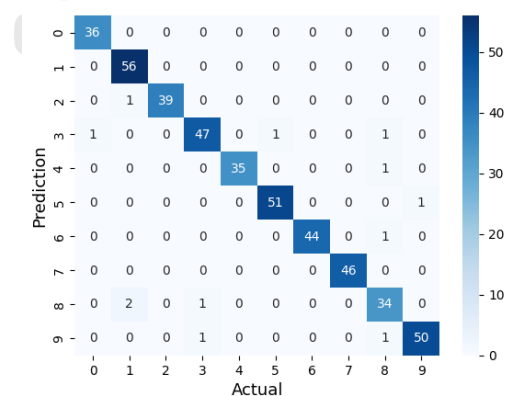
1. กำหนดค่าและพารามิเตอร์ ต่าง ๆ โดยเรียกใช้ไลบรารี KERAS และ TensorFlow จากนั้นจึงกำหนดค่าพารามิเตอร์ ต่าง ๆ เช่น Embedding, SpatialDropout1D หรือ Dense เป็นต้น
2. เขียนฟังก์ชันสำหรับใช้ในการเทรนโมเดลด้วยชุดข้อมูลแบบสำหรับสอนและทดสอบ หลังการเทรนโมเดลจบในแต่ละรอบโมเดลจะถูกบันทึกเข้าที่จัดเก็บเสมอ เพื่อให้ง่ายต่อการเรียกโมเดลมาใช้
3. เขียนฟังก์ชันสำหรับใช้ในการแบ่งชุดข้อมูลออกเป็น k ส่วนมีค่าเท่ากับ 10 เพื่อเรียกใช้ฟังก์ชันการเทรนโมเดลจากข้อที่ 2.
4. เรียกใช้ฟังก์ชันในข้อที่ 3 เพื่อเทรนโมเดล

### 3.2.6 ระเบียบและวิธีในการทดสอบโมเดล [22]

งานวิจัยชุดนี้ได้ใช้วิธีการสุ่มปิดค่าด้วยมาสก์โทเคนเพื่อจำลองวิธีการทำนายค่าที่ขาดหายไปจากประโยคและวัดประสิทธิภาพความถูกต้องด้วยวิธีการใช้เมทริกซ์คอนฟิวชัน (Confusion matrix) ในการประเมินสมรรถนะความถูกต้องของโมเดลแบบการจำแนกประเภทหลายคลาส (Multiclass Classification) และสามารถแสดงผลลัพธ์ได้หลากหลายรูปแบบดังเช่น ภาพที่ 7 และ ภาพที่ 8

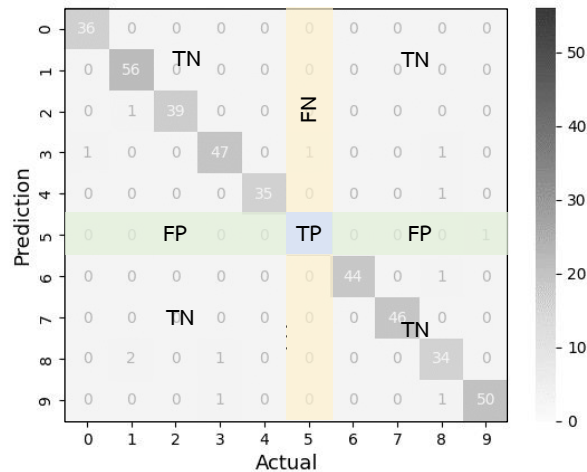
```
[[ [ 0 452 260 ... 4 9 1]
 [ 0 0 0 ... 0 0 0]
 [ 0 0 0 ... 0 0 0]
 ...
 [ 0 0 0 ... 0 0 0]
 [ 0 1 0 ... 0 0 0]
 [ 0 0 0 ... 0 0 0]]
```

ภาพที่ 7 เมทริกซ์คอนฟิวชันแบบการจำแนกประเภทหลายคลาสในรูปแบบอาร์เรย์



ภาพที่ 8 เมทริกซ์คอนฟิวชันแบบการจำแนกประเภทหลายคลาส

จากภาพที่ 9 ในแต่ละคอลัมน์จะแสดงป้ายกำกับที่ถูกต้องและในแต่ละแถวจะแสดงป้ายกำกับที่ถูกทำนายออกมาข้อมูลที่อยู่ในแถวจะแสดงตามลำดับของป้ายชื่อและในตำแหน่งสุดท้ายในแต่ละแถวจะแสดงผลลัพธ์ของตัวอย่างที่สามารถแสดงผลออกเป็นแผนภาพที่เข้าใจได้ง่ายในภาพที่ 10



ภาพที่ 9 จำแนกค่าความถูกต้องจากเมทริกซ์คอนฟิวชัน

True positive (TP) คือจำนวนข้อมูลที่จำแนกถูกว่าเป็นคำตอบถูกต้อง  
 True negative (TN) คือจำนวนข้อมูลที่จำแนกถูกว่าเป็นคำตอบที่ไม่ถูกต้อง  
 False positive (FP) คือจำนวนข้อมูลที่จำแนกผิดว่าเป็นคำตอบถูกต้องซึ่งคำตอบที่แท้จริงเป็นคำตอบที่ไม่ถูกต้อง  
 False negative (FN) คือจำนวนข้อมูลที่จำแนกผิดว่าเป็นคำตอบที่ไม่ถูกต้องซึ่งคำตอบที่แท้จริงเป็นคำตอบถูกต้อง

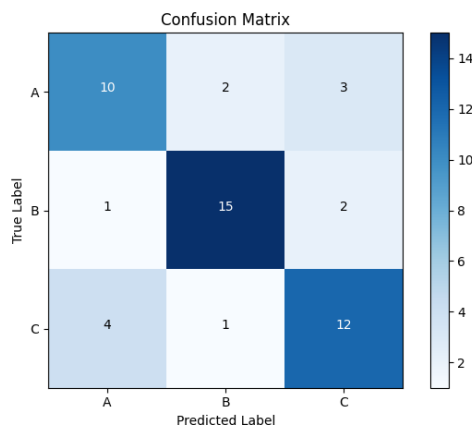
$$\text{ความแม่นยำ (Precision)} = \frac{TP}{TP+FP} \quad (3.1)$$

$$\text{ความระลึก (Recall)} = \frac{TP}{TP+FN} \quad (3.2)$$

$$\text{ความถูกต้อง (Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.3)$$

$$\text{คะแนน F1 (F1-score)} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.4)$$

ตัวอย่างที่ 3-1 วิธีการคำนวณค่าจากเมทริกซ์คอนฟิวชันจากการทำนายตัวอักษร 3 ตัว



ภาพที่ 10 ตัวอย่างเมทริกซ์คอนฟิวชันแบบการจำแนกประเภทหลายคลาส

การคำนวณคลาส “A” อ้างอิงข้อมูลจากภาพที่ 10

$$TP = 10$$

$$FN = 2+3 = 5$$

$$FP = 1+4 = 5$$

$$TN = 15+2+1+12 = 30$$

$$\text{ความแม่นยำ} = \frac{TP}{TP+FP} = \frac{10}{10+5} = 0.667$$

$$\text{ความถูกต้อง} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{10+30}{10+30+5+5} = 0.8$$

$$\text{ความระลึก} = \frac{TP}{TP+FN} = \frac{10}{10+5} = 0.667$$

$$\text{คะแนน F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{0.667 \times 0.667}{0.667 + 0.667} = 0.667$$

ซึ่งสามารถคำนวณคลาสอื่น ๆ โดยใช้วิธีเดียวกันกับ การคำนวณคลาส “A”

ในงานการจำแนกประเภทแบบหลากหลาย (multi-class) สามารถสังเกตค่าความแม่นยำสำหรับแต่ละคลาสได้ในขณะที่เราต้องการประเมินโมเดลที่เป็นการจำแนกประเภทแบบหลากหลายด้วยตัวชี้วัดสากล (one global metric) ซึ่งแบ่งวิธีคำนวณค่าความแม่นยำ 3 ประเภท คือ ไมโคร (micro) มาโคร (macro) และ ถ่วงน้ำหนัก (weighted) ซึ่งสามารถใช้ชี้วัดผลรวมของงานประเภทการจำแนกประเภทแบบหลากหลายได้ดังต่อไปนี้

ค่าความแม่นยำแบบไมโครถูกคำนวณโดยพิจารณาค่าผลรวมของ TP, TN, FN และ TN โดยไม่คำนึงถึงคลาส

$$TP_{micro} = TP_A + TP_B + TP_C = 10+15+12 = 37$$

$$FP_{micro} = FP_A + FP_B + FP_C = (1+4) + (2+1) + (3+2) = 13$$

$$Micro\ Precision = \frac{TP_{micro}}{TP_{micro}+FP_{micro}} = 37 / (37+13) = 0.74$$

ค่าความแม่นยำแบบมาโคร อ้างอิงการคำนวณค่าเฉลี่ยแบบไม่ได้ถ่วงน้ำหนักของตัวชี้วัดในแต่ละคลาส อ้างอิงตัวอย่างผลลัพธ์ของการจำแนกประเภทดังภาพที่ 11

	precision	recall	f1-score	support
A	<b>0.67</b>	0.67	0.67	15
B	<b>0.56</b>	0.56	0.56	18
C	<b>0.53</b>	0.53	0.53	17
accuracy			0.58	50
macro avg	0.58	0.58	0.58	50
weighted avg	0.58	0.58	0.58	50

ภาพที่ 11 รายงานการจำแนกประเภทแบบมาโคร ของ A, B และ C

โดยค่าความแม่นยำของคลาส A, B และ C ใช้วิธีการมาจาก การคำนวณคลาส “A”

$$Macro\ Precision = \frac{Precision_A + Precision_B + Precision_C}{3}$$

$$= (0.67+0.56+0.53)/3 = 0.58$$

ค่าความแม่นยำแบบถ่วงน้ำหนักใช้การคำนวณค่าเฉลี่ยแบบถ่วงน้ำหนักของตัวชี้วัดในแต่ละคลาส

$$Weight\ Precision = \frac{support_1 \times precision_1 + \dots + support_n \times precision_n}{\sum support}$$

$$= \frac{15 \times 0.67 + 18 \times 0.56 + 17 \times 0.53}{50}$$

$$= 0.58$$

### 3.2.6.1 การประเมินผลลัพธ์ความถูกต้องของโมเดลการจำแนกประเภทแบบ

หลากหลาย

แบ่งวิธีการเป็น 4 วิธี และอ้างอิงวิธีการมาจากจำนวนความถี่ที่ถูกรับว่าอักขระหายไปจากตำแหน่ง ๆ นั้น จากตารางที่ 1 ดังต่อไปนี้

ปิด 1 ตำแหน่ง: <mask> คำโดยการสุ่ม 1 character

ปิด 2 ตำแหน่งติดกัน: <mask> คำโดยการสุ่ม 2 character โดยที่ตำแหน่งของ <mask> จะต้องติดกันทั้งสองตำแหน่ง

ปิด 5 ตำแหน่งติดกัน: <mask> คำโดยการสุ่ม 5 character โดยที่ตำแหน่งของ <mask> จะต้องติดกันทั้งสองตำแหน่ง

ปิด 2 ตำแหน่งไม่ติดกัน: <mask> คำโดยการสุ่ม 2 character โดยที่ตำแหน่งของ <mask> จะต้องไม่ติดกันทั้งสองตำแหน่ง

ตารางที่ 1 ความถี่ที่ถูกรับว่ามีตำแหน่งขาดหายไปจากประโยค

ความถี่ (ตำแหน่ง)	1	2	3	4	5	6	7	8
จำนวนทั้งหมด	41	44	15	12	9	7	3	1
	2 <sup>nd</sup>	1 <sup>st</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>			

ตัวอย่าง รูปแบบในการประเมินผลลัพธ์

ปิด 1 ตำแหน่ง: ชนึ่เก็ลออนเซ<mask>ไฟร์ฝ่าหน้าใส

ปิด 2 ตำแหน่งติดกัน: ูกโต<mask><mask>มากส์หมากหวาน

ปิด 5 ตำแหน่งติดกัน: ูกโต<mask><mask><mask><mask><mask>ส์หมากหวาน

ปิด 2 ตำแหน่งไม่ติดกัน: <mask>ูกพร่าบ<mask>อแก็พ

### 3.2.7 กระบวนการปรับปรุงประสิทธิภาพในการประเมินผลลัพธ์

#### 3.2.7.1 การวิเคราะห์ผลลัพธ์ของแต่ละโมเดล

ในกระบวนการนี้จะเป็นการนำผลการประเมินที่โมเดลได้ทำนายอักขระออกมาทำการวิเคราะห์เพื่อตรวจสอบว่าตำแหน่งที่ทายผิดอยู่ตรงตำแหน่งใดของคำที่ถูกปิดด้วยโทเคน

#### 3.2.7.2 การเพิ่มคำศัพท์เฉพาะเข้าไปในตัวแบ่งคำ

1. คำศัพท์ - ปรับปรุงคำศัพท์เฉพาะเพิ่มเติม
2. การตัดคำ - แก้ไขปัญหาของคำศัพท์และประโยคที่ถูกแบ่งคำผิด

#### 3.2.7.3 การกรองและคัดแยกคำศัพท์ในตัวแบ่งคำ

1. กรองโทเคนที่มีตัวอักษรมากกว่า 1 ตัว หรือสัญลักษณ์ออกจากตัวแบ่งคำ
2. กรองโทเคนที่ไม่ใช่ภาษาไทยออกจากตัวแบ่งคำ

#### 3.2.7.4 การประยุกต์ใช้คลังคำศัพท์เฉพาะในการทำนาย

1. โมเดลจะนำอักขระที่ได้จากโทเคนทั้งหมดไปทำการจับคู่ตามจำนวนโทเคน
2. โมเดลนำคำที่ถูกจับคู่ไปค้นหาในคลังคำศัพท์เฉพาะ

3. โมเดลเลือกค่าที่มีอยู่ในคลังคำศัพท์เฉพาะที่มีคะแนนสูงสุด

### 3.2.8 การเปรียบเทียบสมรรถนะของโมเดล (model's performance benchmarking)

การเปรียบเทียบประสิทธิภาพของโมเดลของงานวิจัยชุดนี้ จะใช้ผลการวัดจากเมตริกซ์เมทริกซ์คอนฟิวชันในการทดลองดังต่อไปนี้

1. การเปรียบเทียบสมรรถนะของโมเดลด้วยการใช้กลุ่มโมเดลเรียนรู้ล่วงหน้าและอื่นๆ ดังต่อไปนี้
  - 1.1 โมเดลเอ็กซ์แอลเอ็ม-โรเบอตา
  - 1.2 เบิร์ต-เบส-หลายภาษา-เคสต์
  - 1.3 ดิสทิลเบิร์ต-เบส-หลายภาษา-เคสต์
  - 1.4 หน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง
2. การวัดค่าความถูกต้องในการทำนายผลลัพธ์จากโมเดลในข้อที่ 1 ทั้งก่อนและหลังกระบวนการปรับปรุงประสิทธิภาพโดยใช้ระเบียบวิธีการวัดผลอ้างอิงในบทที่ 3.2.6.1 การประเมินผลลัพธ์ความถูกต้องของโมเดลการจำแนกประเภทแบบหลากหลาย

## บทที่ 4

### การทดลองและอภิปรายผลการทดลอง

ผลการทดลองของงานวิจัยนี้แบ่งออกเป็น 2 ส่วนหลัก คือ (1) ผลการการวัดประสิทธิภาพ ความถูกต้องของโมเดลจากเมตริกซ์คอนฟิวชันและผลจากรายงานการจำแนกประเภท (Classification report) (2) ผลลัพธ์ความถูกต้องในการทำนายผลลัพธ์จากโมเดลทั้งก่อนและหลังปรับปรุงกระบวนการทำนาย

#### 4.1 ผลการการวัดประสิทธิภาพความถูกต้องของโมเดล

งานวิจัยชุดนี้ได้ใช้วิธีการวัดประสิทธิภาพของโมเดลด้วยวิธีการแบ่งข้อมูลออกเป็น 10 ส่วนมีขนาดเท่า ๆ กัน ในแต่ละส่วนของข้อมูลที่แบ่งออกมาจะถูกเรียกว่าพับและในแต่ละพับจะถูกแบ่งชุดข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลสำหรับการสอนและสำหรับทดสอบคิดอัตราส่วนเป็น 90:10 เปอร์เซ็นต์จากชุดข้อมูลตามลำดับ โดยที่ชุดข้อมูลทีนำมาใช้ในการทดสอบทั้ง 10 ส่วนจะมีข้อมูลไม่ซ้ำกันในแต่ละพับ

ข้อกำหนดในการวัดผล

1. แสดงค่าความถูกต้องจากข้อมูลทั้ง 10 ส่วน
2. การเปรียบเทียบประสิทธิภาพของโมเดลในกลุ่มการเรียนรู้ล่วงหน้ากับโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง

#### 4.1.1 รายงานผลการทดลอง

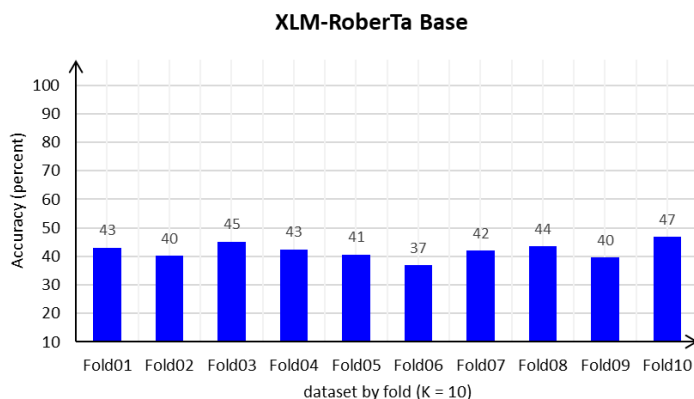
##### 4.1.1.1 รายงานผลการทดลองกลุ่มโมเดลเรียนรู้ล่วงหน้า

1. ผลการทดลองของโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา

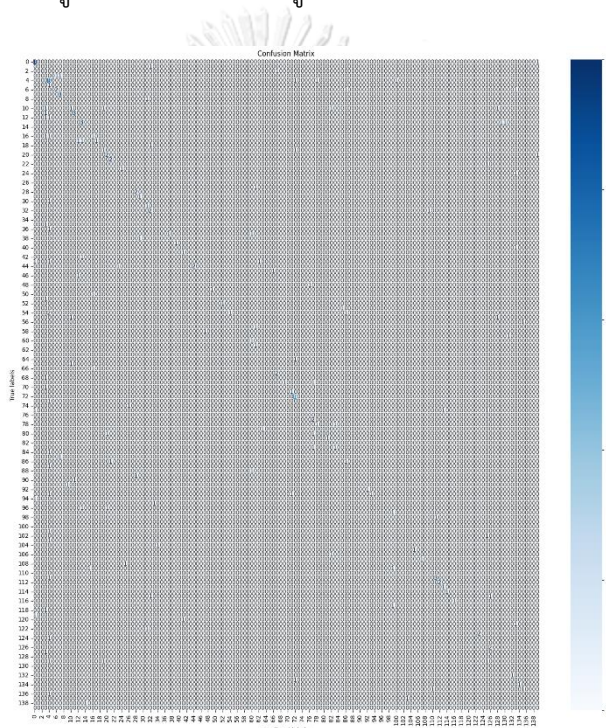
ตารางที่ 2 ผลลัพธ์ค่าความถูกต้องของโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา ทั้ง 10 พับ

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
43	40	45	43	41	37	42	44	40	47





ภาพที่ 12 แผนภูมิแท่งแสดงค่าความถูกต้องของโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา



ภาพที่ 13 เมทริกซ์คอนฟิวชันแบบการจำแนกประเภท 139 คลาส ของโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา

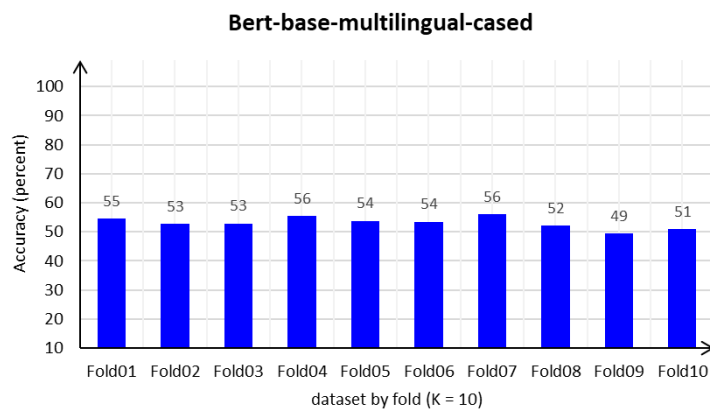
จากผลการทดลองพบว่าโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา มีค่าความถูกต้องที่ดีที่สุดอยู่ที่ร้อยละ 47, ค่าความแม่นยำในการทำนายอักขระได้ถูกต้องอยู่ที่ร้อยละ 43, ความน่าจะเป็นที่โมเดลสามารถทำนายคำออกมาถูกอยู่ที่ร้อยละ 47 และมีค่าคะแนน F1 ร้อยละ 42 อ้างอิงภาพที่ 12 และจากภาพที่ 13 ในแต่ละคอลัมน์ (แกน Y) จะแสดงป้ายกำกับที่ถูกต้องและในแต่ละแถว (แกน X) จะแสดงป้ายกำกับที่ถูกทำนายออกมาข้อมูลที่อยู่ในแถวจะแสดงตามลำดับของป้ายชื่อและในตำแหน่งสุดท้ายในแต่ละแถวจะแสดงผลลัพธ์ ตัวอย่างเช่นป้ายกำกับหมายเลขที่ 4 จำนวนข้อมูลที่ทายคำตอบที่ถูกว่าเป็นคำตอบถูกต้อง ( $TP_4$ ) มีค่าเท่ากับ 8 จำนวนข้อมูลที่ทายคำตอบว่าเป็นคำตอบที่ไม่ถูกต้องซึ่ง

คำตอบที่แท้จริงเป็นคำตอบถูกต้อง ( $FN_d$ ) มีค่าเท่ากับ 3 จำนวนข้อมูลที่ทายคำตอบว่าถูกต้องซึ่งคำตอบที่แท้จริงเป็นคำตอบไม่ถูกต้อง ( $FP_d$ ) มีค่าเท่ากับ 18 และจำนวนข้อมูลที่ทายถูกว่าเป็นเป็นคำตอบถูกต้อง ( $TN_d$ ) มีค่าเท่ากับ 178

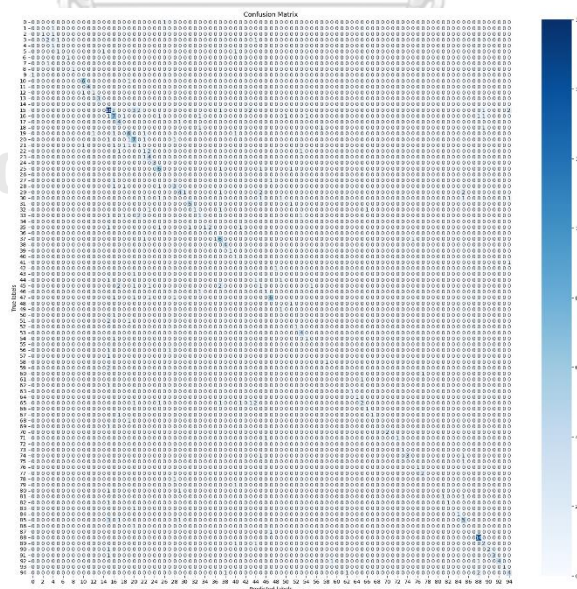
2. ผลการทดลองของโมเดลเบิร์ต-เบส-หลายภาษา-เคสดี

ตารางที่ 3 ผลลัพธ์ค่าความถูกต้องของโมเดลเบิร์ต-เบส-หลายภาษา-เคสดี ทั้ง 10 พับ

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
55	53	53	56	54	54	56	52	49	51



ภาพที่ 14 แผนภูมิแท่งแสดงค่าความถูกต้องของโมเดลเบิร์ต-เบส-หลายภาษา-เคสดี



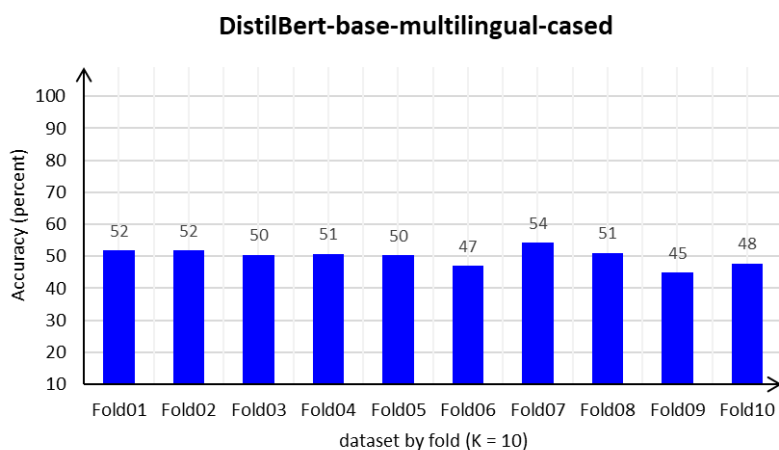
ภาพที่ 15 เมทริกซ์คอนฟิวชันแบบการจำแนกประเภท 94 คลาส ของโมเดลเบิร์ต-เบส-หลายภาษา-เคสดี

จากผลการทดลองพบว่าโมเดลเบิร์ต-เบส-หลายภาษา-เคสดี มีค่าความถูกต้องที่ดีที่สุดอยู่ที่ร้อยละ 56, ค่าความแม่นยำในการทำนายอักขระได้ถูกต้องอยู่ที่ร้อยละ 53, ความน่าจะเป็นที่โมเดลสามารถทำนายคำออกมาถูกอยู่ที่ร้อยละ 55 และมีค่าคะแนน F1 ร้อยละ 52 อ่างอิงภาพที่ 14 และจากภาพที่ 15 ในแต่ละคอลัมน์ (แกน Y) จะแสดงป้ายกำกับที่ถูกต้องและในแต่ละแถว (แกน X) จะแสดงป้ายกำกับที่ถูกทำนายออกมาข้อมูลที่อยู่ในแถวจะแสดงตามลำดับของป้ายชื่อและในตำแหน่งสุดท้ายในแต่ละแถวจะแสดงผลลัพธ์ ตัวอย่างเช่นป้ายกำกับหมายเลขที่ 15 จำนวนข้อมูลที่ทายคำตอบที่ถูกว่าเป็นคำตอบถูกต้อง ( $TP_{15}$ ) มีค่าเท่ากับ 16 จำนวนข้อมูลที่ทายคำตอบว่าเป็นคำตอบที่ไม่ถูกต้องซึ่งคำตอบที่แท้จริงเป็นคำตอบถูกต้อง ( $FN_{15}$ ) มีค่าเท่ากับ 13 จำนวนข้อมูลที่ทายคำตอบว่าถูกต้องซึ่งคำตอบที่แท้จริงเป็นคำตอบไม่ถูกต้อง ( $FP_{15}$ ) มีค่าเท่ากับ 20 และจำนวนข้อมูลที่ทายถูกว่าเป็นคำตอบถูกต้อง ( $TN_{15}$ ) มีค่าเท่ากับ 241

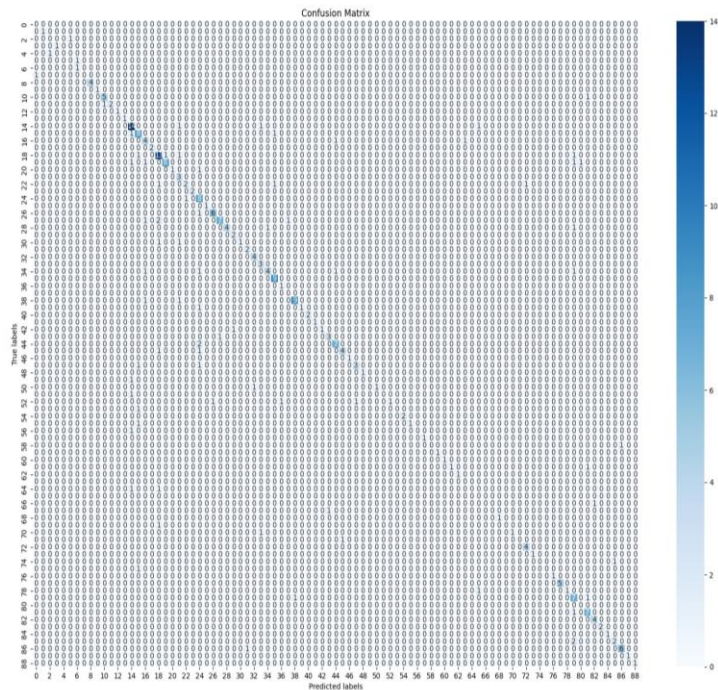
### 3. ผลการทดลองของโมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสดี

ตารางที่ 4 ผลลัพธ์ค่าความถูกต้องของโมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสดี ทั้ง 10 พับ

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
52	52	50	51	50	47	54	51	45	48



ภาพที่ 16 แผนภูมิแท่งแสดงค่าความถูกต้องของโมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสดี



ภาพที่ 17 เมทริกซ์คอนฟิวชันแบบการจำแนกประเภท 88 คลาส ของโมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสดี

จากผลการทดลองพบว่าโมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสดี มีค่าความถูกต้องที่ดีที่สุดอยู่ที่ร้อยละ 52, ค่าความแม่นยำในการทำนายอักษรได้ถูกต้องอยู่ที่ร้อยละ 55, ความน่าจะเป็นที่โมเดลสามารถทำนายคำออกมาถูกต้องอยู่ที่ร้อยละ 52 และมีค่าคะแนน F1 ร้อยละ 50 อ้างอิงภาพที่ 16 และจากภาพที่ 17 ในแต่ละคอลัมน์ (แกน Y) จะแสดงป้ายกำกับที่ถูกต้องและในแต่ละแถว (แกน X) จะแสดงป้ายกำกับที่ถูกทำนายออกมาข้อมูลที่อยู่ในแถวจะแสดงตามลำดับของป้ายชื่อและในตำแหน่งสุดท้ายในแต่ละแถวจะแสดงผลลัพธ์ ตัวอย่างเช่นป้ายกำกับหมายเลขที่ 14 จำนวนข้อมูลที่ทายคำตอบที่ถูกว่าเป็นคำตอบถูกต้อง ( $TP_{14}$ ) มีค่าเท่ากับ 14 จำนวนข้อมูลที่ทายคำตอบว่าเป็นคำตอบที่ไม่ถูกต้องซึ่งคำตอบที่แท้จริงเป็นคำตอบถูกต้อง ( $FN_{14}$ ) มีค่าเท่ากับ 3 จำนวนข้อมูลที่ทายคำตอบว่าถูกต้องซึ่งคำตอบที่แท้จริงเป็นคำตอบไม่ถูกต้อง ( $FP_{14}$ ) มีค่าเท่ากับ 6 และจำนวนข้อมูลที่ทายถูกว่าเป็นคำตอบถูกต้อง ( $TN_{14}$ ) มีค่าเท่ากับ 241

#### 4.1.1.2 รายงานผลการทดลองโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง

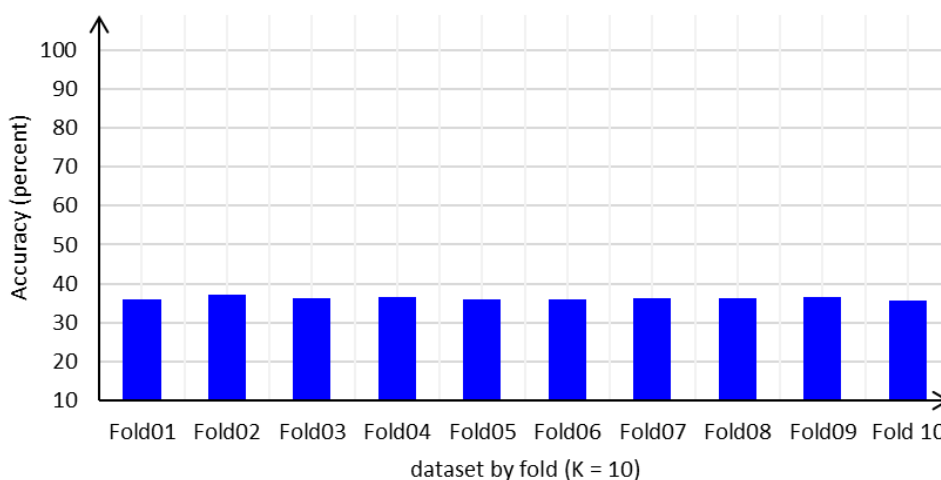
ผลการทดลองของโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง



ตารางที่ 5 ผลลัพธ์ค่าความถูกต้องของโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง ทั้ง 10 พับ

FOLD 1	FOLD 2	FOLD 3	FOLD 4	FOLD 5	FOLD 6	FOLD 7	FOLD 8	FOLD 9	FOLD 10
35.91	37.21	36.24	36.54	35.93	35.95	36.14	36.31	36.52	35.61

### Bidirectional LSTM



ภาพที่ 18 แผนภูมิแท่งแสดงค่าความถูกต้องของโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง

จากผลการทดลองพบว่าโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทางมีค่าความถูกต้องที่ดีที่สุดอยู่ที่ร้อยละ 37.21 ค่าความแม่นยำในการทำนายอักขระได้ถูกต้องอยู่ที่ร้อยละ 36 ความน่าจะเป็นที่โมเดลสามารถทำนายคำออกมาถูก (Recall) อยู่ที่ร้อยละ 36 และมีค่าคะแนน F1 ร้อยละ 35 อ้างอิงภาพที่ 18

#### 4.1.2 ผลการเปรียบเทียบสมรรถนะของโมเดลในกลุ่มโมเดลเรียนรู้ล่วงหน้าทีละและโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง

ตารางที่ 6 การเปรียบเทียบสมรรถนะของโมเดลทั้ง 4 โมเดล

โมเดล	ค่าความถูกต้อง +- ค่าเบี่ยงเบนมาตรฐาน (หน่วย เปอร์เซ็นต์)
เอ็กซ์แอลเอ็ม-โรเบอตา	42 +- 4.65
เบิร์ด-เบส-หลายภาษา-เคสดี	53 +- 3.52

โมเดล	ค่าความถูกต้อง +- ค่าเบี่ยงเบนมาตรฐาน (หน่วย เปอร์เซนต์)
ดิสทิลเบิร์ต-เบส-หลายภาษา-เคสต์	50 +- 3.74
หน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง	36 +- 4.29

จากตารางที่ 6 ผลการทดลองการนำโมเดลเรียนรู้ล่วงหน้าแบบหลากหลายภาษาทั้ง 3 โมเดลมาเรียนรู้เพื่อให้โมเดลมีความเข้าใจในเนื้อหาของภาษาสุโขทัยพบว่าโมเดลเบิร์ต-เบส-หลายภาษา-เคสต์มีค่าความถูกต้องที่ดีที่สุดอยู่ที่ร้อยละ 53 +- 3.52 และโมเดลมีค่าความถูกต้องที่ใกล้เคียงกับโมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสต์ และในการนำหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทางมาประยุกต์ในการทำนายคำที่ถูกสุ่มปิดพบว่าโมเดลมีค่าความถูกต้องในการทำนายน้อยที่สุดอยู่ที่ร้อยละ 36 +- 4.29

#### 4.2 ผลลัพธ์ความถูกต้องในการทำนายผลลัพธ์จากโมเดลทั้งก่อนและหลังปรับปรุงกระบวนการทำนาย

การประเมินผลลัพธ์ความถูกต้องของโมเดลทั้ง 4 ประเภทอ้างอิงรายละเอียดจาก 3.2.6.1 การประเมินผลลัพธ์ความถูกต้องของโมเดลการจำแนกประเภทแบบหลากหลาย ดังต่อไปนี้

**ปิด 1 ตำแหน่ง:** <mask> คำโดยการสุ่ม 1 character

**ปิด 2 ตำแหน่งติดกัน:** <mask> คำโดยการสุ่ม 2 character โดยที่ตำแหน่งของ <mask> จะต้องติดกันทั้งสองตำแหน่ง

**ปิด 5 ตำแหน่งติดกัน:** <mask> คำโดยการสุ่ม 5 character โดยที่ตำแหน่งของ <mask> จะต้องติดกันทั้งสองตำแหน่ง

**ปิด 2 ตำแหน่งไม่ติดกัน:** <mask> คำโดยการสุ่ม 2 character โดยที่ตำแหน่งของ <mask> จะต้องไม่ติดกันทั้งสองตำแหน่ง

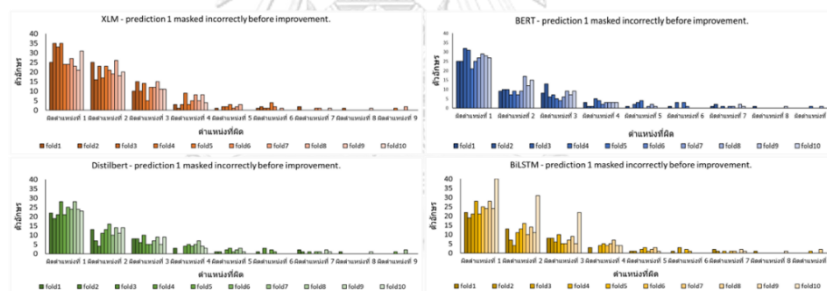
ตารางที่ 7 รายงานผลลัพธ์ความถูกต้องของการทำนายอักขระในประโยคของโมเดลทั้ง 4 โมเดลก่อนปรับปรุง

โมเดล (Model) วิธีการ (Method)	ผลลัพธ์ความถูกต้องของโมเดล (หน่วย เปอร์เซนต์)			
	เอ็กซ์แอลเอ็ม-โรเบอตา	เบิร์ต-เบส-หลายภาษา-เคสต์	ดิสทิลเบิร์ต-เบส-หลายภาษา-เคสต์	หน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง
ปิด 1 ตำแหน่ง	47	50	51	25
ปิด 2 ตำแหน่งติดกัน	15	29	10	17
ปิด 5 ตำแหน่งติดกัน	2	6	5	18

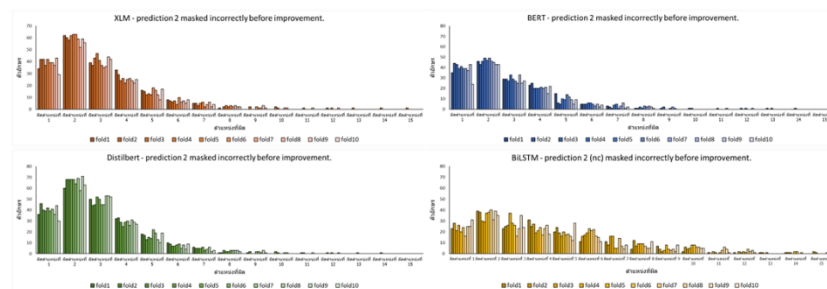
โมเดล (Model) วิธีการ (Method)	ผลลัพธ์ความถูกต้องของโมเดล (หน่วย เปอร์เซนต์)			
	เอ็กซ์แอลเอ็ม-โรเบอตา ๓๗	เบิร์ต-เบส-หลาย ภาษา-เคสดี	ดิสทิลเบิร์ต-เบส-หลาย ภาษา-เคสดี	หน่วยความจำ ระยะสั้นระยะยาว แบบสองทิศทาง
ปิด 2 ตำแหน่งไม่ติดกัน	28	11	15	17

#### 4.2.1 การวิเคราะห์ผลลัพธ์ของแต่ละโมเดลก่อนการปรับปรุง

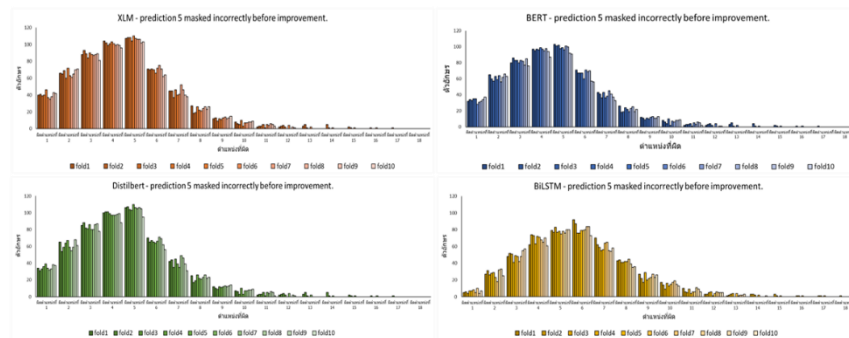
จากตารางที่ 7 พบว่าการสุ่มปิดที่ 1 2 และ 5 ตำแหน่งติดกัน และ 2 ตำแหน่งที่ไม่ติดกัน มีอัตราการทายตัวอักษรผิดในแต่ละโมเดลดังนี้ (1) โมเดลเอ็กซ์แอลเอ็ม-โรเบอตา อยู่ที่ร้อยละ 53 85 98 และ 72 ตามลำดับ (2) โมเดลเบิร์ต-เบส-หลายภาษา-เคสดี อยู่ที่ร้อยละ 50 71 94 และ 89 ตามลำดับ (3) โมเดลดิสทิลเบิร์ต-เบส-หลายภาษา-เคสดี อยู่ที่ร้อยละ 49 90 95 และ 85 ตามลำดับ และโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง อยู่ที่ร้อยละ 75 83 82 และ 83 ตามลำดับ จากนั้นจึงทำการตรวจสอบโมเดลเพื่อศึกษาหาวิธีการแก้ไขจุดที่โมเดลมีโอกาสทายผิดสูงสุดอยู่ตรงตำแหน่งใด จากภาพที่ 19 โมเดลที่ถูกสุ่มปิดคำทำนาย 1 ตำแหน่งมีโอกาสในการทายคำผิดสูงสุดในตำแหน่งตัวอักษรลำดับที่ 1 2 และ 3 ไปในทิศทางเดียวกันทั้ง 4 ประเภท



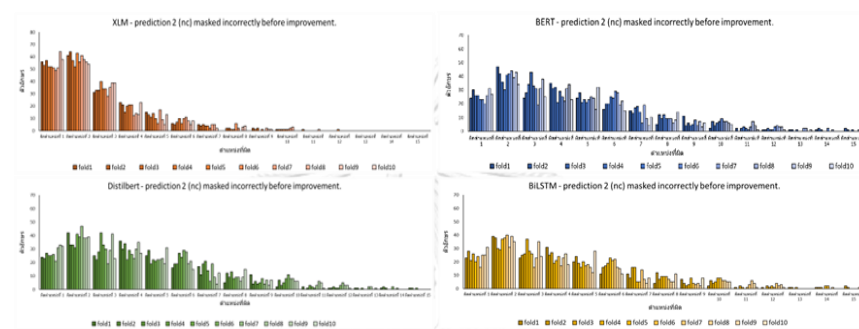
ภาพที่ 19 ผลลัพธ์ของตำแหน่งสุ่มปิด 1 อักขระในคำที่โมเดลทำนายผิด



ภาพที่ 20 ผลลัพธ์ของตำแหน่งสุ่มปิด 2 อักขระในคำที่โมเดลทำนายผิด



ภาพที่ 21 ผลลัพธ์ของตำแหน่งสุ่มปิด 5 อักขระในคำที่โมเดลทำนายผิด



ภาพที่ 22 ผลลัพธ์ของตำแหน่งสุ่มปิด 2 อักขระแบบไม่ติดกันในโมเดลทำนายผิด

ต่อมาจึงทำการตรวจสอบโมเดลที่สุ่มปิด 2 อักขระแบบติดกันในภาพที่ 20 พบว่าโมเดลในกลุ่มทรานฟอร์เมอร์ มีแนวโน้มทำนายคำผิดในตำแหน่งตัวอักษรลำดับที่ 2 สูงสุดและมีแนวโน้มทำนายคำผิดลดลงเรื่อย ๆ หลังจากตำแหน่งที่ 2 ส่วนโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง มีผลการทำนายคล้ายกับกลุ่มโมเดลกลุ่มทรานฟอร์เมอร์แต่มีจำนวนการทำผิดในแต่ละตำแหน่งใกล้เคียงกัน ต่อมาในโมเดลที่สุ่มปิด 5 อักขระแบบติดกันในภาพที่ 21 พบว่ามีแนวโน้มที่โมเดลทำนายผิดไล่ระดับไปจากตำแหน่งที่ 1 จนสูงสุดถึงตำแหน่งที่ 5 แล้วลดจำนวนต่ำลงไปเรื่อย ๆ และเมื่อทำการวิเคราะห์ตำแหน่งสุ่มปิด 2 อักขระแบบไม่ติดกันในภาพที่ 22 จะเห็นว่ามีแค่โมเดลเอ็กซ์แอลเอ็ม-โรเบอตาเท่านั้นที่มีแนวโน้มในการทำนายแบบสุ่มปิด 2 ตำแหน่งไปในทิศทางเดียวกันในภาพที่ 20 แต่โมเดลที่เหลืออีก 3 ประเภทมีระดับจำนวนการทำผิดแบบลดหลั่นในแต่ละตำแหน่งที่ใกล้เคียงกัน

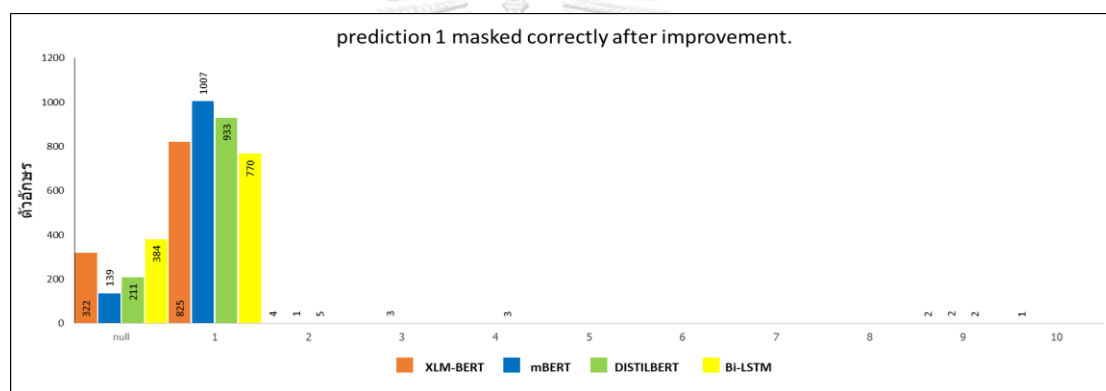
#### 4.2.2 การวิเคราะห์ผลลัพธ์ของแต่ละโมเดลหลังการปรับปรุง

หลังทำการปรับปรุงกระบวนการทำนายคำด้วยวิธีการเพิ่มคำศัพท์เฉพาะและการประยุกต์ใช้คลังคำศัพท์เฉพาะผลลัพธ์หลังการปรับปรุงอ้างอิงจากตารางที่ 8 พบว่าผลลัพธ์ความถูกต้องในการทำนายเพิ่มสูงขึ้นมากกว่า 30 เปอร์เซ็นต์จากค่าเฉลี่ยก่อนปรับปรุง



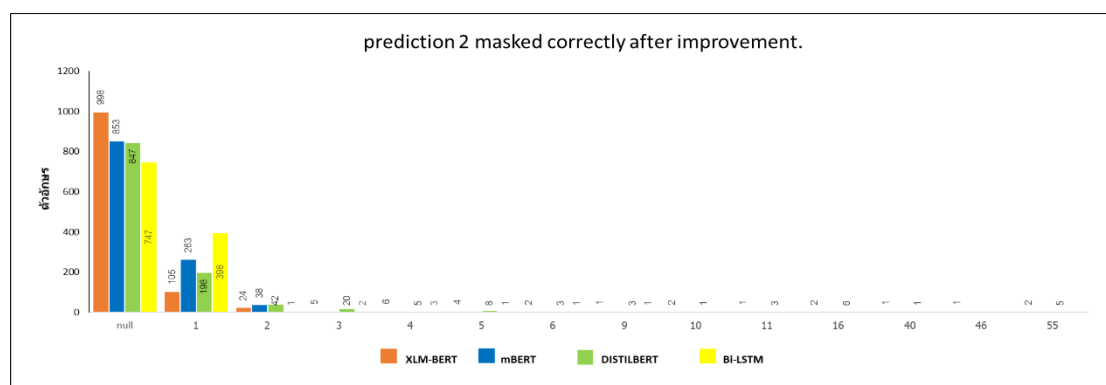
ตารางที่ 8 รายงานผลลัพธ์ความถูกต้องของการทำนายอักขระในประโยคของโมเดลทั้ง 4 โมเดลหลังปรับปรุง

โมเดล วิธีการ	ผลลัพธ์ความถูกต้องของโมเดล (หน่วย เปอร์เซ็นต์)			
	เอ็กซ์แอลเอ็ม-โรเบอตา	เบิร์ต-เบส-หลายภาษา-เคสดี	ดิสทิลเบิร์ต-เบส-หลายภาษา-เคสดี	หน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง
ปิด 1 ตำแหน่ง	72	87.9	81.7	66.7
ปิด 2 ตำแหน่งติดกัน	45.2	61.1	60.8	52.3
ปิด 5 ตำแหน่งติดกัน	6	27.3	26.8	42.6
ปิด 2 ตำแหน่งไม่ติดกัน	45.3	61.3	62.3	41.8



ภาพที่ 23 ผลลัพธ์ค่าความถูกต้องของตำแหน่งสุ่มปิด 1 ตำแหน่ง

จากภาพที่ 23 โมเดลทำการจับคู่คำด้วยโทเคนที่ถูกทำนายและนำคำศัพท์ที่ได้ไปค้นหาในคลังคำศัพท์พบว่าโมเดลมีการทำนายคำศัพท์ได้ถูกต้องตั้งแต่การเจอคำศัพท์ตัวนั้นที่ถูกต้องตั้งแต่ลำดับแรกและจากจำนวนคำที่โมเดลทำนายออกมาไม่ถูกต้องเลยนั้นมีค่าลดลงตั้งแต่ 4 ถึง 51 เปอร์เซ็นต์

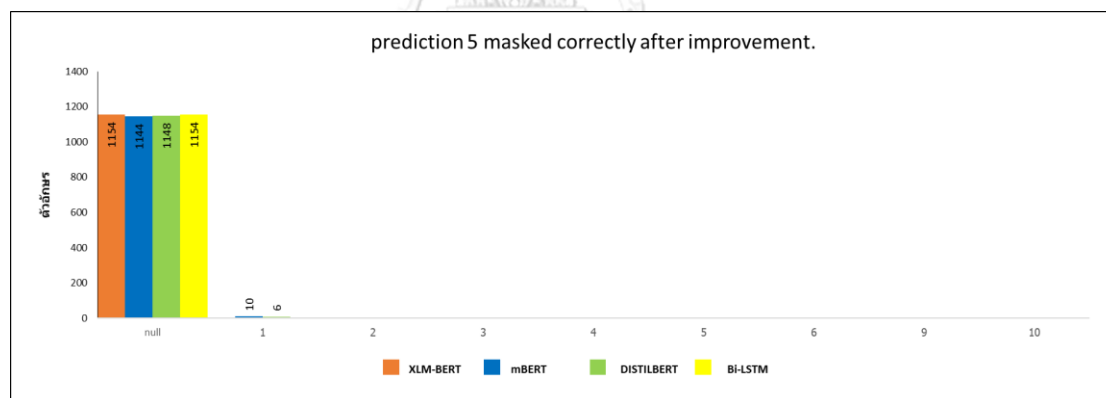


ภาพที่ 24 ผลลัพธ์ค่าความถูกต้องของการทำนายอักขระในประโยคของตำแหน่งสุ่มปิด 2 ตำแหน่ง

จากภาพที่ 24 โมเดลทำการจับคู่คำด้วยโทเคนที่ถูกทำนายและนำคำศัพท์ที่ได้ไปค้นหาในคลังคำศัพท์ พบว่าโมเดลมีแนวโน้มในการทำนายคำให้ถูกต้องทั้ง 2 ตำแหน่งที่ถูกสุ่มปิดน้อยมากอยู่ที่ 10 ถึง 35 เปอร์เซ็นต์เมื่อเทียบกับจำนวนคำที่ทายผิดและหากวิเคราะห์ที่ข้อมูลจากตารางที่ 8 สามารถแบ่งแยกอัตราการถูกผิดได้จากตารางที่ 9 พบว่าโมเดลมีความแม่นยำในการทำนายคำที่ถูกต้องเพียง 1 อักษรต่ำสุดที่ 3.4 เปอร์เซ็นต์จนถึงสูงสุดที่ 68.7 เปอร์เซ็นต์ และอัตราการทำนายคำที่ไม่ถูกเลยจะอยู่ที่ 3.8 ถึง 30.7 เปอร์เซ็นต์

ตารางที่ 9 ผลลัพธ์ความถูกต้องของโมเดลในแต่ละตำแหน่งของวิธีการปิด 2 ตำแหน่งติดกัน

โมเดล ตำแหน่ง	ผลลัพธ์ความถูกต้องของโมเดล (หน่วย เปอร์เซ็นต์)			
	เอ็กซ์แอลเอ็ม-โรเบอตา	เบิร์ต-เบส-หลายภาษา-เคสดี	ดิสทิลเบิร์ต-เบส-หลายภาษา-เคสดี	หน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง
ถูก 1 ตำแหน่ง	68.7%	70.1%	68.5%	34.1%
ถูก 2 ตำแหน่ง	10.8%	26.1%	26.6%	35.3%
ไม่ถูกเลย	20.5%	3.8%	4.9%	30.7%



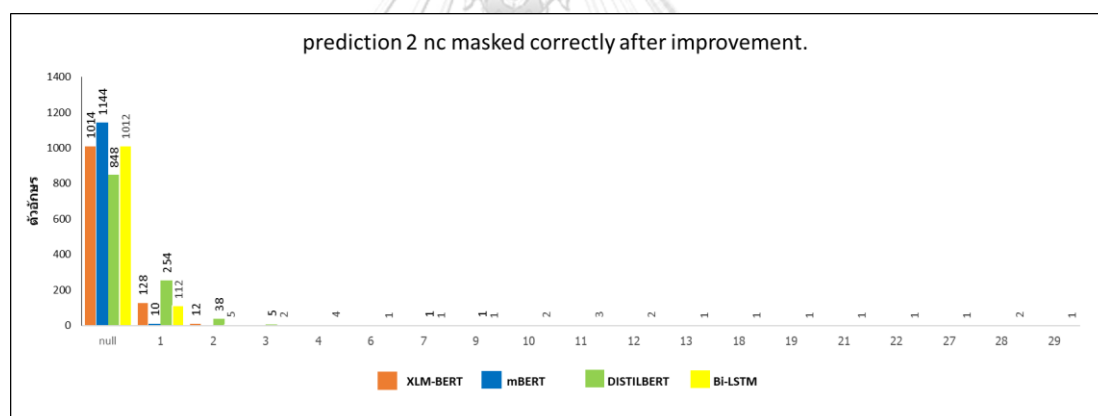
ภาพที่ 25 ผลลัพธ์ความถูกต้องของการทำนายอักษรในประโยคของตำแหน่งสุ่มปิด 5 ตำแหน่ง

จากภาพที่ 25 หลังจากการพยายามปรับปรุงโมเดลพบว่าการสุ่มปิดตัวอักษรที่ 5 ตำแหน่งมีแนวโน้มสูงที่โมเดลไม่สามารถทำนายผลลัพธ์ได้ถูกต้องทั้งหมดและค่าความถูกต้องของโมเดลจะอยู่ในช่วงระหว่าง 6 ถึง 42.6 เปอร์เซ็นต์ อ้างอิงข้อมูลจากตารางที่ 8 บ่งชี้ให้เห็นว่าโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา มีประสิทธิภาพต่ำที่สุดในการทำนายส่วนโมเดลอื่น ๆ ล้วนมีความสามารถในการทำนายถูกแค่บางตำแหน่งเท่านั้นโดยอ้างอิงข้อมูลจากตารางที่ 10 โอกาสสูงสุดที่โมเดลในกลุ่มทรานฟอเมอร์สามารถ

ทำนายออกมาได้ผลลัพธ์ที่ดีที่สุดคือ ตำแหน่งที่ถูกสุ่มปิด 1 ตำแหน่ง และในกลุ่มของโมเดล หน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง คือตำแหน่งที่ถูกสุ่มปิด 2 ตำแหน่ง

ตารางที่ 10 ผลลัพธ์ความถูกต้องของโมเดลในแต่ละตำแหน่งของวิธีการปิด 5 ตำแหน่งติดกัน

ตำแหน่ง (Positions) \ โมเดล (Model)	ผลลัพธ์ความถูกต้องของโมเดล (หน่วย เปอร์เซ็นต์)			
	XLM-BERT	mBERT	DISTILBERT	Bi-LSTM
ถูก 1 ตำแหน่ง	21.2%	41.6%	42.8%	0.2%
ถูก 2 ตำแหน่ง	3.3%	25.0%	24.0%	84.7%
ถูก 3 ตำแหน่ง	0%	10.5%	10.1%	13.6%
ถูก 4 ตำแหน่ง	0%	2.6%	2.6%	0.7%
ถูก 5 ตำแหน่ง	0%	0.9%	0.5%	0%
ไม่ถูกเลย	75.5%	19.8%	19.9%	0.9%



ภาพที่ 26 ผลลัพธ์ความถูกต้องของการทำนายอักขระในประโยคของตำแหน่งสุ่มปิด 2 ตำแหน่งที่ไม่ต่อเนื่อง

หลังจากทำการวิเคราะห์ข้อมูลในกลุ่มของการสุ่มปิดอักขระที่มีติดกันตามลำดับแล้วต่อไปจะเป็นการวิเคราะห์ข้อมูลของการสุ่มปิดอักขระ 2 ตำแหน่งที่ไม่ติดกันจากภาพที่ 26 พบว่าโมเดลมีแนวโน้มในการทำนายคำให้ถูกทั้ง 2 ตำแหน่งที่ถูกสุ่มปิดน้อยมากอยู่ที่ 12 ถึง 26 เปอร์เซ็นต์เมื่อเทียบกับจำนวนคำที่ทายผิดและหากวิเคราะห์ข้อมูลจากตารางที่ 8 สามารถแบ่งแยกอัตราการถูกผิดได้จากตารางที่ 11 พบว่าโมเดลมีความแม่นยำในการทำนายคำที่ถูกต้องเพียง 1 อักขระต่ำสุดที่ 59.1 เปอร์เซ็นต์จนถึงสูงสุดที่ 71.6 เปอร์เซ็นต์ และอัตราการทำนายคำที่ไม่ถูกเลยจะอยู่ที่ 1.9 ถึง 28.6 เปอร์เซ็นต์

ตารางที่ 11 ผลลัพธ์ความถูกต้องโมเดลในแต่ละตำแหน่งของวิธีการปิด 2 ตำแหน่งที่ไม่ต่อเนื่อง

ตำแหน่ง โมเดล	ผลลัพธ์ความถูกต้องของโมเดล (หน่วย เปอร์เซ็นต์)			
	เอ็กซ์แอลเอ็ม-โรเบอตา	เบิร์ต-เบส-หลายภาษา-เคสดี	ดิสทิลเบิร์ต-เบส-หลายภาษา-เคสดี	หน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง
ถูก 1 ตำแหน่ง	66.4%	70.1%	71.6%	59.1%
ถูก 2 ตำแหน่ง	12.1%	26.1%	26.5%	12.3%
ไม่ถูกเลย	21.5%	3.8%	1.9%	28.6%

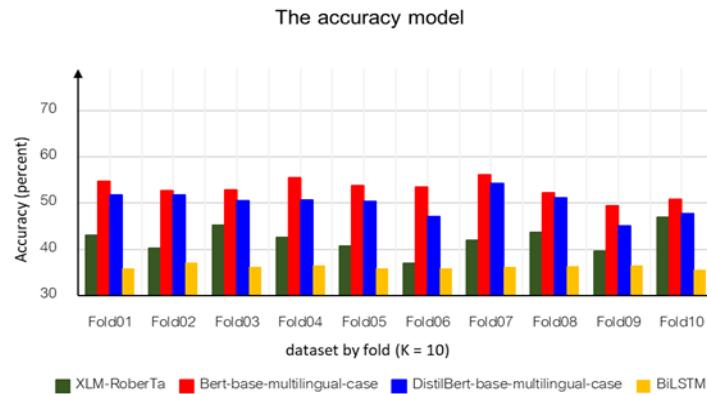


## บทที่ 5

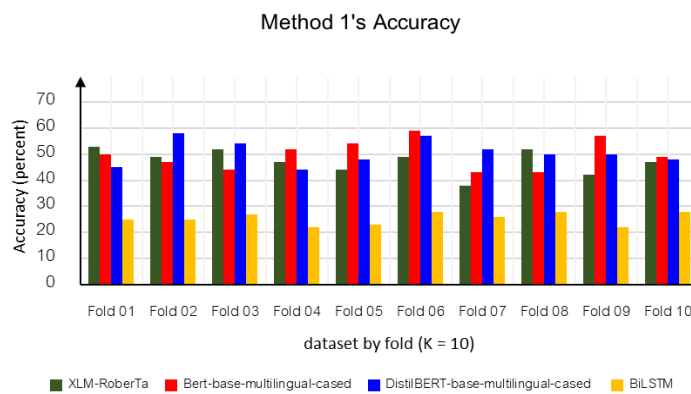
### สรุปผลการวิจัยและข้อเสนอแนะ

#### 5.1 สรุปผลการวิจัย

รายงานผลการทดลองจากตารางที่ 6 และ 7 ของงานวิจัยฉบับนี้ พบว่าสมรรถนะของโมเดลในการทำนายอักขระที่ถูกสุ่มปิดคำด้วยโทเคน <mask> ด้วยการใช้โมเดลเรียนรู้ล่วงหน้าทั้ง 3 ประเภทได้แก่ (1) เอ็กซ์แอลเอ็ม-โรเบอตา (2) เบิร์ต-เบส-หลายภาษา-เคสต์ (3) ดิสทิลเบิร์ต-เบส-หลายภาษา-เคสต์ และการใช้โมเดลในกลุ่มแบบจำลองโครงข่ายประสาทเทียมแบบลึกแบบเรียนรู้สองทิศทางได้แก่ (4) หน่วยความจำระยะสั้นระยะยาวแบบสองทิศทางมีจำนวนคลาสทั้งสิ้น 72 คลาส โดยที่ค่าความถูกต้องของโมเดลจากภาพที่ 27 อยู่ที่ค่าเฉลี่ยร้อยละ 42 ร้อยละ 53 ร้อยละ 50 และ ร้อยละ 36 ตามลำดับ จากนั้นนำโมเดลที่ผ่านการเรียนรู้ล่วงหน้าทั้งสามประเภทและโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง มาทำการวัดประสิทธิภาพในการทำนายอักขระที่ถูกสุ่มปิดคำด้วยโทเคน <mask> พบว่าผลลัพธ์ของการทำนายโดยใช้การสุ่มอักขระหนึ่งตำแหน่งได้ประสิทธิภาพใกล้เคียงกับค่าความถูกต้องของเมทริกซ์คอนฟิวชันอ้างอิงจากภาพที่ 28 โดยที่โมเดล ดิสทิลเบิร์ต-เบส-หลายภาษา-เคสต์ มีผลการทำนายที่สูงที่สุดอยู่ที่ร้อยละ 51 สำหรับโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา โมเดลเบิร์ต-เบส-หลายภาษา-เคสต์ และโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทางจะอยู่ที่ร้อยละ 47 50 และ 25 ตามลำดับ เมื่อนำผลการทดลองในกลุ่มของโมเดลทรานฟอเมอร์ไปเปรียบเทียบกับโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง มีค่าความถูกต้องของโมเดลอยู่ที่ร้อยละ 36 แสดงให้เห็นว่าโมเดลในกลุ่มของโมเดลทรานฟอเมอร์มีสมรรถนะในการทำนายคำที่ถูกปิดสูงกว่าร้อยละ 6 ถึง 17 อีกทั้งเมื่อมีการนำโมเดลทั้ง 4 ประเภทไปทำการทดสอบทำนายคำกับชุดข้อมูล อ้างอิงจากตารางที่ 6 พบว่าประสิทธิภาพในการทำนายของทั้ง 4 โมเดลมีความสามารถในการทำนายอักขระที่หายไป 1 ตำแหน่งติดกันได้ดีกว่าแบบอื่น ๆ อีกทั้งผลการทำนายแสดงให้เห็นแนวโน้มของทั้งสามโมเดลในกลุ่มของทรานฟอเมอร์อยู่ในทิศทางใกล้เคียงกันและมีความสามารถในการทายคำที่ถูกปิดกลับลดลงเรื่อย ๆ เมื่อมีการเพิ่มจำนวนอักขระที่ถูกสุ่มปิดด้วย <mask> กลับกันในส่วนของโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง แม้ว่าตัวโมเดลจะได้ค่าความถูกต้องที่ต่ำกว่าโมเดลอื่น ๆ แต่ความสามารถในการทายคำที่ถูกปิดมากกว่า 1 ตำแหน่งกลับสูงกว่าโมเดลในกลุ่มทรานฟอเมอร์ อ้างอิงความถูกต้องในการทำนายอักขระดังต่อไปนี้

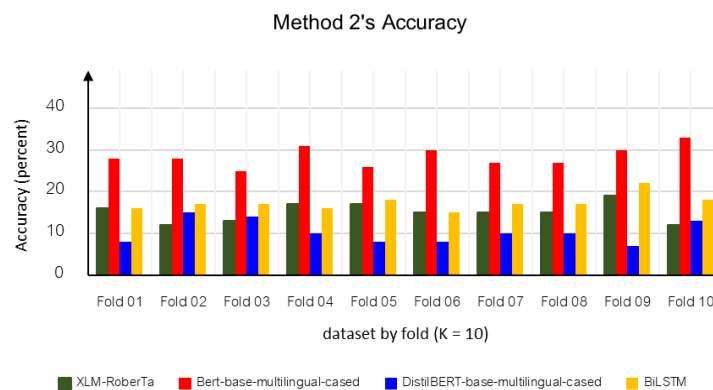


ภาพที่ 27 ค่าความถูกต้องของโมเดลทั้งสี่ประเภทจากเมทริกซ์คอนฟิวชัน



ภาพที่ 28 ค่าความถูกต้องของโมเดลที่ตำแหน่งของ <mask> หนึ่งตำแหน่ง

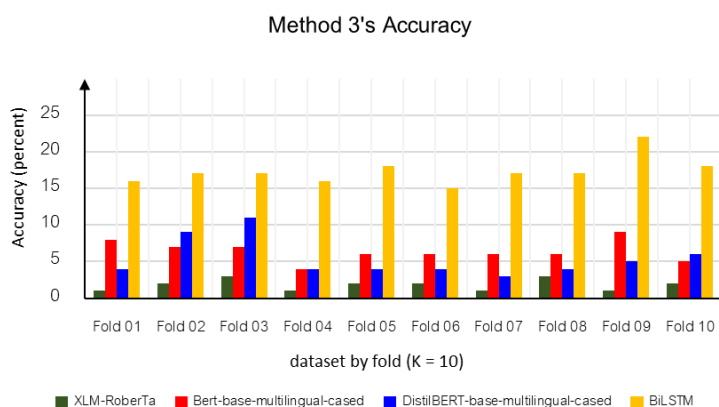
1. โมเดลทำนายอักขระที่ถูกปิดโดยที่ตำแหน่งของ <mask> จะต้องติดกันทั้งสองตำแหน่ง โมเดลเบิร์ต-เบส-หลายภาษา-เคสดี แสดงความสามารถในการทำนายอักขระที่ถูกปิดสูงกว่าโมเดลตัวอื่น ๆ อยู่ที่ระดับ 17 – 18 เปอร์เซ็นต์ จากภาพที่ 29



ภาพที่ 29 ค่าความถูกต้องของโมเดลที่ตำแหน่งของ <mask> จะต้องติดกันทั้งสองตำแหน่ง

2. โมเดลทำนายอักขระที่โดยที่ตำแหน่งของ <mask> จะต้องติดกันทั้งห้าตำแหน่ง

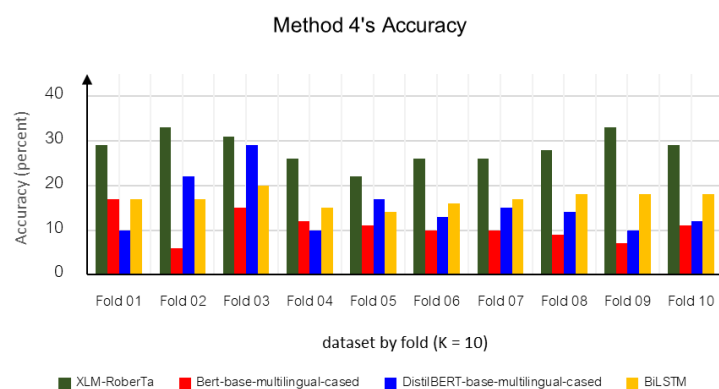
โมเดลทั้งสามประเภทในกลุ่มทรานฟอร์เมอร์จากภาพที่ 30 แสดงความสามารถในการทำนายอักขระที่ถูกปิดติดกันห้าตำแหน่งในระดับความถูกต้องที่อยู่ระดับ 1 – 2 เปอร์เซ็นต์ ถือว่าอยู่ในเกณฑ์ที่ต่ำและมีประสิทธิภาพไม่เพียงพอที่จะทำนายอักขระและโมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทางมีประสิทธิภาพทำนายอักขระอยู่ที่ 17 เปอร์เซ็นต์



ภาพที่ 30 ค่าความถูกต้องของโมเดลที่ตำแหน่งของ <mask> จะต้องติดกันทั้งห้าตำแหน่ง

3. โมเดลทำนายอักขระที่ถูกปิดไป 2 ตำแหน่งโดยที่ตำแหน่งของ <mask> จะต้องไม่ติดกันทั้งสองตำแหน่ง

จากภาพที่ 31 โมเดลเอ็กซ์แอลเอ็ม-โรเบอตา มีความสามารถในการทำนายอักขระที่ถูกปิดโดยที่อักขระไม่ติดกันทั้งสองตำแหน่งได้ดีกว่าโมเดลตัวอื่น ๆ และรองลงมาคือ โมเดลหน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง



ภาพที่ 31 ค่าความถูกต้องของโมเดลที่ตำแหน่งของ <mask> จะต้องไม่ติดกันทั้งสองตำแหน่ง

จากที่ได้กล่าวมาข้างต้นผลลัพธ์ที่ได้ยังมีประสิทธิภาพไม่เพียงพอต่อการทำนายนงานวิจัยนี้จึงได้นำเสนอวิธีการปรับปรุงการการนายนเพื่อให้อาประสิทธิภาพของการการนายนมีความถูกต้องมากขึ้น โดยหลังการปรับปรุงค่าความถูกต้องของโมเดลเพิ่มสูงขึ้นอ้างอิงข้อมูลจากตารางที่ 8 ค่าความถูกต้องของการสุ่มปิดที่ 1 ตำแหน่งมีแนวโน้มในการการนายนผลลัพธ์อยู่ในช่วง 66.7 ถึง 87.9 เปอร์เซ็นต์ มากกว่าก่อนปรับปรุงประมาณ 33 เปอร์เซ็นต์ ในขณะที่การสุ่มปิดที่ 2 ตำแหน่งทั้งติดกันและไม่ติดกันมีระดับความแม่นยำอยู่ในช่วงตั้งแต่ 41.8 ถึง 62.3 เปอร์เซ็นต์ มากกว่าก่อนปรับปรุงประมาณ 36 เปอร์เซ็นต์ และในตำแหน่งที่มีการสุ่มปิดติดกัน 5 ตำแหน่งมีแนวโน้มในการการนายนผลลัพธ์ได้ดีในโมเดลต่อไปนี้ (1) เบิร์ต-เบส-หลายภาษา-เคสดี (2) ดิสทิลเบิร์ต-เบส-หลายภาษา-เคสดี (3) หน่วยความจำระยะสั้นระยะยาวแบบสองทิศทาง อยู่ในช่วงตั้งแต่ 26.8 ถึง 42.6 เปอร์เซ็นต์ มากกว่าก่อนปรับปรุงประมาณ 25 เปอร์เซ็นต์ ในส่วนของโมเดลเอ็กซ์แอลเอ็ม-โรเบอตา ผลลัพธ์อยู่ที่ร้อยละ 6 เปอร์เซ็นต์ บ่งชี้ให้เห็นว่าโมเดลมีความสามารถในการการนายนที่ต่ำในกรณีที่มีการสุ่มปิด 5 อักษรติดกันแบบต่อเนื่อง

## 5.2 ข้อเสนอแนะ

งานวิจัยฉบับนี้เป็นงานวิจัยเชิงปฏิบัติใช้ชุดข้อมูลจริงที่ถูกลดความมาจากศิลาจารึกโดยหน่วยงานฐานข้อมูลจารึกแห่งประเทศไทย ศูนย์มานุษยวิทยาสิรินธร โดยใช้ชุดข้อมูลจากศิลาจารึก 5 หลัก เพื่อศึกษาวิธีการเรียกคืนระดับตัวอักษรที่ขาดหายไปจากข้อความที่ถูกป้อนเข้ามา ดังนั้นผลของงานวิจัยนี้อาจจะยังไม่สามารถครอบคลุมกับการเรียกคืนระดับตัวอักษรจากศิลาจารึกสุโขทัยทั้งหมดได้จึงควรเพิ่มเติมชุดข้อมูลเข้าไปในโมเดลมากขึ้นเพื่อยกระดับประสิทธิภาพในการเรียนรู้ของโมเดล



## บรรณานุกรม

1. Department, F.A., *A practical manual on ancient documents*. 2548, Bangkok: Fine Arts Department: Bangkok: Fine Arts Department. 68.
2. Cheoklin, N., *The digitization of the inscription report of the Department of Eastern Languages, Faculty of Archeology*. 2022, Faculty of Archeology. p. 6.
3. Coedès, G., *wat-pa-mamuang*, in *WAT-PA-MAMUANG INSCRIPTIONS*, V.U. Trongjai Hutangkura, Dokrak Payaksri, Editor. 1963: THE PRINCESS MAHA CHAKRI SIRINDHORN ANTHROPOLOGY CENTRE.
4. Coedès, G., *nakhon chum*, in *NAKHON-CHUM INSCRIPTIONS*, V.U. Trongjai Hutangkura, Dokrak Payaksri, Editor. 1983, Trongjai Hutangkura, Vasharabhorn Ungkunshutchai, Dokrak Payaksri: THE PRINCESS MAHA CHAKRI SIRINDHORN ANTHROPOLOGY CENTRE.
5. Coedès, G., *pho khun ram khamheng*, in *pho khun ram khamheng*, V.U. Trongjai Hutangkura, Dokrak Payaksri, Editor. 1983, Trongjai Hutangkura, Vasharabhorn Ungkunshutchai, Dokrak Payaksri: THE PRINCESS MAHA CHAKRI SIRINDHORN ANTHROPOLOGY CENTRE.
6. Coedès, G., *wat si chum*, in *wat si chum inscriptions*, V.U. Trongjai Hutangkura, Dokrak Payaksri, Editor. 1978, Trongjai Hutangkura, Vasharabhorn Ungkunshutchai, Dokrak Payaksri: THE PRINCESS MAHA CHAKRI SIRINDHORN ANTHROPOLOGY CENTRE.
7. thongkhamwann, C., *pu khun chit khun chot inscriptions*, in *pu khun chit khun chot*, V.U. Trongjai Hutangkura, Dokrak Payaksri, Editor. 1957, Trongjai Hutangkura, Vasharabhorn Ungkunshutchai, Dokrak Payaksri: THE PRINCESS MAHA CHAKRI SIRINDHORN ANTHROPOLOGY CENTRE.
8. Alpa Reshamwala, P.P., Dharendra S Mishra, *REVIEW ON NATURAL LANGUAGE PROCESSING*. IRACST – Engineering Science and Technology: An International Journal (ESTIJ), 2013. **3**: p. 2.
9. Jacob Devlin, M.-W.C., Kenton Lee, Kristina Toutanova *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2018. **1**, 1-2 DOI:

<https://doi.org/10.48550/arXiv.1810.04805>.

10. Vaswani, A., et al., *Attention Is All You Need*. arXiv pre-print server, 2017.
11. Sun, C., et al., *How to Fine-Tune BERT for Text Classification?* arXiv pre-print server, 2020.
12. Courville, I.G.a.Y.B.a.A., *Deep Learning*. 2016, MIT Press. p. 716.
13. Zhang, W., *Music Genre Classification Based on Deep Learning*. Mobile Information Systems, 2022. **2022**: p. 1-11.
14. Sepp Hochreiter, J.S., *Long Short-term Memory*. Neural Computation, 1997. **9**: p. 1735-80.
15. Soam, M. and S. Thakur. *Next Word Prediction Using Deep Learning: A Comparative Study*. IEEE.
16. Devlin, J., et al., *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv pre-print server, 2019.
17. Taylor, W.L., *"Cloze Procedure": A New Tool For Measuring Readability*. Journalism quarterly, 1953. **30**: p. 415-433.
18. Conneau, A., et al., *Unsupervised Cross-lingual Representation Learning at Scale*. arXiv pre-print server, 2020: p. 1-3.
19. Zhou, Y., et al., *TopicBERT: A Topic-Enhanced Neural Language Model Fine-Tuned for Sentiment Classification*. IEEE Transactions on Neural Networks and Learning Systems, 2023. **34**(1): p. 380-393.
20. Platform, H.F., *Supported frameworks*. Hugging Face: Hugging Face Website.
21. Platform, H.F., *Data Collator*. Hugging Face: Hugging Face Website. p. <https://huggingface.co/docs/>.
22. Andreas C. Müller, S.G., *Introduction to Machine Learning with Python*. first ed. Introduction to Machine Learning with Python, ed. Dawn Schanafelt. Vol. 1. 2016, Gravenstein Highway North, Sebastopol: O'Reilly Media. 376.

## ประวัติผู้เขียน

ชื่อ-สกุล	นางสาวสุจิตรา ทองคำ
วัน เดือน ปี เกิด	31 กรกฎาคม ค.ศ. 1993
สถานที่เกิด	ประเทศไทย
วุฒิการศึกษา	ปริญญาตรี วิศวกรรมอุตสาหการ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ กรุงเทพฯ ปริญญาโท วิทยาศาสตร์มหาบัณฑิต สาขาวิทยาการคอมพิวเตอร์ ภาควิชา วิศวกรรม คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ที่อยู่ปัจจุบัน	38/2 ตำบล ท่าสะท้อน อำเภอ พุนพิน จังหวัด สุราษฎร์ธานี



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY