

ระบบประมวลคำวิจารณ์ภาพยนตร์ของผู้ชมทางทวิตเตอร์

MOVIE CRITICISM SYSTEM FOR VIEWERS ON TWITTER



ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2562

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบประมวลคำวิจารณ์ภาพยนตร์ของผู้ชมทางวิดีโอ

ปีการศึกษา 2562



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบประมวลคำวิจารณ์ภาพยนตร์ของผู้ชมทางทวิตเตอร์

นายจรูญกร อยู่เจริญทรัพย์ 59011146
นายอภิรัตน์ เชาว์เครือ 59011497
อาจารย์บุญฤทธิ์ เครือตราชู อาจารย์ที่ปรึกษา
ปีการศึกษา 2562

บทคัดย่อ

สืบเนื่องมาจากการเปิดตัวของแพลตฟอร์มการดูภาพยนตร์ออนไลน์ภายในประเทศไทยในเมื่อไม่กี่ปีที่ผ่านมา ตัวอย่างเช่น Netflix, Line TV เป็นต้น ส่งผลให้คนไทยนิยมชมภาพยนตร์ทางออนไลน์มากขึ้น แต่เนื่องจากภาพยนตร์ตามแพลตฟอร์มต่าง ๆ มีความหลากหลายมาก จนอาจทำให้ผู้ใช้ไม่สามารถตัดสินใจที่จะเลือกรับชมภาพยนตร์ที่ทางแพลตฟอร์มต่าง ๆ แนะนำมาให้ได้ ประกอบกับในปัจจุบันคนไทยมีอัตราการเล่นทวิตเตอร์ (Twitter) เพิ่มขึ้น และมีการทวิต (Tweet) ข้อความที่เกี่ยวกับการวิจารณ์หรือการรีวิวภาพยนตร์เรื่องต่าง ๆ ซึ่งสามารถนำทวิตดังกล่าวช่วยในการแนะนำภาพยนตร์ให้กับผู้ชมได้ แต่เว็บไซต์ภายในประเทศไทยที่มีการใช้ข้อความจากทวิตเตอร์เพื่อช่วยผู้ใช้ตัดสินใจในการรับชมภาพยนตร์ รวมถึงการแนะนำ การตอบคำถาม หรือการพูดคุยทางภาพยนตร์ยังไม่เป็นที่รู้จักหรือมีน้อยเว็บไซต์ที่จะสามารถรวบรวมหัวข้อดังกล่าวเข้าไว้ด้วยกันได้

เพื่อเป็นแหล่งรวบรวมของผู้ที่มีใจรักในภาพยนตร์ หรือผู้ที่ต้องการรับการแนะนำภาพยนตร์ ทางคณะผู้จัดทำจึงได้ออกแบบและพัฒนาเว็บไซต์ โดยภายในเว็บไซต์จะแบ่งเป็น 4 หัวข้อหลัก ซึ่งหัวข้อแรกคือการแนะนำภาพยนตร์จากผู้ชมทางทวิตเตอร์ โดยจะนำข้อความ (Tweet) ของภาพยนตร์เรื่องนั้น มาแบ่งประเภทเป็นความคิดเห็นแง่บวก เป็นกลาง หรือแง่ลบ เพื่อช่วยในการตัดสินใจในการเลือกรับชมภาพยนตร์ อีกทั้งผู้ใช้ยังสามารถเลือกภาพยนตร์ที่ตนเองชอบหรือไม่ชอบเพื่อให้ระบบช่วยแนะนำภาพยนตร์ หัวข้อต่อมาจะเป็นการรวบรวมรีวิวภาพยนตร์เรื่องต่าง ๆ เพื่อช่วยผู้ใช้ตัดสินใจในการเลือกรับชมภาพยนตร์ หัวข้อที่สามจะเป็นการรวบรวมบทความหรือสาระเกร็ดความรู้ของภาพยนตร์เรื่องต่าง ๆ และหัวข้อสุดท้ายเปรียบเสมือนบอร์ดสำหรับแลกเปลี่ยนความรู้ ความคิดเห็น หรือตอบคำถามที่เกี่ยวข้องกับภาพยนตร์ จากหัวข้อที่กล่าวมาทำให้ครอบคลุมผู้ใช้ที่ต้องการรับการแนะนำภาพยนตร์ หรือผู้ที่หลงใหลในภาพยนตร์

Movie Criticism System For Viewers On Twitter

Mr. Rujikorn Yoocharoensub 59011146

Mr. Apirut Chaokrua 59011497

Mr. Boontee Kruatrachue Advisor

Academic Year 2019

ABSTRACT

A few years ago, there are opening video streaming in Thailand, such as Netflix, Line TV, etc. cause Thai people to watch a movie on an online platform increasingly. However, there is a diversity of films that make viewers cannot decide which of the recommends movie by the movie platform to watch. In the present, the number of Thai accounts on twitter has grown significantly, and they post a lot of tweets about movie criticism or movie reviews, which can use to recommend movies to viewers. But Thailand hardly has websites that use tweets to help viewers make a decision include a recommendation, questionnaires, and discussion.

To be a gathering website for people interested in the movie or those who want to get movie recommendations. So, we decide to design and develop the site. It has four main parts. The first part is to recommend a movie by taking the movie's tweet from Twitter and classify them into positive, neutral, and negative comments to help them decide on the film. Also, users can choose movies that they like or dislike then the system will recommend movies. The second part will be a collection of movie reviews to help users make decisions. The third part will be a collection of articles or knowledge about movies, and the last part is a Webboard for exchanging knowledge, opinions, or discussions related to movies. Each section covers users who want to get movie recommendations and people who are passionate about movies.

กิตติกรรมประกาศ

ทางคณะผู้จัดทำขอขอบพระคุณอาจารย์บุญฤทธิ์ เครือตราชู ซึ่งเป็นอาจารย์ที่ปรึกษา
โครงการ โดยช่วยให้คำปรึกษา คำชี้แนะ อีกทั้งยังให้ความรู้ในส่วนของการพัฒนาชิ้นงาน ตลอดจน
แนวทางในการแก้ไขปัญหาที่เกิดขึ้น

ขอขอบคุณเพื่อน ๆ ภาควิชาคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ เป็นอย่างยิ่ง ที่คอยช่วยเหลือ
ไม่ว่าจะเป็นการให้คำปรึกษา การชี้แนะแนวทางในการดำเนินงาน หรือเมื่อเกิดปัญหาที่คอย
ช่วยเหลือซึ่งกันและกัน และสุดท้ายนี้ต้องขอขอบคุณสถานที่จาก ภาควิชาคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ ที่ให้สถานที่ในการจัดทำโครงการชิ้นนี้

ทางคณะผู้จัดทำขออภัย สำหรับคำหยาบคาย หรือประโยคไม่เหมาะสมต่าง ๆ ที่ปรากฏ
เนื่องจากการยกตัวอย่างด้วยคำและประโยคจากโลกสังคมออนไลน์ โดยมีจุดประสงค์เพื่อ
การศึกษาเท่านั้น

รุจิกร อยู่เจริญทรัพย์
อภิรัตน์ เชาว์เครือ

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญรูป.....	X
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาของปัญหา.....	1
1.2 วัตถุประสงค์ของโครงการ.....	2
1.3 ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.4 ขอบเขตของโครงการ.....	2
1.5 ส่วนประกอบของวิทยานิพนธ์.....	3
1.6 ตารางเวลาของโครงการ.....	3
บทที่ 2 ทฤษฎีและเครื่องมือที่เกี่ยวข้อง.....	7
2.1 เทคโนโลยีที่ใช้ด้านเว็บแอปพลิเคชันฝั่งหน้าบ้าน (Front-End).....	7
2.2 เทคโนโลยีที่ใช้ด้านเว็บแอปพลิเคชันฝั่งหลังบ้าน (Back-End).....	8
2.3 เทคโนโลยีที่ใช้ด้านฐานข้อมูล.....	9
2.4 เทคโนโลยีด้านการเรียนรู้ของเครื่อง (Machine Learning).....	10
2.5 หลักภาษาไทย.....	11
2.6 การประมวลผลภาษาธรรมชาติ (Natural Language Processing).....	14
2.7 การเรียนรู้ของเครื่อง (Machine Learning).....	18
บทที่ 3 การออกแบบและพัฒนา.....	25
3.1 การออกแบบ.....	25
3.2 การพัฒนา.....	41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
บทที่ 4 การใช้งานและการทดลอง.....	45
4.1 เว็บไซต์.....	45
4.2 ผลการทดสอบการวิเคราะห์อารมณ์จากข้อความ.....	53
บทที่ 5 สรุปผลและข้อเสนอแนะ.....	66
5.1 บทสรุป.....	66
5.2 ความรู้ที่ได้จากทำโครงการ.....	67
5.3 ปัญหาอุปสรรคและแนวทางการแก้ไข.....	67
5.4 แนวทางการพัฒนาต่อ.....	68
บรรณานุกรม.....	69
ภาคผนวก.....	71
ก. รายการตารางอธิบายชุดสเปคที่ 3.....	71
ข. รายการตารางรายละเอียดฐานข้อมูลบทที่ 3.....	90

สารบัญตาราง

ตาราง	หน้า
1.1 การดำเนินการภาคเรียนที่ 1	3
1.2 การดำเนินการภาคเรียนที่ 1 (ต่อ)	4
1.3 การดำเนินการภาคเรียนที่ 1 (ต่อ)	5
1.4 การดำเนินการภาคเรียนที่ 2	6
4.1 จำนวนข้อมูลที่จัดเก็บในแต่ละแง่มุม	53
4.2 จำนวนข้อมูลที่ใช้ในการฝึกโมเดล	54
4.3 จำนวนข้อมูลที่ใช้ในการทดสอบโมเดล	54
4.4 ตัวอย่างข้อมูลที่ใช้ในการฝึกโมเดล	55
4.5 คอนฟิวชันเมตริกซ์แสดงผลการฝึกการวิเคราะห์อารมณ์ด้วยโมเดล Logistic Regression กับชุดข้อมูลฝึกสอน	56
4.6 คอนฟิวชันเมตริกซ์แสดงผลการทดสอบการวิเคราะห์อารมณ์ด้วยโมเดล Logistic Regression กับชุดข้อมูลทดสอบ	56
4.7 recall, precision และ F1-score ของ โมเดล Logistic Regression ของชุดข้อมูลทดสอบ	57
4.8 คอนฟิวชันเมตริกซ์แสดงผลการฝึกการวิเคราะห์อารมณ์ด้วย โมเดล Naïve Bayes	57
4.9 คอนฟิวชันเมตริกซ์แสดงผลการทดสอบการวิเคราะห์อารมณ์ด้วยโมเดล Naïve Bayes กับชุดข้อมูลทดสอบ	58
4.10 recall, precision และ F1-score ของโมเดล Naïve Bayes ของชุดข้อมูลทดสอบ	58
4.11 คอนฟิวชันเมตริกซ์แสดงผลการฝึกการวิเคราะห์อารมณ์ด้วย โมเดล SVM	59
4.12 คอนฟิวชันเมตริกซ์แสดงผลการทดสอบการวิเคราะห์อารมณ์ด้วยโมเดล SVM กับชุดข้อมูลทดสอบ	59
4.13 recall, precision และ F1-score ของโมเดล SVM ของชุดข้อมูลทดสอบ	59
4.14 คอนฟิวชันเมตริกซ์แสดงผลการฝึกการวิเคราะห์อารมณ์ด้วยโมเดล ULMFiT กับชุดข้อมูลฝึกสอน	60
4.15 คอนฟิวชันเมตริกซ์แสดงผลการทดสอบการวิเคราะห์อารมณ์ด้วย ULMFiT กับชุดข้อมูลทดสอบ	60
4.16 recall, precision และ F1-score ของโมเดล ULMFiT ของชุดข้อมูลทดสอบ	61

สารบัญตาราง (ต่อ)

ตาราง	หน้า
4.17 การเปรียบเทียบความแม่นยำในการเลือกใช้ n-gram ในช่วงต่าง ๆ สำหรับการทำให้ TF-IDF กับ โมเดล Logistic Regression และ SVM.....	61
4.18 ข้อความที่เป็นปัญหา.....	63
4.19 ข้อความที่เป็นปัญหา (ต่อ).....	64
4.20 ความแม่นยำชุดทดสอบของแต่ละโมเดล.....	65
ก.1 อธิบายยูสเคส ลงทะเบียน.....	71
ก.2 อธิบายยูสเคส เข้าสู่ระบบ.....	71
ก.3 อธิบายยูสเคส แก้ไขโปรไฟล์.....	71
ก.4 อธิบายยูสเคส ผ่านทาง Facebook.....	72
ก.5 อธิบายยูสเคส ดูโปรไฟล์.....	72
ก.6 อธิบายยูสเคส ติดตามผู้ใช้งาน.....	73
ก.7 อธิบายยูสเคส ยกเลิกการติดตามผู้ใช้งาน.....	73
ก.8 อธิบายยูสเคส ลบรายการหนังสือ.....	73
ก.9 อธิบายยูสเคส ลบรายการบทความ.....	74
ก.10 ยูสเคส ลบรายการรีวิว.....	74
ก.11 อธิบายยูสเคส ลบรายการกระดานสนทนา.....	74
ก.12 อธิบายยูสเคส ดูรายการหนังสือ.....	75
ก.13 อธิบายยูสเคส ดูรายการทวิต.....	75
ก.14 อธิบายยูสเคส เพิ่มรายการหนังสือที่อยากดู.....	76
ก.15 อธิบายยูสเคส ลบรายการหนังสือที่อยากดู.....	76
ก.16 อธิบายยูสเคส เพิ่มรายการหนังสือที่ชอบ/ไม่ชอบ/เฉย ๆ.....	76
ก.17 อธิบายยูสเคส ลบรายการหนังสือที่ชอบ/ไม่ชอบ/เฉย ๆ.....	77
ก.18 อธิบายยูสเคส ดูรายการบทความ.....	77
ก.19 อธิบายยูสเคส ถูกใจบทความ.....	78
ก.20 อธิบายยูสเคส บัญชีมาร์คบทความ.....	78
ก.21 อธิบายยูสเคส แก้ใจบทความ.....	78
ก.22 อธิบายยูสเคส ลบบทความ.....	79

สารบัญตาราง (ต่อ)

ตาราง	หน้า
ก.23 อธิบายยูสเคส คู่มือการรีวิว	79
ก.24 อธิบายยูสเคส คู่มือใจรีวิว	80
ก.25 อธิบายยูสเคส บัญชีรีวิว	80
ก.26 อธิบายยูสเคส แก้ไขรีวิว	80
ก.27 อธิบายยูสเคส ลบรีวิว	81
ก.28 อธิบายยูสเคส คู่มือการกระดานสนทนา	81
ก.29 อธิบายยูสเคส คู่มือใจกระดานสนทนา	82
ก.30 อธิบายยูสเคส บัญชีกระดานสนทนา	82
ก.31 อธิบายยูสเคส แก้ไขกระดานสนทนา	82
ก.32 อธิบายยูสเคส ลบกระดานสนทนา	83
ก.33 อธิบายยูสเคส สร้างบทความ/รีวิว/กระดานสนทนา	83
ก.34 อธิบายยูสเคส ดูความคิดเห็น	84
ก.35 อธิบายยูสเคส สร้างความคิดเห็น	84
ก.36 อธิบายยูสเคส แก้ไขความคิดเห็น	85
ก.37 อธิบายยูสเคส ลบความคิดเห็น	85
ก.38 อธิบายยูสเคส ค้นหา	85
ก.39 อธิบายยูสเคส จัดการสมาชิก	86
ก.40 อธิบายยูสเคส ลบผู้ใช้งาน	86
ก.41 อธิบายยูสเคส แก้ไขผู้ใช้งาน	86
ก.42 อธิบายยูสเคส จัดการข้อความทวิต	87
ก.43 อธิบายยูสเคส เพิ่มข้อความทวิต	87
ก.44 อธิบายยูสเคส แก้ไขข้อความทวิต	87
ก.45 อธิบายยูสเคส จัดการรายการหนัง/รีวิว/บทความ/กระดานสนทนา	88
ก.46 อธิบายยูสเคส เพิ่มรายการ	88
ก.47 อธิบายยูสเคส แก้ไขรายการ	89
ก.48 อธิบายยูสเคส ลบรายการ	89
ข.1 รายละเอียดตารางผู้ใช้งาน (User)	90

สารบัญตาราง (ต่อ)

ตาราง	หน้า
ข.2 รายละเอียดตารางการอนุญาต (Permission).....	90
ข.3 รายละเอียดตารางเอนทิตี (Entity).....	90
ข.4 รายละเอียดตารางความชอบผู้ใช้ (User_Entity).....	91
ข.5 รายละเอียดตารางการเก็บเอนทิตี (User_Keep_Entity).....	91
ข.6 รายละเอียดตารางทวิตเตอร์ (Twitters).....	91
ข.7 รายละเอียดตารางหนัง (Movies).....	92
ข.8 รายละเอียดตารางประเภทหนัง (Genres).....	93
ข.9 รายละเอียดตารางประเภทของหนัง (Movie_Genres).....	93
ข.10 รายละเอียดตารางคำสำคัญ (Plot_Keywords).....	93
ข.11 รายละเอียดตารางคำสำคัญของหนัง (Movie_Plot_Keys).....	93
ข.12 รายละเอียดตารางทีมงาน (People).....	94
ข.13 รายละเอียดตารางเครดิต (Credits).....	94
ข.14 รายละเอียดตารางรีวิว (Reviews).....	95
ข.15 รายละเอียดตารางบทความ (Articles).....	95
ข.16 รายละเอียดตารางการสนทนา (Discussions).....	96
ข.17 รายละเอียดตารางความคิดเห็น (Comments).....	96
ข.18 รายละเอียดตารางแท็ก (Tags).....	97
ข.19 รายละเอียดตารางการแท็ก (Tag_Entity).....	97
ข.20 รายละเอียดตารางการติดตาม (Followable).....	97
ข.21 รายละเอียดตารางผู้ติดตาม (Follower).....	97

สารบัญรูป

รูป	หน้า
2.1 สัญลักษณ์ ReactJS.....	7
2.2 วงจรชีวิตของ React เวอร์ชัน 16.4 (React life cycle).....	7
2.3 สัญลักษณ์ Next.JS.....	8
2.4 สัญลักษณ์ Node.JS.....	8
2.5 สัญลักษณ์ Express.JS.....	9
2.6 สัญลักษณ์ PyThaiNLP.....	10
2.7 ขั้นตอนการเข้าใจภาษาธรรมชาติ.....	15
2.8 การตัดประโยคแบบตัดตามโครงสร้างไวยากรณ์.....	17
2.9 การตัดประโยคแบบตัดตามหน่วยปริจเฉท.....	17
2.10 กราฟของเส้นโค้งของการถอดยลิจิสติกแสดงที่น่าจะเป็นในการ ผ่านการสอบเทียบกับชั่วโมงที่เรียน.....	19
2.11 กราฟแสดงหลักการของซัพพอร์ตเวกเตอร์แมชชีน.....	20
2.12 แสดง Recurrent Neural Networks แบบการวนลูป.....	22
2.13 แสดง Recurrent Neural Networks แบบแยกในหนึ่งโหนด.....	22
2.14 LSTM Forget Gate.....	23
2.15 LSTM Update Cell State และ Input Gate.....	23
2.16 LSTM Output Gate.....	23
3.1 ภาพรวมของระบบ.....	25
3.2 แผนภาพยูสเคส (Use Case Diagram).....	27
3.3 แผนภาพแสดงความสัมพันธ์ของข้อมูล.....	28
3.4 หน้า Landing page.....	32
3.5 ป๊อปอัพการเข้าสู่ระบบ.....	32
3.6 หน้าลงทะเบียน.....	33
3.7 หน้าหลัก ซึ่งแสดงรายการหนังสือ และเมื่อกดที่หนังสือเรื่องนั้น ๆ.....	33
3.8 หน้าแสดงรายการบทความ.....	34
3.9 หน้าแสดงบทความ.....	34
3.10 หน้าแสดงรายการรีวิว.....	35

สารบัญรูป (ต่อ)

รูป	หน้า
3.11 หน้าแสดงรีวิว.....	35
3.12 หน้าแสดงรายการหัวข้อพูดคุย.....	36
3.13 หน้าแสดงหัวข้อการพูดคุย.....	36
3.14 หน้าแสดงหนัง (แบบเต็ม).....	37
3.15 หน้าโปรไฟล์ของผู้ใช้งาน โดยสามารถดูหัวข้อที่ทำการเก็บบันทึกไว้ได้.....	38
3.16 ป๊อปอัพสำหรับการแก้ไขโปรไฟล์ และดูการติดตามของผู้ใช้.....	38
3.17 เมื่อกดที่นักแสดงหรือบุคคลที่เกี่ยวข้อง จะแสดงหน้าที่มีรายละเอียดของบุคคลนั้น.....	39
3.18 เมื่อทำการค้นหาจากหน้ารีวิว.....	39
3.19 แสดงหน้าการค้นหา เมื่อทำการค้นหาจากแท็ก.....	40
3.20 การออกแบบสำหรับมือถือ.....	40
3.21 การฝึกโมเดล.....	44
4.1 หน้า Landing page.....	45
4.2 หน้าการยืนยันตัวตน.....	45
4.3 หน้าการลงทะเบียนเข้าใช้งาน.....	46
4.4 หน้าหลัก โดยแสดงรายการภาพยนตร์.....	47
4.5 หน้าหลัก เมื่อกดที่โปรเตอร์์ของภาพยนตร์.....	48
4.6 หน้าแสดงรายละเอียดของนักแสดง.....	49
4.7 หน้าแสดงโปรไฟล์ของผู้ใช้งาน.....	49
4.8 หน้าบทความ.....	50
4.9 หน้ารีวิว.....	50
4.10 หน้ากระดานการพูดคุย.....	51
4.11 ฟังก์ชันการกดชอบ เฉยๆ หรือไม่ชอบสำหรับภาพยนตร์เรื่องต่าง ๆ.....	51
4.12 หน้าสำหรับจัดการภาพยนตร์.....	52
4.13 หน้าสำหรับจัดการข้อความทวิต.....	52
4.14 หน้าสำหรับจัดการสมาชิก.....	53
4.15 กราฟแสดงค่าเฉลี่ยของคะแนน TF-IDF ของชุดฝึก.....	62
4.16 กราฟความสัมพันธ์ระหว่าง TF-IDF และสมการ Logistic Function.....	62

บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

ในช่วงไม่กี่ปีที่ผ่านมา การใช้ทวิตเตอร์ (Twitter) ของคนไทยมีอัตราการเติบโตสูงมาก ซึ่งถือได้ว่าเติบโตที่สุดในเอเชียตะวันออกเฉียงใต้ โดยทวิตเตอร์เป็นแพลตฟอร์มสำหรับส่งข้อความที่เรียกว่า ทวิต (Tweet) ผู้ใช้สามารถทวิตข้อความพร้อมติดแฮชแท็ก (hashtag) ในแต่ละวัน ทวิตเตอร์มีผู้ทวิตข้อความเป็นจำนวนมาก หนึ่งในประเภทของการทวิตที่ได้รับความนิยมอยู่เสมอ คือ การทวิตที่เกี่ยวกับวิจารณ์หรือรีวิวกาพย์ยนตร์ โดยทวิตที่เกี่ยวกับภาพยนตร์ส่วนมากจะทำการทวิตพร้อมกับติดแฮชแท็กของภาพยนตร์เรื่องนั้น ๆ เพื่อให้ผู้ที่สนใจเข้ามาอ่านได้ง่าย ซึ่งการที่ผู้ใช้จะอ่านข้อความรีวิว หรือเพื่อเข้าใจภาพรวมความคิดเห็นของผู้ชมทั้งหมดเป็นเรื่องยาก

นอกจากนี้ในปัจจุบันมีผู้ชมภาพยนตร์ทางแพลตฟอร์มออนไลน์เพิ่มมากขึ้น แม้ว่าจะมีระบบแนะนำจากทางแพลตฟอร์มแล้ว แต่ผู้ใช้งานจำนวนมากยังไม่สามารถเลือกได้ว่า จะรับชมภาพยนตร์เรื่องอะไร ทำให้ใช้เวลานานในการตัดสินใจเลือกชม ซึ่งในเว็บไซต์หรือแอปพลิเคชันต่าง ๆ ไม่ได้ทำการรวบรวมหัวข้อที่ช่วยประกอบการตัดสินใจ อาทิเช่น บทความ รีวิว กระดานสนทนา หรือการแนะนำภาพยนตร์เข้าด้วยกัน จึงทำให้ผู้ใช้งานไม่มีความสะดวก เมื่อต้องการหาข้อมูลหรืออ่านบทความของภาพยนตร์ที่ต้องการรับชม

จากปัญหาที่กล่าวมาข้างต้น ทางคณะผู้จัดทำเล็งเห็นว่าในประเทศไทยยังไม่มีเว็บไซต์แห่งรวมภาพยนตร์ โดยเฉพาะที่เป็นสำหรับคนไทยที่รวบรวมบทความ รีวิว หรือกระดานสนทนาของภาพยนตร์เรื่องนั้นเพื่อช่วยในการตัดสินใจ คณะผู้จัดทำจึงสนใจที่จะสร้างเว็บไซต์สำหรับรวบรวมหัวข้อที่เกี่ยวกับภาพยนตร์ โดยแบ่งเป็นส่วนต่าง ๆ ดังนี้

- 1) การแนะนำภาพยนตร์ ซึ่งจะนำทวิตของภาพยนตร์เรื่องนั้น ๆ ไปแบ่งกลุ่มเพื่อตีความว่า ทวิตดังกล่าวเป็นแง่บวก แง่กลาง หรือแง่ลบ แล้วนำไปแสดงหน้าเว็บไซต์เพื่อช่วยผู้ใช้งานในการตัดสินใจ และระบบจะทำการแนะนำ จากภาพยนตร์ที่ผู้ใช้งานกดถูกใจ
- 2) รีวิวภาพยนตร์ ผู้ใช้สามารถเขียนหรืออ่านรีวิวของผู้ใช้คนอื่นได้ เพื่อใช้ในการประกอบการตัดสินใจในการเลือกรับชมภาพยนตร์
- 3) บทความภาพยนตร์ ผู้ใช้สามารถเขียนหรืออ่านบทความของผู้ใช้คนอื่นได้ ตัวอย่างเช่น บทความตีความทางภาพยนตร์ หรือสาระความรู้ต่าง ๆ เป็นต้น
- 4) กระดานสนทนา เป็นที่สำหรับการตั้งคำถามหรือตอบคำถามทางภาพยนตร์ เช่น การถามเรื่องที่ไม่เข้าใจภายในภาพยนตร์ เป็นต้น

ในการแก้ไขปัญหาที่กล่าวมาด้วยวิธีการข้างต้น ทางคณะผู้จัดทำเชื่อว่าจะสามารถทำให้เว็บไซต์ดังกล่าวเป็นเว็บไซต์แรกของประเทศไทยที่เป็นศูนย์รวมของผู้ชื่นชอบในการภาพยนตร์

1.2 วัตถุประสงค์ของโครงการ

- 1) เพื่อสร้างเว็บไซต์ที่เป็นแหล่งศูนย์รวมของผู้ใช้ที่ชื่นชอบในการชมภาพยนตร์
- 2) เพื่อเป็นนำความรู้ด้านการประมวลผลภาษาธรรมชาติ (Natural Language Programming) มาประยุกต์ใช้กับคำวิจารณ์ภาพยนตร์ภาษาไทย เพื่อช่วยแนะนำและประกอบการตัดสินใจในการเลือกชมภาพยนตร์

1.3 ประโยชน์ที่คาดว่าจะได้รับ

- 1) สามารถสร้างเว็บไซต์ที่เป็นแหล่งศูนย์รวมของผู้ใช้ที่ชื่นชอบในการชมภาพยนตร์ได้
- 2) ได้เรียนรู้การใช้โมเดลภาษา (NLP) เพื่อนำข้อความจากทวิตเตอร์มาแบ่งกลุ่มด้านต่าง ๆ
- 3) ได้เรียนรู้การออกแบบส่วนต่อประสานกับผู้ใช้ (Use Interface)
- 4) ได้เรียนรู้และมีความเข้าใจหลักการออกแบบและการใช้งานฐานข้อมูล
- 5) ได้เรียนรู้การพัฒนาเว็บไซต์แอปพลิเคชัน

1.4 ขอบเขตของโครงการ

1.4.1 เว็บไซต์

- 1) ระบบสามารถแนะนำภาพยนตร์ โดยการแบ่งกลุ่มทวิตของภาพยนตร์เป็นแง่ต่าง ๆ เพื่อให้ผู้ใช้เห็นภาพรวมของผู้ชมส่วนใหญ่มากขึ้น หรือผู้ใช้สามารถเลือกภาพยนตร์ที่ชอบ แล้วระบบจะทำการแนะนำภาพยนตร์จากรายการภาพยนตร์ของผู้ใช้
- 2) ผู้ใช้สามารถเขียนรีวิวก่อนภาพยนตร์เรื่องต่าง ๆ หรืออ่านรีวิวจากผู้ใช้คนอื่น โดยสามารถเก็บรีวิวไว้เพื่ออ่านทีหลังได้
- 3) ผู้ใช้สามารถเขียนหรืออ่านบทความของผู้ใช้คนอื่นได้ และเช่นเดียวกับหัวข้อรีวิว ผู้ใช้สามารถเก็บบทความต่าง ๆ ไว้เพื่ออ่านทีหลังได้
- 4) กระดานสนทนา ผู้ใช้สามารถตั้งคำถามในกระดานสนทนาของเว็บไซต์ หรือตอบคำถามของผู้ใช้คนอื่นได้

1.4.2 การจำแนกประเภทข้อความจากทวิตเตอร์

- 1) ระบบสามารถวิเคราะห์ข้อความที่เกี่ยวข้องกับภาพยนตร์ในเรื่องต่าง ๆ จากทวิตเตอร์ มาแบ่งประเภทว่าเป็นแง่บวก แง่กลาง หรือแง่ลบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5 ส่วนประกอบของวิทยานิพนธ์

ปริญญาานิพนธ์นี้ ได้แบ่งเนื้อหาออกเป็น 5 บท ได้แก่ บทนำ ทฤษฎีที่เกี่ยวข้อง การออกแบบ และการพัฒนา การทดลองและผลการทดลอง และ บทสรุปและข้อเสนอแนะ โดยมีรายละเอียด ดังนี้

บทที่ 1 บทนำ กล่าวถึงความเป็นมาของปัญหา วัตถุประสงค์ ประโยชน์ที่คาดว่าจะได้รับ ส่วนประกอบของวิทยานิพนธ์ ขอบเขตของโครงการ และตารางเวลาของโครงการ

บทที่ 2 ทฤษฎีและเครื่องมือที่เกี่ยวข้อง กล่าวถึงทฤษฎี เทคโนโลยีที่เกี่ยวข้อง

บทที่ 3 การออกแบบและพัฒนา กล่าวถึงองค์ประกอบภาพรวม และการพัฒนาระบบ

บทที่ 4 การใช้งานและการทดลอง

บทที่ 5 สรุปผลและข้อเสนอแนะ

1.6 ตารางเวลาของโครงการ

ตารางที่ 1.1 การดำเนินการภาคเรียนที่ 1

ลำดับ	งานที่ลงมือทำ	สิงหาคม				กันยายน				ตุลาคม				พฤศจิกายน				
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	
1	วางแผน ออกแบบ ส่วนประสานงาน ผู้ใช้ และสมัครใช้งานทวิตเตอร์ สำหรับผู้พัฒนา																	
2	ศึกษาหาความรู้ ทางด้าน machine learning และ NLP																	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 1.2 การดำเนินการภาคเรียนที่ 1 (ต่อ)

ลำดับ	งานที่ลงมือทำ	สิงหาคม				กันยายน				ตุลาคม				พฤศจิกายน			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
3	ออกแบบ database และ API ของเว็บไซต์ รวมถึงเริ่มต้นทำ ในส่วนของ front-end																
4	นำข้อมูลที่ได้จาก twitter มา วิเคราะห์ และทำ ความสะอาด ข้อมูล																
5	นำข้อมูลที่ได้จาก ทวิตเตอร์ มา ทดลองนำเข้า โมเดล Logistic Regression และ Naive Bayes																
6	ดึงข้อมูลภาพ ยนตร์สำหรับใช้ สร้างฐานข้อมูล และ implement back-end อีกทั้ง ทำระบบสมาชิก (authentication)																

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 1.3 การดำเนินการภาคเรียนที่ 1 (ต่อ)

ลำดับ	งานที่ลงมือทำ	สิงหาคม				กันยายน				ตุลาคม				พฤศจิกายน			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
7	ทดลองและปรับปรุงโมเดล Deep learning โดยใช้โมเดล ULMFiT และทำโมเดลตัดคำ สำหรับใช้ในการฝึกโมเดล																
8	ทำ component และหน้าเพจของ front-end กำหนด API และฟังก์ชันต่าง ๆ ของ back-end																
9	เชื่อมต่อระหว่าง front-end และ back-end ในส่วนของโปรไฟล์ผู้ใช้ ทำระบบกดถูกใจ และทำรายงานโครงการ																

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 1.4 การดำเนินการภาคเรียนที่ 2

ลำดับ	งานที่ลงมือทำ	มกราคม				กุมภาพันธ์				มีนาคม			
		1	2	3	4	1	2	3	4	1	2	3	4
1	ทำระบบรีวิ บทความ รวมถึง ระบบผู้ดูแล เว็บไซต์												
2	เรียนรู้และใช้งาน Amazon S3 ใน การเก็บรูปภาพ และ Heroku ใน การ deploy												
3	นำระบบหน้าบ้าน และหลังบ้านขึ้น เซิร์ฟเวอร์ (deploy) อีกทั้งเทรนโมเดล โดยมีข้อมูลเพิ่ม มากขึ้น												
4	ปรับปรุงโมเดล การตัดคำ และ ทดลองโมเดล SVM												
5	ปรับปรุงและแก้ เว็บไซต์ เช่น https , admin system, responsive website และ search bar เป็นต้น												

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและเครื่องมือที่เกี่ยวข้อง

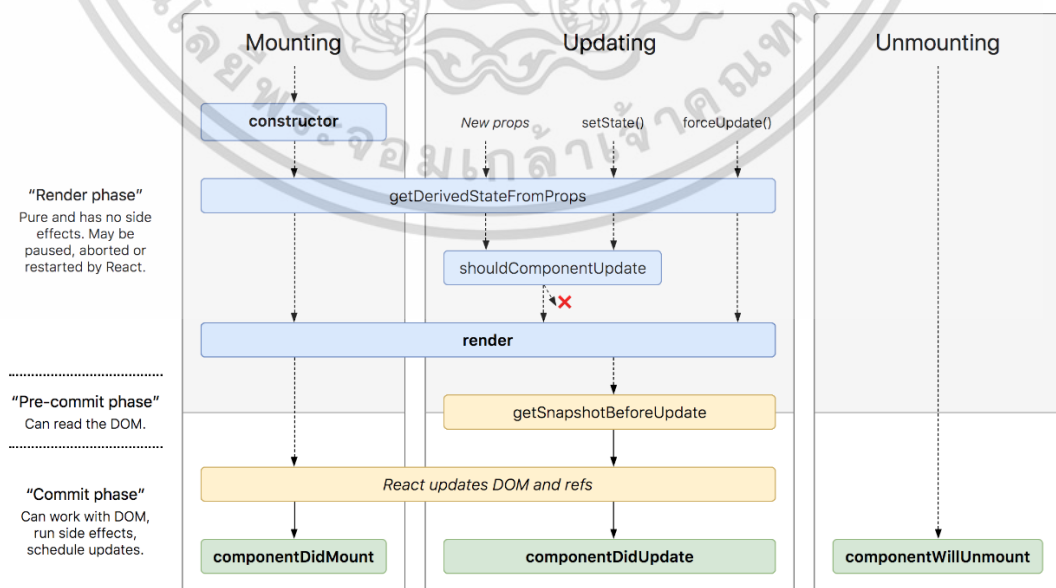
2.1 เทคโนโลยีที่ใช้ด้านเว็บแอปพลิเคชันฝั่งหน้าบ้าน (Front-End)

2.1.1 ReactJS



รูปที่ 2.1 สัญลักษณ์ ReactJS

เป็น Javascript framework สำหรับการสร้างส่วนต่อประสานกับผู้ใช้ (user interface) และรองรับการเขียน JavaScript syntax extension (JSX) โดยสามารถใช้ ReactJS ในการพัฒนาเว็บ Single Page Application (SPA) คือเมื่อเว็บไซต์เปลี่ยนหน้า จะไม่ทำการโหลด HTML ใหม่ นอกจากนี้ยังสามารถแยกส่วนต่าง ๆ ของเว็บไซต์ออกเป็นคอมโพเนนต์ (component) เพื่อสามารถนำกลับมาใช้ใหม่ได้ (reused) โดยสามารถรีเ็นเดอร์คอมโพเนนต์เมื่อมีการเปลี่ยนแปลงข้อมูลหรือสแตต (state) นอกจากนี้ยังมีวงจรชีวิต (life cycle) คือ ฟังก์ชันต่างที่คอยควบคุมเหตุการณ์ที่จะทำการแสดงผล ตามรูปด้านล่าง



รูปที่ 2.2 วงจรชีวิตของ React เวอร์ชัน 16.4 (React life cycle)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 Next.JS



รูปที่ 2.3 สัญลักษณ์ Next.JS

เป็น React framework ที่มีฟีเจอร์ต่าง ๆ ที่เพิ่มเข้ามาเพื่อให้มีความสะดวกในการทำงานมากยิ่งขึ้นเช่น การทำ routing เป็นต้น ซึ่ง Next.JS จะมีการทำ Server Side Rendering เป็นการประมวลผลข้อมูลโดยไม่ต้องรอให้จาวาสคริปต์บนฝั่งไคลเอนต์ (Client) ทำงาน โดยจะเป็นประโยชน์สำหรับการทำ Search Engine Optimization (SEO) คือสามารถติดการค้นหา เมื่อค้นหาเว็บไซต์จาก Google

2.1.3 AdobeXD

โปรแกรมสำหรับใช้ในการออกแบบต้นแบบ (Prototype) ของแอปพลิเคชันต่าง ๆ เช่น แอปพลิเคชันบนมือถือ เว็บไซต์ เป็นต้น ในการออกแบบสามารถเลือกใช้ UI KIT ต่าง ๆ เพื่อนำไปเลือกใช้ให้เข้ากับแอปพลิเคชันที่ต้องการ ซึ่งผู้ใช้สามารถแชร์การออกแบบร่วมกับเพื่อนร่วมทีม หรือส่งปริวิิวให้ผู้อื่นดูได้ นอกจากนี้ยังสามารถบันทึกการออกแบบเพื่อนำไปใช้เป็นแบบในการพัฒนาแอปพลิเคชันต่อไป

2.2 เทคโนโลยีที่ใช้ด้านเว็บแอปพลิเคชันฝั่งหลังบ้าน (Back-End)

2.2.1 Node.JS



รูปที่ 2.4 สัญลักษณ์ Node.JS

เป็น Cross Platform Runtime Environment ซึ่งเป็นแพลตฟอร์มยอดนิยมสำหรับทำ sever side ถูกพัฒนาด้วยภาษาจาวาสคริปต์โดย Google Chrome's JavaScript Engine (V8 Engine) ซึ่งมีจุดเด่นที่สำคัญ เช่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 1) มีไลบรารีเป็นจำนวนมาก และมีการใช้ทรัพยากรน้อย
- 2) ไลบรารีของ Node.JS เป็น อะซิงโครนัส (Asynchronous) ทำให้ไม่ต้องรอการตอบกลับของ API สามารถไปทำงานอื่นได้ หรือเรียกว่า non - blocking I/O
- 3) ไลบรารีมีความเร็วสูงในการประมวลผลของซอร์ซโค้ด (Source code)
- 4) Node.js สามารถใช้โมเดล single thread ที่มีการวนลูปเหตุการณ์ต่าง ๆ โดยจะช่วยให้เซิร์ฟเวอร์ตอบสนองในลักษณะ non-blocking และทำให้เซิร์ฟเวอร์สามารถปรับขนาดได้สูง

2.2.2 Express

express

รูปที่ 2.5 สัญลักษณ์ Express.JS

เป็น web application framework ที่ทำงานบนแพลตฟอร์มของ Node.js ซึ่งมีหน้าที่สำคัญในการสร้าง routing หรือ restful API เช่น ฟังก์ชัน POST เป็นต้น หรือสร้าง middleware ซึ่งเป็นฟังก์ชันที่สามารถแก้ไข HTTP request หรือ HTTP response ให้กับระบบต่าง ๆ เป็นต้น นอกจากนี้ยังสามารถเขียน Express.js ในรูปแบบของ MVC เพื่อเชื่อมต่อกับฐานข้อมูลได้อีกด้วย

2.2.3 Postman

เป็นโปรแกรมสำหรับทดสอบ API ไม่ว่าจะเป็นการ GET POST หรือ PUT เป็นต้น โดยเป็นการส่ง request และรอ response กลับมา ซึ่งสามารถใช้ Postman ในการออกแบบ API ก่อนเริ่มลงมือทำจริง และยังสามารถ Automated Testing ได้

2.3 เทคโนโลยีที่ใช้ด้านฐานข้อมูล

2.3.1 PostgreSQL

เป็นระบบการจัดการฐานข้อมูลเชิงวัตถุ-สัมพันธ์ (RDBMS) ซึ่งสามารถใช้งานได้ฟรี และเป็น โอเพนซอร์ซ (Open Source) โดยรองรับรูปแบบคำสั่งของภาษา SQL ได้เกือบทั้งหมด สามารถทำงานได้หลายระบบปฏิบัติการ และถูกออกแบบมาเพื่อจัดการปริมาณงานหลากหลาย ตั้งแต่เครื่องเดียวไปจนถึงคลังข้อมูล หรือบริการบนเว็บที่มีผู้ใช้งานพร้อมกันจำนวนมาก

2.3.2 DataGrip

เป็นเครื่องมือที่ช่วยในการฐานข้อมูล ที่ช่วยให้นักพัฒนาสามารถทำงานกับฐานข้อมูล SQL ได้ง่ายขึ้น มีไคลเอนต์ที่รองรับฐานข้อมูลต่างๆมากมาย เช่น MariaDB, Microsoft SQL Server, MySQL, Oracle , PostgreSQL เป็นต้น

2.4 เทคโนโลยีด้านการเรียนรู้ของเครื่อง (Machine Learning)

2.4.1 Python

เป็นภาษาโปรแกรมคอมพิวเตอร์ระดับสูง ถูกออกแบบมาให้เป็นภาษาสคริปต์ที่อ่านง่าย สร้างโดย Guido van Rossum และเปิดตัวครั้งแรกในปี 1991 โดยหลักการออกแบบของภาษาที่เน้นอ่านโค้ดได้ง่าย เป็นภาษาที่มีโครงสร้างและวิธีการเขียนโปรแกรมเชิงวัตถุ ซึ่งช่วยให้การเขียนโค้ดชัดเจน และเหมาะสมสำหรับทั้งโครงการขนาดเล็กและขนาดใหญ่ นอกจากนี้เป็นโอเพนซอร์ซ โดยมีไลบรารีที่สนับสนุนคำนวณทางคณิตศาสตร์ชั้นสูงมากมาย และถูกประยุกต์ใช้งานทางด้านวิทยาศาสตร์อย่างมาก

2.4.2 Tensorflow

เป็นแพลตฟอร์มโอเพนซอร์ซของ Google มีเครื่องมือและไลบรารีที่มีความยืดหยุ่นและครอบคลุมสำหรับการทำ machine learning ซึ่งสามารถทำงานทั้งบนซีพียูและจีพียู นอกจากนี้ยังมี TensorBoard ซึ่งเป็นการจำลองการทำงานของ TensorFlow และ TensorFlow Lite สำหรับการประมวลผล deep learning บนมือถือ

2.4.3 Keras

เป็น deep learning library ภาษาไพทอน (Python) โดยส่วนมากแอปพลิเคชัน deep learning ที่สร้างโดย Tensorflow สามารถทำได้ด้วย Keras ซึ่ง API ง่ายต่อการเข้าใช้งาน โดยสามารถทำงานบน TensorFlow CNTK หรือ Theano อีกทั้งยังสามารถทำงานได้ทั้งซีพียูและจีพียู นอกจากนี้ช่วยให้การสร้างต้นแบบที่ง่ายและรวดเร็วโดยเป็นมิตรต่อผู้ใช้งาน หรือการขยายขอบเขตของงาน เป็นต้น ทำให้ผู้เริ่มต้นสามารถทำการศึกษาได้ไม่ยาก

2.4.4 PyThaiNLP



รูปที่ 2.6 สัญลักษณ์ PyThaiNLP

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เป็นโมดูลภาษาไทยที่พัฒนาขึ้นเพื่อใช้ในการประมวลผลภาษาธรรมชาติในภาษาไทย โดยมีฟังก์ชันสำหรับการตัดคำ การตรวจสอบคำผิด การหาประเภทของคำ หรือการวิเคราะห์อารมณ์ทางข้อความ เป็นต้น

2.5 หลักภาษาไทย

2.5.1 ชนิดของคำ

คำ เป็นหน่วยที่เล็กที่สุดซึ่งมีความหมายในตัว ใช้ประกอบหน้าคำอื่นมีความหมาย เช่นนั้น คำในภาษาไทย แบ่งออกเป็น 7 ชนิด คือ

2.5.1.1 คำนาม คือ คำที่ใช้เรียกชื่อ คน สัตว์ สิ่งของ สถานที่ อาคาร สภาพ และลักษณะทั้งสิ่งที่มีชีวิตและไม่มีชีวิต ทั้งที่เป็นรูปธรรมและนามธรรมคำนามแบ่งออกเป็น 5 ชนิด

- 1) สามานยนาม คือ คำนามสามัญที่ใช้เป็นชื่อทั่วไป เช่น คน รถ ปลา
- 2) วิสามานยนาม คือ คำนามที่เป็นชื่อเฉพาะของคน สัตว์ สถานที่ หรือเป็นคำเรียกบุคคล สถานที่เพื่อเจาะจง เช่น โรงเรียนสตรีวิทยา วัดพระแก้ว
- 3) ลักษณะนาม คือ คำนามที่ทำหน้าที่ประกอบนามอื่น เพื่อบอกรูปร่าง ลักษณะขนาดหรือปริมาณ เช่น แท่ง เรือ น หลัง อัน
- 4) สมุหนาม คือ คำนามบอกหมวดหมู่ของสามานยนาม และวิสามานยนามที่รวมกันมาก ๆ เช่น ผุ่่นก คณะนักเรียน เหล่าลูกเสือ
- 5) อาการนาม คือ คำเรียกสิ่งที่ไม่รูปร่าง ไม่มีขนาด จะมีคำว่า "การ" และ "ความ" นำหน้าคำกริยา หรือคำวิเศษณ์ เช่น การกิน การนอน การเรียน ความสวย ความคิด

2.5.1.2 คำสรรพนาม คือ คำที่ใช้แทนนามในประโยคสื่อสาร แบ่งเป็น 6 ชนิด ได้แก่

- 1) บุรุษสรรพนาม (สรรพนามที่ใช้แทนคน) เช่น ฉัน ดิฉัน ผม ท่าน มัน
- 2) ประพันธสรรพนาม (สรรพนามที่ใช้เชื่อมประโยค) เช่น ที่ ซึ่ง อัน
- 3) วิภาคสรรพนาม (สรรพนามบอกความชี้ชัด) ใช้แทนคำนามที่อยู่ข้างหน้า ได้แก่ บ้าง ต่าง กัน
- 4) นิยมสรรพนาม (สรรพนามชี้เฉพาะ) ใช้แทนคำนามที่กล่าวถึงที่อยู่ เช่น นี่ นี้ นั่น นั้น โน้น อย่างนี้
- 5) อนิยมสรรพนาม (สรรพนามบอกความไม่เจาะจง) เช่น ใคร อะไร ที่ไหน ผู้ใด สิ่งใด ใคร ๆ อะไร ๆ
- 6) ปจฉาสรรพนาม (สรรพนามที่เป็นคำถาม) เช่น ใคร อะไร ไหน ผู้ใด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.1.3 คำกริยา คือ คำที่แสดงอาการ สภาพ หรือการกระทำของคำนาม และคำสรรพนาม ในประโยค คำกริยาบางคำอาจมีความหมายสมบูรณ์ในตัวเอง บางคำต้องมีคำอื่นมาประกอบและ บางคำต้องไปประกอบคำอื่นเพื่อขยายความชนิดของคำกริยา คำกริยาแบ่งออกเป็น 5 ชนิด ดังนี้

- 1) **อกรรมกริยา** (กริยาที่ไม่ต้องมีกรรมมารับ) เช่น แม่ยื่น รถติด
- 2) **สกรรมกริยา** (กริยาที่ต้องมีกรรมมารับ) เช่น น้องกินเค้ก สมใจล้างจาน
- 3) **วิกตรรถกริยา** (คำที่มารับไม่ใช่กรรมแต่เป็นส่วนเติมเต็ม) เช่น เป็น เหมือน คล้าย เท่า คือ เสมือน ดูจ
- 4) **กริยานุเคราะห์** (กริยาช่วย) คือ คำที่เติมหน้าคำกริยาหลักในประโยคเพื่อช่วยขยายความหมายของคำกริยาสำคัญ เช่น กำลัง จะ ได้ แล้ว ต้อง อย่า จง โปรด ช่วย ควร คงจะ อาจจะ
- 5) **กริยาสภามลา** คือ กริยาที่ทำหน้าที่เป็นคำนามอาจเป็นประธาน หรือกรรม หรือทขยายของประธานเช่น เขาซ้อมร้องเพลงทุกวัน

2.5.1.4 คำวิเศษณ์ คือ คำที่ใช้ขยายคำอื่น ได้แก่ คำนาม คำสรรพนาม คำกริยา หรือคำวิเศษณ์ ให้มีความหมายชัดเจนขึ้น บอกลักษณะ ชนิด ขนาด สี เสียง กลิ่น รส อาการ เป็นต้น

2.5.1.5 คำบุพบท คือ คำที่มีหน้าที่เชื่อมคำ หรือกลุ่มคำเพื่อแสดงความสัมพันธ์กับคำอื่น ๆ

2.5.1.6 คำสันธาน คือ คำที่ทำหน้าที่เชื่อมประโยคกับประโยค เชื่อมข้อความกับข้อความให้สละสลวย เช่น ทั้ง...และ กว่า...ก็ เพราะ...จึง หรือ...ก็

2.5.1.7 คำอุทาน คือ คำที่เปล่งออกมาเพื่อแสดงอารมณ์หรือความรู้สึกของผู้พูด มักจะเป็นคำที่ไม่มีความหมาย แต่เน้นความรู้สึกและอารมณ์ของผู้พูด เฮ้ ดายจริง ้วย ลืมหูลืมตา

2.5.2 การสร้างคำ

2.5.2.1 คำมูล คือ ที่มีความหมายสมบูรณ์ในตัวเอง เป็นคำดั้งเดิมที่มีในภาษาเดิมต้องไม่ใช่คำที่เกิดจากการประสมกับคำอื่น ๆ เช่น ช้าง ป่า ม้า วัว ควาย ใจ ขนม ทะเล นริ กะทิ

2.5.2.2 คำประสม คือ การนำคำมูลที่มีความหมายไม่เหมือนกัน ตั้งแต่สองคำขึ้นไปนำมา รวมกันแล้วเกิดความหมายใหม่ แต่ยังมีเค้าความหมายเดิมอยู่ เช่น หมอความ แม่ยาย กินที่ น้ำหวาน โรงเรียน หมดตัว เตารีด

2.5.2.3 คำซ้อน คือ การนำคำมูลที่มีความหมายหรือเสียงใกล้เคียงกัน หรือเหมือนกันมา ซ้อนกันแล้วทำให้เกิดความหมายใหม่หรือความหมายใกล้เคียงกับความหมายเดิม บ้านเรือน อ้วนพี จิตใจ รูปร่าง ชัดดู ใจคอ หูตา ญาติโยม ถ้วยชาม คัดเลือก ทูตดี คุณคิม คับแค้น อ่อนหวาน อบรม ชั่วดี ผิดถูก เป็นตาย ร้ายดี ขับขี้ แข็งขัน งอแง

2.5.2.4 คำซ้ำ คือการนำคำมูลคำเดียวกันมากล่าวซ้ำ เพื่อเน้นน้ำหนักของคำ เช่น พี่ ๆ น้อง ๆ ไป ๆ มา ๆ เบา ๆ กลัว ๆ เป็นต้น

2.5.2.5 คำสมาส คือ วิธีการสร้างคำที่มาจากภาษาบาลีและสันสกฤต ซึ่งเกิดเป็นคำใหม่ มีความหมายใหม่ เช่น ประวัติกาศาสตร์ วิศวกรรม วัฒนธรรม

2.5.2.6 คำสนธิ คือ การสมาสโดยการเชื่อมคำเข้าระหว่างพยางค์หลังของคำหน้ากับพยางค์หน้าของคำหลัง เรียกว่า การสมาสที่มีสนธิ หรือ คำสมาสแบบกลมกลืนเสียง เป็นการย่ออักษรให้น้อยลง เช่น ธรรมาธิปไตย (ธรรม + อธิปไตย) นโยบาย (นย + อุบาย)

2.5.3 วากยสัมพันธ์

วากยสัมพันธ์ คือ ความสัมพันธ์ของคำในประโยค ซึ่งภาษาไทยเป็นภาษาคำโดด คำในภาษาไทยจะไม่มี การเปลี่ยนแปลงรูปไม่ว่าจะอยู่ในกาล (tense) การก (case) มาลา (mood) หรือ วาก (voice) ใดก็ตาม คำในภาษาไทยไม่มีลิงก์ (gender) ไม่มีพจน์ (number) คำในภาษาไทยหลายคำจึงไม่สามารถกำหนดหน้าที่ของคำตายตัวลงไปได้ ต้องอาศัยบริบทเข้าช่วยในการพิจารณา เมื่อต้องการจะผูกประโยคจะนำเอาคำแต่ละคำมาเรียงติดต่อกันเข้า โดยภาษาไทยมีโครงสร้างแตกต่างกันไป ทางขวา และคำคุณศัพท์จะถูกลวางไว้หลังคำนาม ซึ่งลักษณะทางวากยสัมพันธ์โดยรวมแล้วจึงเป็นแบบ “ประธาน-กริยา-กรรม”

2.5.3.1 วลี

วลี คือ คำหลายคำที่นำมาเรียงชิดติดกัน มีใจความไม่สมบูรณ์เพราะขาดส่วนใดส่วนหนึ่งในภาคประธานหรือภาคแสดง อาจใช้โดยลำพังได้ หรือใช้ประกอบประโยค เช่น นักเรียนโรงเรียนสตรีวิทยา พี่และน้อง ประเทศในแถบเอเชีย การใช้สิทธิ

2.5.3.2 ประโยค

ประโยค คือ ถ้อยคำที่นำมาเรียงกันแล้วเกิดใจความที่สมบูรณ์ ซึ่งประกอบไปด้วยภาคประธานและภาคแสดง โดยภาคประธาน คือ คำหรือกลุ่มคำที่ทำหน้าที่เป็นประธาน เป็นผู้กระทำ ผู้แสดงของประโยคซึ่งจะมี บทประธาน บทขยายประธาน โดยบทขยายอาจจะไม่มีหรือมีก็ได้ และภาคแสดง คือ คำหรือกลุ่มคำที่ประกอบไปด้วย บทกริยา บทกรรมและส่วนเติมเต็ม โดยบทกริยาจะทำหน้าที่เป็นตัวแสดงของประโยค บทกรรมจะทำหน้าที่เป็นผู้ถูกกระทำ และส่วนเติมเต็มทำหน้าที่เสริมใจความของประโยคให้สมบูรณ์

ชนิดของประโยค

- 1) **ประโยคความเดียว** คือ ประโยคที่มีเนื้อความเดียว มีบทกริยาเพียงบทเดียว เช่น นกบิน พ่อกินผลไม้ เธออ่านหนังสือในห้องสมุด

- 2) **ประโยคความรวม** คือ ประโยคที่นำเอาประโยคความเดียวตั้งแต่สองประโยคขึ้นไปมารวมกัน โดยมีคำสันธานเป็นตัวเชื่อม โดยอาจมีเนื้อความคล้ายตามกัน ขัดแย้งกัน เป็นเหตุเป็นผลกัน หรือ ให้เลือกเอาอย่างใดอย่างหนึ่ง
- 3) **ประโยคความซ้อน** คือ ประโยคที่มีใจความสำคัญเพียงใจความเดียว ซึ่งประกอบด้วยประโยคหลัก และประโยคย่อย โดยประโยคย่อยทำหน้าที่ประกอบขยายส่วนใดส่วนหนึ่งในประโยคหลัก โดยอาจมีคำเชื่อม เช่น ให้ ว่า ที่ ซึ่ง อัน ผู้ อัน เมื่อ เพราะ จน ตาม

2.6 การประมวลผลภาษาธรรมชาติ (Natural Language Processing)

การประมวลผลภาษาธรรมชาติ เป็นส่วนหนึ่งของสาขาปัญญาประดิษฐ์ (Artificial Intelligence) และภาษาศาสตร์คอมพิวเตอร์ (Computational Linguistics) ที่ศึกษาถึงปัญหาในการประมวลผลและการใช้งานภาษาธรรมชาติ ซึ่งคือภาษาทั่วไปที่เกิดจากวัฒนธรรมของมนุษย์ที่ใช้สื่อสาร และมีการเปลี่ยนแปลงไปตามวัฒนธรรมของผู้ใช้ภาษา โดยไม่มีรูปแบบหรือกฎเกณฑ์ของภาษาตายตัวที่แน่นอน เช่น ภาษาไทย, ภาษาอังกฤษ, ภาษาจีน เป็นต้น เพื่อให้คอมพิวเตอร์สามารถเข้าใจภาษาธรรมชาติเพื่อใช้ประโยชน์ในด้านการสื่อสาร และการวิเคราะห์ข้อมูล เช่น เครื่องแปลภาษา (Machine translation), การรู้จำเสียงพูด (Speech recognition), การค้นคืนสารสนเทศ (Information retrieval), การจดจำตัวอักษร (Optical character recognition), การทำเหมืองข้อความ (Text mining) เป็นต้น

2.6.1 ขั้นตอนการเข้าใจภาษาธรรมชาติ

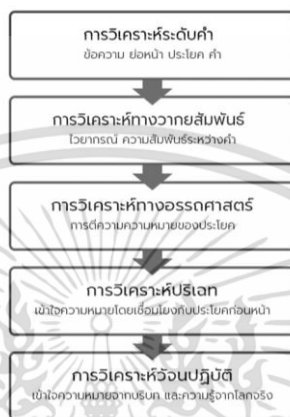
แบ่งเป็นขั้นตอนใหญ่ 5 ขั้นตอน ดังนี้

- 1) **การวิเคราะห์ระดับคำ (Lexical Analysis)** การระบุและวิเคราะห์โครงสร้างของคำศัพท์ ชุดของคำและวลีในภาษา เกี่ยวข้องกับการตัดคำ (Word segmentation) การระบุคำที่เป็นนิพจน์ระบุนาม (Named Entities Recognition - NER) การกำหนดหน้าที่ของคำ (Part-Of-Speech Tagging)
- 2) **การวิเคราะห์ทางวากยสัมพันธ์ (Syntactic Analysis)** เกี่ยวข้องกับการวิเคราะห์โครงสร้างประโยคตามหลัก และการจัดเรียงคำในลักษณะที่แสดงความสัมพันธ์ระหว่างคำเข้าใจประโยคโครงสร้าง การแบ่งประโยค
- 3) **การวิเคราะห์ทางอรรถศาสตร์ (Semantic Analysis)** การวิเคราะห์ความหมายของคำในประโยค เข้าใจความหมายของคำและความหมายของประโยค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) การวิเคราะห์ปริบท (Discourse Integration) การวิเคราะห์ความเชื่อมโยงของประโยค เข้าใจความหมายโดยเชื่อมโยงกับความหมายของประโยคก่อนหน้า

5) การวิเคราะห์วจนปฏิบัติ (Pragmatic Analysis) การตีความหมายของของคำและประโยคโดยอิงจากความรู้เดิมและขึ้นอยู่กับสถานการณ์ เข้าใจความหมายที่รูปประโยคไม่สื่อความหมายโดยตรง



รูปที่ 2.7 ขั้นตอนการเข้าใจภาษาธรรมชาติ

2.6.2 การประมวลผลภาษาไทย

ภาษาไทยเป็นภาษาคำโดด โดยใช้คำวิเศษณ์ เช่น คุณศัพท์ และกริยาวิเศษณ์ เพื่อบอกคุณสมบัติทางไวยากรณ์แทนการผันคำ ซึ่งภาษาไทยมีลำดับคำแน่นอน คือ ประธาน-กริยา-กรรม และคำที่มีความหมายซับซ้อนเกิดจากการประกอบสร้างของคำที่มีหน่วยความหมายเล็ก ๆ ซึ่งทำให้เกิดความเรียบง่าย แต่ซับซ้อนของภาษา ภาษาไทยจึงมีความกำกวมในทุกระดับของการประมวลผลดังนี้

2.6.2.1 การวิเคราะห์ระดับคำ (Lexical Analysis)

ภาษาไทยไม่ใช่เว้นวรรคเพื่อบ่งบอกขอบเขตของคำ ไม่มีการใช้ตัวอักษรใด ๆ ในการบ่งบอกขอบเขตของคำอย่างชัดเจน ระดับพื้นฐานที่สุดจึงต้องเริ่มจากการตัดคำ (word segmentation) ซึ่งความกำกวมของขอบเขตของคำในภาษาไทยนั้นมีสูงมาก ตัวอย่างเช่น ตามหามเหสี สามารถตัดคำได้เป็น ตาม|หาม|เหสี, ตาม|หาม|เหสี, และ ตาม|หาม|เหสี

เทคนิคที่ประยุกต์ใช้ในการตัดคำสามารถแบ่งออกเป็น 3 วิธีหลัก ได้แก่

- 1) การใช้กฎไวยากรณ์ทางภาษา (Rule-based) วิธีนี้ไม่จำเป็นต้องเก็บพจนานุกรมไว้ในหน่วยความจำ และค่อนข้างรวดเร็ว แต่ยากที่จะสร้างกฎที่สมบูรณ์แบบ

ตามหลักไวยากรณ์ ซึ่งไม่สามารถจัดการคำที่ไม่แปลได้อย่างมีประสิทธิภาพ โดยเฉพาะคำที่ทับศัพท์

- 2) การอ้างอิงคำจากพจนานุกรม (Dictionary-based) เป็นวิธีที่มีอัตราความแม่นยำสูง โดยแก้ปัญหาคำที่ไม่รู้จักโดยการอัปเดตพจนานุกรม แต่ต้องใช้หน่วยความจำขนาดใหญ่เพื่อจัดเก็บพจนานุกรม และยากที่จะดูแลให้พจนานุกรมมีความทันสมัย
- 3) การสร้างโมเดลเรียนรู้จากฐานข้อความขนาดใหญ่ (Machine Learning or Corpus based) แก้ปัญหาคำที่ไม่รู้จักบางส่วนโดยเรียนรู้จากคลังข้อมูลขนาดใหญ่ ซึ่งสามารถสร้างกฎได้เอง แต่ต้องการคลังในการเรียนรู้ขนาดใหญ่และเวลาการเรียนรู้กับอัลกอริทึมการเรียนรู้ที่ดี

2.6.2.2 ระดับไวยากรณ์ (Syntax)

ไวยากรณ์ภาษาไทยมีความซับซ้อนในการใช้หน่วยคำที่มีความหมายขนาดเล็กมาประกอบเป็นคำใหม่ที่มีความหมายขนาดใหญ่ จึงทำให้ภาษาไทยมีความกำกวมระหว่างนามวลีและประโยคภายใต้รูปแบบของ นาม-กริยา-นาม เช่น คนขับรถชนเด็ก ซึ่งอาจแบ่งได้เป็น คนขับรถ ชนเด็ก หรือ คน ขับรถ ชน เด็ก เป็นต้น

2.6.2.3 ระดับบรรดาศาสตร์ (Semantics)

ในด้านความหมายของคำ ภาษาไทยเป็นภาษาที่มีคำพ้องรูป คำพ้องเสียง การตีความความหมายระดับลึกจึงเป็นเรื่องที่ยาก ซึ่งในการตีความด้านทัศนคตินั้นมีเทคนิคในการเลือกคำอยู่ เช่น การเลือกคำที่สื่อถึงทัศนคติ ได้แก่ คำคุณศัพท์ เช่น อร่อย, สวย, ดี คำนาม เช่น หนังสือเล่มนี้เป็นงานชิ้นโบว์แดง คำกริยา เช่น ชอบ ปลื้ม รัก เกลียด ชอบ คำวิเศษณ์ เช่น อย่างช้าของ แต่ในบางครั้งเราก็ต้องใช้การตีความจากบริบทนอกเหนือจากรูปของภาษาในการตีความหมายด้วย โดยสิ่งที่ต้องระมัดระวังของการตีความหมายมีดังนี้

- คำที่ทำความเป็นจริงเปลี่ยนไป (Nonveridicality) เช่น หนังสือเล่มนี้น่าจะจบได้ดีกว่านี้
- คำที่ทำให้กลายเป็นประโยคปฏิเสธ (Negation) เช่น ไม่เคยผิดหวังเลยจนมาภาคนี้
- การเติมคำลบแล้วความหมายไม่เปลี่ยนซ้ำ (Negating Adverb) เช่น ไม่ดีเยี่ยมหมายความว่า ไม่ได้ดีมาก แต่ก็ยังมีความหมายคืออยู่ ไม่อุบาทว์ ก็ยังมีความหมายในแง่ลบอยู่
- การเสียดสี (Sarcasm) เช่น อร่อยจนอยากจะทำหนังสือเล่มนี้เหมาะแก่การใช้ทับกระดาษ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6.2.4 ระดับปริจเฉท (Discourse)

ในการศึกษาความเชื่อมโยงและขยายกันของหน่วยประโยค เพื่อประกอบสร้างเป็นเอกสารขนาดใหญ่ แต่ภาษาไทยมีปัญหาด้านการตัดประโยค เพราะไม่มีเครื่องหมายจบประโยค ทำให้ขอบเขตของประโยคมีความกำกวม แม้ว่าภาษาไทยจะมีเครื่องหมายเว้นวรรค แต่เครื่องหมายวรรคมีหน้าที่หลายประการ ได้แก่ จบประโยค พักหายใจ คั่นไม้ยมก ยกตัวอย่าง คั่นชื่อ ตัวเลข และ คำในภาษาต่างประเทศเพื่อความสวยงาม

เกณฑ์ในการตัดประโยค 2 แบบ ได้แก่

แบบตัดตามโครงสร้างไวยากรณ์ โดยเน้นความสัมพันธ์ระหว่างประโยคใหญ่ (main clause) กับประโยคย่อย (subordinate clause) ซึ่งทำให้โครงสร้างประโยคมีขนาดใหญ่ มาก เช่น

อย่างไรก็ตาม อดีตประธาน ทปอ. **กล่าว**ว่า_c [[มีการ**หักท้วง**เรื่องนี้อย่าง**หมดจด**ว่า_c [มี**ช่วงเวลาว่างนานขนาดนี้** **ทำไมถึง**ยังต้องมีการ**จัด**สอบนอกเหนือจากนี้**อีก**] **เพราะ**การสอบ**ล่วงหน้า**ไปในเวลาระหว่างเรียนมัธยมนั้น**กระทบ**กับเรื่องอื่นๆ **โดยเฉพาะ**การเรียนในชั้น**เป็น**วงจรรูกลูกโซ่] [แนวโน้มที่เข้ามาแก้เรื่องนี้ **เป็น**ความคิดที่ดี **แต่**ยังไม่**เห็น**เรื่องใช้ผลการเรียนในชั้นมาเป็นองค์ประกอบ**รับตรง** ซึ่ง**อาจ**ทำให้เด็กไม่สนใจห้องเรียน **และ**มุ่ง**กวดวิชา** ทำให้**การสอบเข้าอุดมศึกษา**เป็น**จำเลย**ข้อหาทำลายระบบการศึกษาขั้นพื้นฐาน **จน**ไป**สู่**ปัญหา**ต่างๆ** ได้]]

รูปที่ 2.8 การตัดประโยคแบบตัดตามโครงสร้างไวยากรณ์

แบบตัดตามหน่วยปริจเฉท โดยเน้นแยกเนื้อความเป็นใจความย่อย โดยจะทำให้ไม่เห็นความสัมพันธ์ระหว่างประโยคใหญ่กับประโยคย่อย เช่น

อย่างไรก็ตาม อดีตประธาน ทปอ. **กล่าว**ว่า_c | มีการ**หักท้วง**เรื่องนี้อย่าง**หมดจด**ว่า_c | มี**ช่วงเวลาว่างนานขนาดนี้** | **ทำไมถึง**ยังต้องมีการ**จัด**สอบนอกเหนือจากนี้**อีก** | **เพราะ**การสอบ**ล่วงหน้า**ไปในเวลาระหว่างเรียนมัธยมนั้น**กระทบ**กับเรื่องอื่นๆ | **โดยเฉพาะ**การเรียนในชั้น**เป็น**วงจรรูกลูกโซ่ | แนวโน้มที่เข้ามาแก้เรื่องนี้ **เป็น**ความคิดที่ดี | **แต่**ยังไม่**เห็น**เรื่องใช้ผลการเรียนในชั้นมาเป็นองค์ประกอบ**รับตรง** | ซึ่ง**อาจ**ทำให้เด็กไม่สนใจห้องเรียน | **และ**มุ่ง**กวดวิชา** | ทำให้**การสอบเข้าอุดมศึกษา**เป็น**จำเลย**ข้อหาทำลายระบบการศึกษาขั้นพื้นฐาน | **จน**ไป**สู่**ปัญหา**ต่างๆ** ได้

รูปที่ 2.9 การตัดประโยคแบบตัดตามหน่วยปริจเฉท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6.2.5 ระดับวจนปฏิบัติ (Pragmatics)

การตีความข้อความให้เห็นความหมายระดับลึกตาม การวิเคราะห์ในระดับนี้จึง ต้องใช้ความรู้ทางโลก และสามัญสำนึกเข้าช่วย ซึ่งแทบอยู่เกินวิสัยที่คอมพิวเตอร์จะอนุมานจาก บริบทได้ เช่น การกล่าวว่า “ร้อนจังเลย” สื่อถึงการให้เปิดหน้าต่าง หรือการกล่าวว่า “ก็โหมงแล้ว” ในบางสถานการณ์คือ การว่ากล่าว ซึ่งการสร้างคลังข้อมูลภาษาที่เก็บข้อมูลวจนปฏิบัติระดับนี้ทำได้ ยากมาก เพราะต้องอาศัยมนุษย์ตีความแบบลึกซึ้ง

ภาษาในสังคมออนไลน์

ภาษาในสังคมออนไลน์จัดเป็นทำเนียบภาษา (language register) แบบหนึ่ง ซึ่งมีผู้ใช้ ภาษาเฉพาะกลุ่มอยู่ในโลกออนไลน์ โดยภาษาเหล่านี้มีความเปลี่ยนแปลงสูงมาก เพราะมีคำใหม่ เกิดขึ้นและหายไปอย่างรวดเร็ว ซึ่งในการทำการประมวลผลทางภาษาไทยจึงต้องมีการทำให้คำ เหล่านี้กลายเป็นภาษาปกติ (normalization) ก่อนที่จะเข้าสู่กระบวนการปกติต่อไป โดยคำใหม่ เหล่านี้มีที่มาจากหลายสาเหตุ เช่น

การทำให้ภาษาใช้ง่ายขึ้น (simplification) เช่น ซิมิ (ใช้ใหม่) มะรุ (ไม่รู้)

เหตุการณ์โด่งดัง เช่น เสียงสูง ใหญ่กว่าที่คิด

พิมพ์ผิด (typos) เช่น เมพ ครัช คนับ ลาก่อย

การลากเสียงเข้า (assimilation) เช่น แวรั้ง อัลไล ตัลลลลล

ลักษณะเฉพาะ เช่น นก

การทำให้น่ารัก (dramatization) เช่น ตั้ลลลล จู่มู จีมู

การทำให้รื่นหู (euphemism) เช่น ห่าน เหี้ยม หมี่

คำพวน เช่น ดิสกว่าที่คีย์ เบอว่ารักแอบ ฯลฯ

คำยืม (loan words) เช่น สก๊อย

การประชด (sarcasm) เช่น ตีออก

การพิมพ์ซ้ำ เช่น คำาาาาาาาา สนุกกกกกก มากกกก

2.7 การเรียนรู้ของเครื่อง (Machine Learning)

2.7.1 Logistic Regression

การถดถอยโลจิสติก เป็นโมเดลทางคณิตศาสตร์โดยใช้โลจิสติกฟังก์ชัน เพื่อทำการ แบ่งกลุ่ม (Classification) ซึ่งสามารถแบ่งออกได้เป็น 2 กลุ่ม คือ การใช้การถดถอยโลจิสติก แบ่งกลุ่มออกเป็นสองกลุ่ม และการใช้การถดถอยโลจิสติกแบ่งกลุ่มมากกว่าสองกลุ่ม เช่น การ วิเคราะห์อารมณ์ของข้อความที่แบ่งออกเป็น แบ่งบวก แบ่งกลาง หรือแบ่งลบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

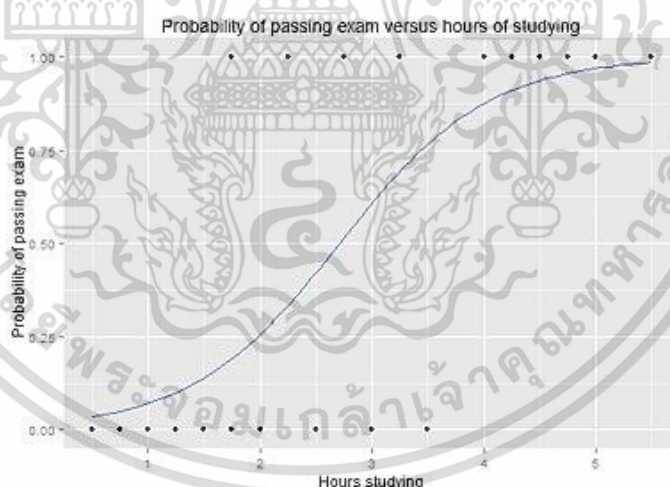
จากสมการการถดถอยแบบเส้นตรง (linear regression) สมการจะอยู่ในรูปแบบของ

$$y = a + bX$$

สำหรับการวิเคราะห์การถดถอยโลจิสติกจะเป็นการแบ่งกลุ่ม โดยไม่อยู่ในรูปแบบของการทำนายค่าประมาณของตัวเลขตามการถดถอยแบบเส้นตรง ซึ่งสามารถเขียนอยู่ในรูปแบบสมการ

$$p(y) = \frac{1}{1 + e^{-f(x)}}$$

ซึ่งฟังก์ชัน $f(x)$ เป็นการทำนายของสมการการถดถอยแบบเส้นตรงหรือการถดถอยแบบง่าย โดยจากสมการข้างต้นจะได้ความน่าจะเป็นในการเกิดเหตุการณ์ y เช่น ความน่าจะเป็นที่ข้อความจะเป็นแฉ่บวค เป็นค้ัน ซึ่งจะมีการตั้งจุดอ้่างอิง (threshold) เพื่อทำการวัดค่าความน่าจะเป็น เช่นเมื่อค่าความน่าจะเป็นที่ได้มีค่ามากกว่าหรือเท่ากับ 0.5 จะทำการทำนายค่า y เป็น 1 หรือกลุ่มที่ 1 โดยสามารถอธิบายโดยรูปได้ดังรูปด้านล่าง



รูปที่ 2.10 กราฟของเส้นโค้งของการถดถอยโลจิสติกแสดงความน่าจะเป็นในการผ่านการสอบเทียบกับชั่วโมงที่เรียน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.2 Naive Bayes Classifier

Naive Bayes เป็นประเภทของการทำการแบ่งกลุ่ม โดยใช้ทฤษฎีของ Bayes เป็นการทำนายประเภทของกลุ่มด้วยความน่าจะเป็นของสิ่งยังไม่เกิดขึ้น ซึ่งจะทำการคาดเดาจากสิ่งที่เคยเกิดขึ้นในอดีตมาก่อน ถ้าคำนวณกลุ่มไหนที่มีความน่าจะเป็นสูงสุด โมเดลจะทำการทำนายให้อยู่กลุ่มนั้น เช่น Maximum A Posteriori (MAP) โดยมีรูปแบบของสมการดังด้านล่าง

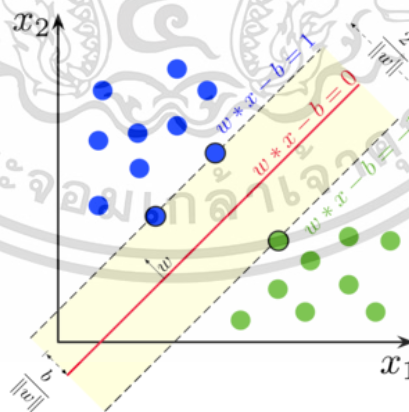
$$MAP(H) = \max(P(H|E))$$

หรือ
$$MAP(H) = \max\left(\frac{P(H|E)P(H)}{P(E)}\right)$$

จากสมการข้างต้นเป็นการหาความน่าจะเป็นสูงสุดที่เหตุการณ์ E เกิดขึ้นก่อนและตามด้วยเหตุการณ์ H ซึ่งสามารถแตกสมการออกได้ตามทฤษฎีของ Bayes

2.7.3 Support Vector Machine (SVM)

ซัพพอร์ตเวกเตอร์แมชชีน เป็นรูปแบบการเรียนรู้ของเครื่องที่ใช้อัลกอริทึมแก้ปัญหาในการจำแนกประเภทของกลุ่ม โดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ฝึกสอนเพื่อให้ได้เส้นแบ่งไฮเปอร์เพลน (hyperplane) ในพื้นที่ N มิติ (N คือจำนวนของฟีเจอร์) ซึ่งเป็นเส้นแบ่งแยกกลุ่มข้อมูลได้ดีที่สุด และหาระยะห่างจากระยะขอบ (margin) ให้มากที่สุดเพื่อรองรับให้จุดข้อมูลในอนาคตสามารถจำแนกได้ ซึ่งการจำแนกประเภทนี้รวดเร็วและเชื่อถือได้ ซึ่งทำงานได้ดีมากกับข้อมูลจำนวนจำกัด



รูปที่ 2.11 กราฟแสดงหลักการของซัพพอร์ตเวกเตอร์แมชชีน

โดยกำหนดให้สมการเส้นแบ่งไฮเปอร์เพลน คือ

$$w^T x + b$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการหาเส้นแบ่งที่ดีที่สุดต้องเพิ่มเส้นขอบให้กับเส้นแบ่งทั้งสองข้างให้สัมผัสกับค่าข้อมูลทั้งสองใกล้ที่สุด โดยจุดสัมผัสข้อมูลที่ใกล้ที่สุดจากเส้นขอบนี้เราเรียกว่า ซัพพอร์ตเวกเตอร์ (support vector) ซึ่งจะเป็นการกำหนดช่วงสำหรับขอบรับความผิดพลาด โดยจะมีการเพิ่มตัวแปร ξ ดังสมการ

$$w^T x + b \geq y - \xi_i; y = 1$$

$$w^T x + b \leq y + \xi_i; y = -1$$

การกำหนดค่า $\xi > 0$ มีเป้าหมายทำให้เพิ่มระยะเส้นขอบให้มากที่สุดและลดข้อผิดพลาดในการทำงานให้ต่ำที่สุด ดังสมการ

$$\text{Minimize } \frac{\|w\|^2}{2} + c \sum_{i=1}^N \xi_i$$

โดยในการทำงานของซัพพอร์ตเวกเตอร์แมชชีนจะเรียกใช้เคอร์เนลฟังก์ชัน (kernel function) เพื่อแปลงข้อมูลให้มีมิติที่มากขึ้นเพื่อให้อยู่ในรูปแบบที่สามารถใช้สมการเส้นตรงแบ่งแยกได้ง่ายขึ้น โดยเคอร์เนลฟังก์ชันที่นิยมใช้อยู่ 3 ชนิดด้วยกันคือ

โพลีโนเมียล (polynomial)

$$k(x_i, x_j) = (x_i^T x_j + r)^\gamma; \gamma > 0$$

เรเดียลเบสิสฟังก์ชัน (Radial Basis Function - RBF)

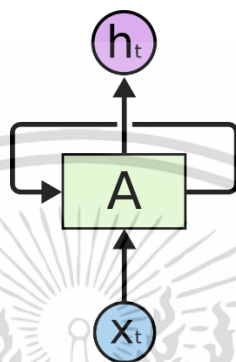
$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2); \gamma > 0$$

ซิกมอยด์ (sigmoid)

$$k(x_i, x_j) = \tanh(-\gamma x_i^T x_j - r)$$

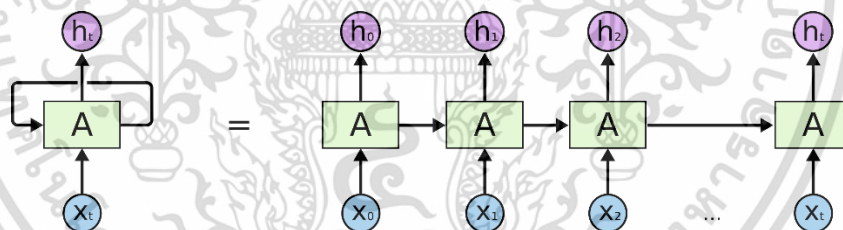
2.7.4 RNN

RNN หรือ Recurrent Neural Networks เป็น Neural Network ชนิดหนึ่งที่ใช้ในการแก้ไขปัญหาค้าง ๆ มากมาย เช่น การรู้จำเสียง การสร้างแบบจำลองภาษาต่าง ๆ หรือการแปลงข้อความ เป็นต้น โดยเครือข่ายประสาทนี้มีการใช้สถานะก่อนหน้าเพื่อสามารถทำนายเหตุการณ์ที่จะเกิดขึ้นต่อไป



รูปที่ 2.12 แสดง Recurrent Neural Networks แบบการวนลูป

จากภาพที่ 2.12 สามารถคิดได้ว่า RNN เป็นการคัดลอกเน็ตเวิร์คซ้ำ ๆ โดยข้อความแต่ละข้อความจะถูกส่งผ่านไปยังเน็ตเวิร์คที่ทำการคัดลอกต่อ ๆ ไปดังรูปที่ 2.13

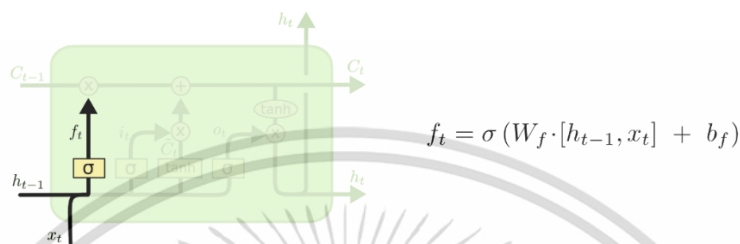


รูปที่ 2.13 แสดง Recurrent Neural Networks แบบแยกในหนึ่งโหนด

จากภาพที่ 2.13 จะพบว่า เป็นลักษณะคล้ายห่วงโซ่ที่แสดงให้เห็นว่าเครือข่ายประสาทที่เกิดขึ้นซ้ำนั้นมีความสัมพันธ์กับลำดับ แต่การทำงานของ RNN สามารถก่อให้เกิดปัญหา Long-Term Dependencies หรือเมื่อช่องว่าง (ช่องว่างระหว่างข้อมูลที่เกี่ยวข้องและจุดที่กำลังดำเนินการ) เพิ่มขึ้น RNN จะไม่สามารถเรียนรู้การเชื่อมต่อข้อมูลกับจุดก่อนๆ ได้ ซึ่งสามารถแก้ไขปัญหาดังกล่าวได้ด้วย Long Short-Term Memory networks (LSTMs)

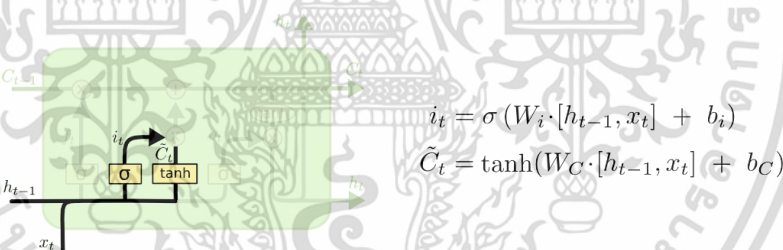
2.7.5 LSTMs

LSTMs หรือ Long Short-Term Memory networks เป็นเครือข่ายประสาทที่แก้ไขปัญหา Long-Term Dependencies ของ RNN โดย LSTMs สามารถทำการเลือกได้ว่าข้อมูลไหนที่ควรจำ ไม่ควรจำ หรือไม่ว่าจะเป็นการอ่าน เขียน อัปเดตข้อมูล โดยใช้เซลล์ที่ใช้ในการจดจำ (memory cell states) โดยเปรียบเสมือนประตูที่คอยควบคุมข้อมูลแต่ละโหนด



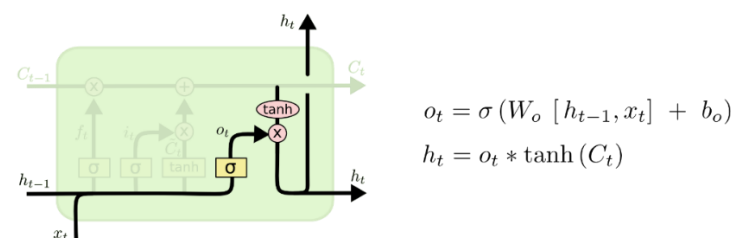
รูปที่ 2.14 LSTM Forget Gate

จากรูปด้านบนมีฟังก์ชัน sigmoid ซึ่งผลลัพธ์ที่ได้จะมีค่าอยู่ระหว่าง 0 และ 1 โดยถ้าผลลัพธ์มีค่าเป็น 0 หมายถึง ไม่มีข้อมูลใดได้เข้าไป และผลลัพธ์เป็น 1 หมายถึง ปล่อยให้ข้อมูลทุกข้อมูลเข้าไป



รูปที่ 2.15 LSTM Update Cell State และ Input Gate

รูปที่ 2.15 เป็นประตูที่ใช้เพื่อรับข้อมูลที่เข้ามาใหม่ เพื่อให้บันทึกลงไป และเป็นการอัปเดต state จากเซลล์เก่า หรืออัปเดตค่าจากเซลล์ C_{t-1} ลงไปในเซลล์ C_t โดยฟังก์ชัน tanh เป็นการแปลงข้อมูลให้มีความใกล้เคียงกับข้อมูลขาออก



รูปที่ 2.16 LSTM Output Gate

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และสุดท้ายรูปที่ 2.16 เป็นขาออกของเซลล์ โดยจะทำการตัดสินใจจาก state ของเซลล์ นั้น ๆ เริ่มต้นโดยเข้าฟังก์ชัน sigmoid จะเป็นการตัดสินใจว่าส่วนใดของ cell state ที่จะทำการส่งออกจากนั้นจะนำค่าจาก cell state เข้าฟังก์ชัน tanh และคูณด้วยผลลัพธ์ของฟังก์ชัน sigmoid สุดท้ายผลลัพธ์ที่ได้จะถูกแบ่งออกเป็นผลลัพธ์ของโหนดนั้น และข้อมูลที่จะส่งเข้าโหนดอื่น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

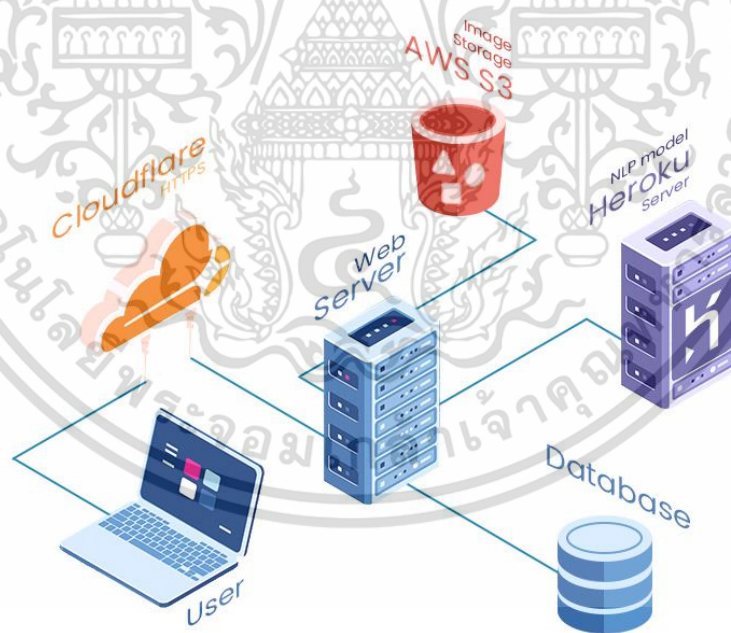
การออกแบบและพัฒนา

การออกแบบและพัฒนาเว็บแอปพลิเคชัน ประกอบด้วย ส่วนการออกแบบ ได้แก่ ภาพรวมของระบบ แผนภาพยูสเคส แผนภาพแสดงความสัมพันธ์ของข้อมูล การออกแบบโมเดลสำหรับการประมวลผลทางภาษา และส่วนต่อประสานกับผู้ใช้ ในส่วนการการพัฒนา ได้แก่ การเก็บข้อมูล การเตรียมข้อมูลก่อนการประมวลผล และการฝึกโมเดล

3.1 การออกแบบ

3.1.1 ภาพรวมของระบบ

การพัฒนาเว็บแอปพลิเคชันภาพยนต์ มีความต้องการทั้งหมด 4 ส่วนคือ การแนะนำภาพยนตร์ การรีวิวภาพยนตร์ บทความภาพยนตร์ และกระดานสนทนา ในส่วนของการแนะนำภาพยนตร์จะนำทวิตของภาพยนตร์เรื่องนั้น ๆ ไปผ่านโมเดลการเรียนรู้ของเครื่องเพื่อตีความและแสดงบนหน้าเว็บไซต์ ซึ่งภาพรวมของระบบจะแบ่งออกเป็นส่วนดังนี้



รูปที่ 3.1 ภาพรวมของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.1.1 ส่วนของเว็บแอปพลิเคชัน

ส่วนของเว็บแอปพลิเคชัน เป็นส่วนต่อประสานกับผู้ใช้ทั้งสำหรับผู้ใช้งานและสำหรับผู้ดูแลระบบ ผู้ใช้งานสามารถดูข้อมูลและความคิดเห็นของหนังเรื่องต่าง ๆ โดยยังสามารถแสดงความคิดเห็น เขียนรีวิว บทความ สนทนาได้ ผู้ดูแลระบบสามารถจัดการเพิ่ม ลบ และแก้ไขรายการหนังหรือรายการอื่น ๆ ที่เกี่ยวข้องของระบบได้ ในส่วนของเว็บแอปพลิเคชันนี้จะใช้เฟรมเวิร์ก Next.js ที่ช่วยในการเรนเดอร์ฝั่งเซิร์ฟเวอร์ซึ่งเป็นประโยชน์สำหรับการทำ SEO

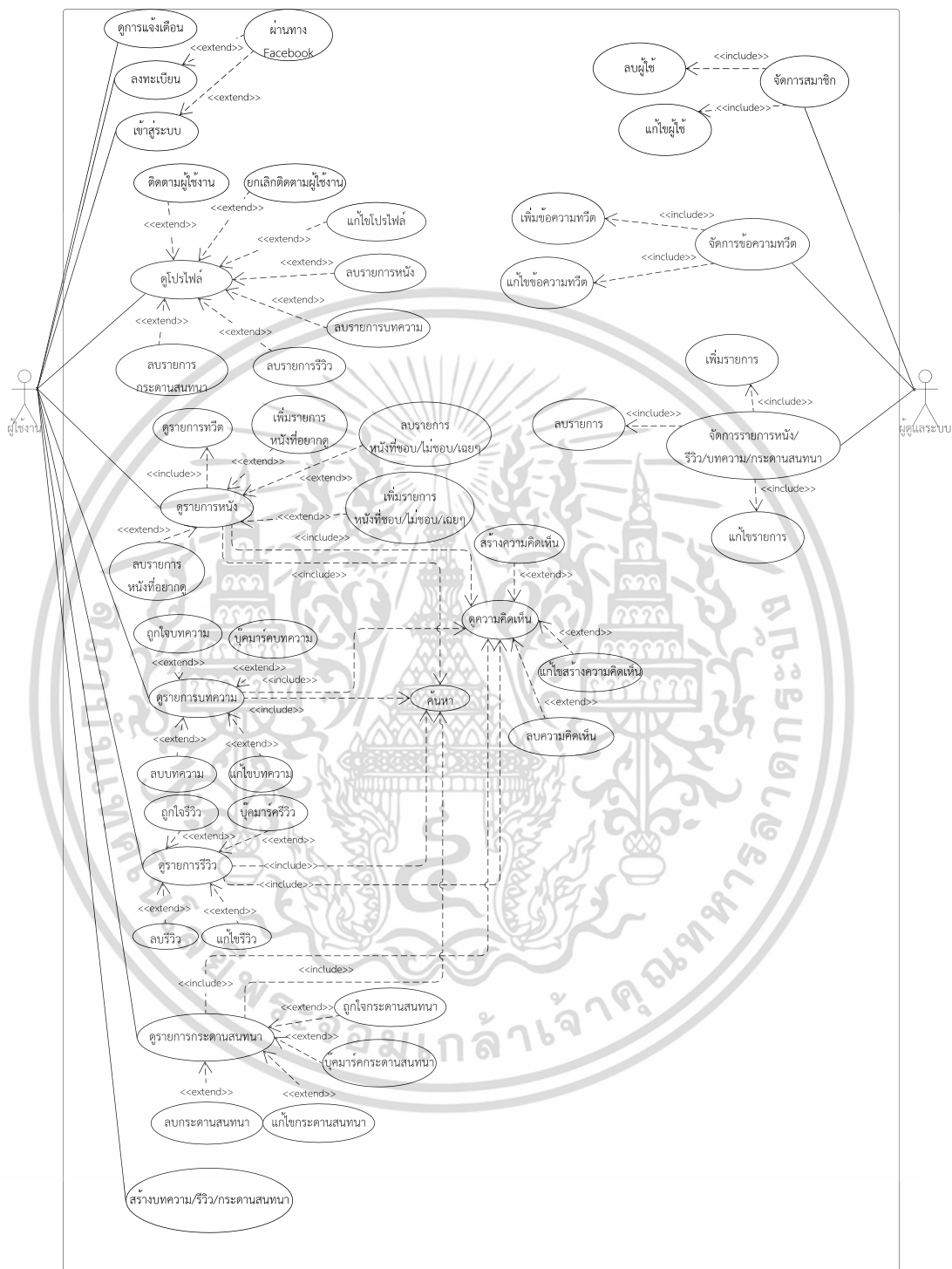
3.1.1.2 ส่วนของเซิร์ฟเวอร์และฐานข้อมูล

ส่วนของเซิร์ฟเวอร์ เป็นส่วนกลางประสานงานกับส่วนเว็บแอปพลิเคชันกับฐานข้อมูล โดยเซิร์ฟเวอร์จะส่งข้อมูลต่าง ๆ จากฐานข้อมูลผ่านการเรียก API จากทางเว็บไซต์ และยังเป็นส่วนที่ประมวลผลในส่วนของความคิดเห็นจากทวิตเตอร์ ในส่วนเซิร์ฟเวอร์ใช้ Node.js ในการพัฒนา และใช้ PostgreSQL เป็นระบบจัดการฐานข้อมูล

3.1.1.3 ส่วนของโมเดล

ส่วนของโมเดลที่จะใช้ทำนายแ่งมของข้อความทวิต ซึ่งถูกนำขึ้นเซิร์ฟเวอร์ด้วย Heroku และใช้ Flask ในการจัดการ API โดยเว็บแอปพลิเคชันสำหรับผู้ดูแลระบบจะทำการยิง API ไปที่เซิร์ฟเวอร์ของโมเดลเพื่อทำนายแ่งมของข้อความทวิต

3.1.2 แผนภาพยูสเคส (Use Case Diagram)

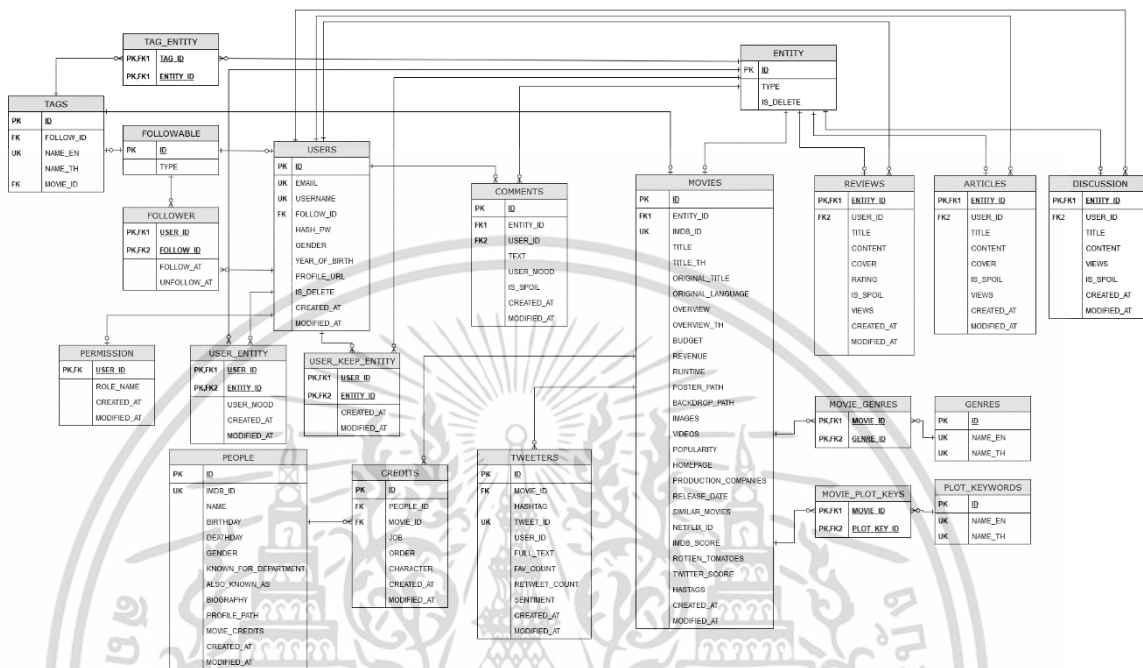


รูปที่ 3.2 แผนภาพยูสเคส (Use Case Diagram)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.3 แผนภาพแสดงความสัมพันธ์ของข้อมูล (Logical Data Model)

การออกแบบฐานข้อมูลของระบบทั้งหมด สามารถแสดงออกมาในรูปแบบภาพความสัมพันธ์ของข้อมูลดังนี้



รูปที่ 3.3 แผนภาพแสดงความสัมพันธ์ของข้อมูล

การออกแบบฐานข้อมูลประกอบด้วยตารางจำนวน 21 ตาราง มีรายละเอียดดังนี้

3.1.3.1 ตารางผู้ใช้งาน (User)

ตารางผู้ใช้งาน เป็นตารางที่เก็บข้อมูลเกี่ยวกับผู้ใช้งาน ประกอบไปด้วย ไอดีผู้ใช้งาน ชื่อผู้ใช้งานอีเมลผู้ใช้งาน ไอดีของการติดตาม รหัสผ่านที่เข้ารหัส เพศ ปีเกิดผู้ใช้งาน URL ของรูปภาพโปรไฟล์ สถานะการลบ เวลาที่สร้างผู้ใช้งาน และเวลาที่มีการแก้ไขล่าสุด

3.1.3.2 ตารางการอนุญาต (Permission)

ตารางการอนุญาต เป็นตารางที่เก็บข้อมูลที่เกี่ยวข้องกับการอนุญาตสิทธิ์ของผู้ใช้งาน ประกอบไปด้วย ไอดีผู้ใช้งาน บทบาทของผู้ใช้งาน เวลาที่สร้าง และเวลาที่มีการแก้ไขล่าสุด

3.1.3.3 ตารางเอนทิตี (Entity)

ตารางเอนทิตี เป็นตารางที่สร้างไอดีเฉพาะของสิ่งที่ต้องการเก็บ โดยในที่นี้คือ หนังสือ รีวิว บทความ และกระดานสนทนา ให้ไม่ซ้ำเพื่อประโยชน์ในการเก็บข้อมูลความคิดเห็น และการเก็บเอนทิตี ประกอบไปด้วย ไอดีของเอนทิตี ชนิดของเอนทิตี และสถานการณลบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.3.4 ตารางความชอบผู้ใช้ (User_Entity)

ตารางความชอบผู้ใช้ เป็นตารางที่เก็บความรู้สึกของผู้ใช้ที่มีต่อเอนทิตีต่าง ๆ ประกอบไปด้วย ไอดีของผู้ใช้งาน ไอดีของเอนทิตี ความรู้สึกของผู้ใช้ เวลาที่สร้างบันทึก และเวลาที่มีการแก้ไขล่าสุด

3.1.3.5 ตารางการเก็บเอนทิตี (User_Keep_Entity)

ตารางการเก็บเอนทิตี เป็นตารางที่เก็บเอนทิตีของผู้ใช้ ประกอบไปด้วย ไอดีของผู้ใช้งาน ไอดีของเอนทิตี เวลาที่สร้างบันทึก และเวลาที่มีการแก้ไขล่าสุด

3.1.3.6 ตารางทวิตเตอร์ (Tweets)

ตารางทวิตเตอร์ เป็นตารางที่เก็บข้อความทวิตเตอร์ที่เกี่ยวกับหนัง ประกอบไปด้วย ไอดีของผู้ใช้งาน ไอดีของหนัง แฮชแท็กที่ดึงข้อมูล ไอดีของข้อความทวิตเตอร์ ไอดีของผู้ใช้ทวิตเตอร์ ข้อความจากทวิตเตอร์ จำนวนการกดชอบ จำนวนการรีทวีต การตีความหมาย เวลาที่สร้างบันทึก และเวลาที่มีการแก้ไขล่าสุด

3.1.3.7 ตารางหนัง (Movies)

ตารางหนัง เป็นตารางที่เก็บข้อมูลของหนัง ประกอบไปด้วย ไอดีของหนัง ไอดีของเอนทิตี ไอดีหนังของ IMDB ชื่อหนัง ชื่อหนังภาษาไทย ชื่อหนังดั้งเดิม ภาษาดั้งเดิม เรื่องย่อหนัง เรื่องย่อหนังภาษาไทย งบประมาณการสร้าง รายได้ ความยาวหนัง (นาที) URL ของโปสเตอร์หนัง URL ของรูปภาพพื้นหลัง URL ของรูปภาพ ข้อมูลวิดีโอ ความนิยม เว็บไซต์หลัก บริษัทที่ผลิตวันที่เข้าฉาย หนังที่คล้ายคลึง ไอดีของหนังจาก Netflix ข้อมูลคะแนนจาก IMDB ข้อมูลคะแนนหนังจากเว็บมะเขือเทศน่า คะแนนจากทวิตเตอร์ แฮชแท็กของหนัง เวลาที่สร้าง และเวลาที่มีการแก้ไขล่าสุด

3.1.3.8 ตารางประเภทหนัง (Genres)

ตารางประเภทหนัง เป็นตารางที่เก็บประเภทหนัง ประกอบไปด้วย ไอดีของประเภทหนัง ชื่อประเภทหนังภาษาอังกฤษ และชื่อประเภทหนังภาษาไทย

3.1.3.9 ตารางประเภทของหนัง (Movie_Genres)

ตารางประเภทของหนัง เป็นตารางที่เก็บประเภทของหนังเรื่องต่าง ๆ ประกอบไปด้วย ไอดีของหนัง และไอดีของประเภทหนัง

3.1.3.10 ตารางคำสำคัญ (Plot_Keywords)

ตารางคำสำคัญ เป็นตารางที่เก็บคำสำคัญ ประกอบไปด้วย ไอดีของคำสำคัญ ชื่อคำสำคัญภาษาอังกฤษ และชื่อคำสำคัญภาษาไทย

3.1.3.11 ตารางคำสำคัญของหนัง (Movie_Plot_Keys)

ตารางคำสำคัญของหนัง เป็นตารางที่เก็บคำสำคัญของหนังต่าง ๆ ประกอบไปด้วย ไอดีของหนัง และไอดีของคำสำคัญ

3.1.3.12 ตารางทีมงาน (People)

ตารางทีมงาน เป็นตารางที่เก็บข้อมูลต่าง ๆ ของทีมงาน ประกอบไปด้วย ไอดีของคน ไอดีของ IMDB ชื่อ วันเกิด วันเสียชีวิต เพศ บทบาทที่เป็นที่รู้จัก ชื่อในภาษาอื่น ชีวประวัติ URL รูปภาพโปรไฟล์ หน้าที่ได้รับเครดิต เวลาที่สร้าง และเวลาที่มีการแก้ไขล่าสุด

3.1.3.13 ตารางเครดิต (Credits)

ตารางเครดิต เป็นตารางที่เก็บเครดิตของทีมงานว่าเกี่ยวข้องกับหนังในหน้าที่อะไร ประกอบไปด้วย ไอดีของเครดิต ไอดีของคน ไอดีของหนัง ลำดับของเครดิต หน้าที่ บทบาทที่แสดง เวลาที่สร้าง และเวลาที่มีการแก้ไขล่าสุด

3.1.3.14 ตารางรีวิว (Reviews)

ตารางรีวิว เป็นตารางที่เก็บรีวิวต่าง ๆ ของหนัง ประกอบไปด้วย ไอดีของเอนทิตี ไอดีของผู้ใช้งาน หัวข้อ เนื้อหา การสปอยล์เนื้อหาสำคัญ คะแนนรีวิว ยอดผู้เข้าชม เวลาที่สร้างรีวิว และเวลาที่มีการแก้ไขล่าสุด

3.1.3.15 ตารางบทความ (Articles)

ตารางบทความ เป็นตารางที่เก็บบทความต่าง ๆ ประกอบไปด้วย ไอดีของเอนทิตี ไอดีของผู้ใช้งาน หัวข้อ เนื้อหา การสปอยล์เนื้อหาสำคัญ ยอดผู้เข้าชม เวลาที่สร้างบทความ และเวลาที่มีการแก้ไขล่าสุด

3.1.3.16 ตารางการสนทนา (Discussions)

ตารางการสนทนา เป็นตารางที่เก็บบทสนทนาต่าง ๆ ประกอบไปด้วย ไอดีของเอนทิตี ไอดีของผู้ใช้งาน เนื้อหา การสปอยล์เนื้อหาสำคัญ ยอดผู้เข้าชม เวลาที่สร้างบทสนทนา และเวลาที่แก้ไขล่าสุด

3.1.3.17 ตารางความคิดเห็น (Comments)

ตารางความคิดเห็น เป็นตารางที่เก็บความคิดเห็นต่าง ๆ ทั้งหมด ประกอบไปด้วย ไอดีของความคิดเห็น ไอดีของเอนทิตี ไอดีของผู้ใช้ ข้อความ ความรู้สึกของผู้ใช้ การสปอยล์เนื้อหาสำคัญ เวลาที่สร้างความคิดเห็น และเวลาที่มีการแก้ไขล่าสุด

3.1.3.18 ตารางแท็ก (Tags)

ตารางแท็ก เป็นตารางที่เก็บชื่อแท็ก ประกอบไปด้วย ไอดีของแท็ก ไอดีของการติดตาม แท็กภาษาอังกฤษ ชื่อแท็กภาษาไทย และไอดีของหนัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.3.19 ตารางการแท็ก (Tag_Entity)

ตารางการแท็ก เป็นตารางที่เก็บการแท็กต่าง ๆ ของบทความ หรือบทสนทนา ประกอบไปด้วย ไอดีของแท็ก และไอดีของเอนทิตี

3.1.3.20 ตารางการติดตาม (Followable)

ตารางการติดตาม เป็นตารางที่เก็บไอดีการติดตามต่าง ๆ ประกอบไปด้วย ไอดีของการติดตาม และชนิดของสิ่งที่ติดตาม

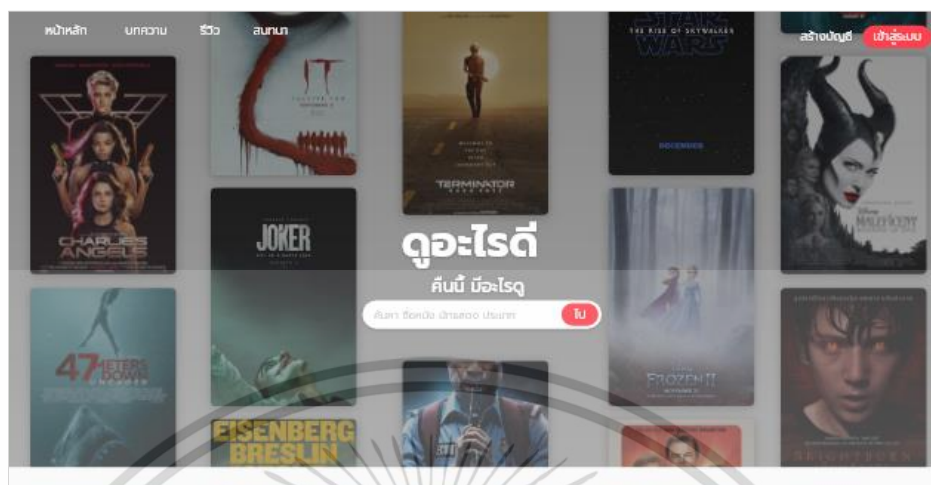
3.1.3.21 ตารางผู้ติดตาม (Follower)

ตารางผู้ติดตาม เป็นตารางที่เก็บการติดตามของผู้ใช้ ประกอบไปด้วย ไอดีของผู้ใช้งาน ไอดีของการติดตาม เวลาที่เริ่มการติดตาม และเวลายกเลิกการติดตาม

3.1.4 การออกแบบโมเดลสำหรับการประมวลผลทางภาษา

เมื่อได้ข้อความจากทวิตเตอร์แล้ว ข้อความจะต้องผ่านกระบวนการทำความสะอาดข้อความ และตัดคำก่อน หลังจากนั้นจากนั้นจึงนำชุดคำดังกล่าวกับชุดคำตอบที่แยกประเภทไว้แล้วเข้าฝึกโดยใช้การเรียนรู้เครื่องในการสร้างโมเดลวิเคราะห์อารมณ์จากข้อความ

3.1.5 การออกแบบส่วนต่อประสานกับผู้ใช้

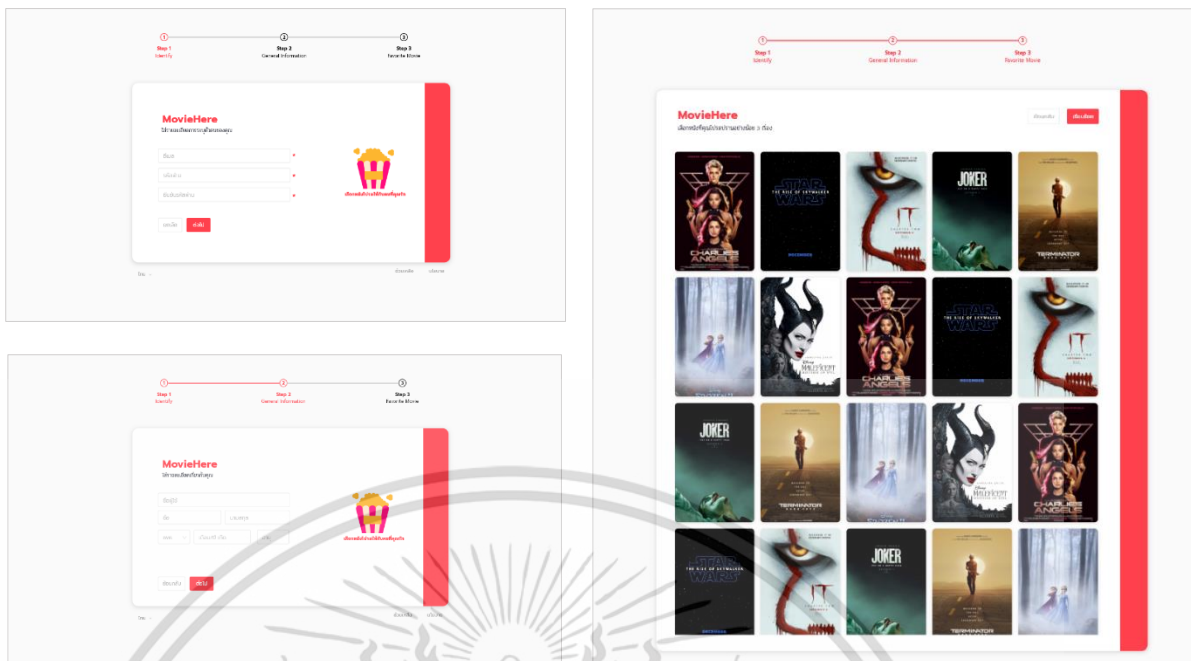


รูปที่ 3.4 หน้า Landing page

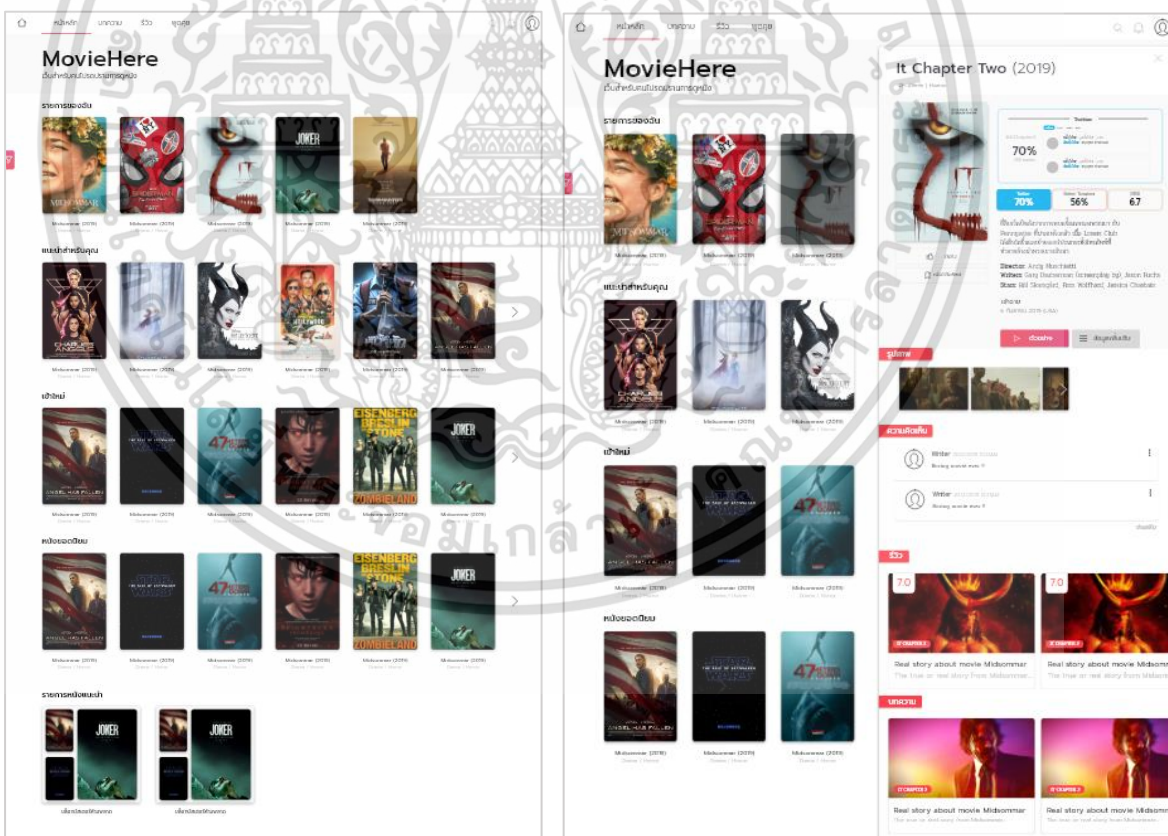


รูปที่ 3.5 ป๊อปอัพการเข้าสู่ระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

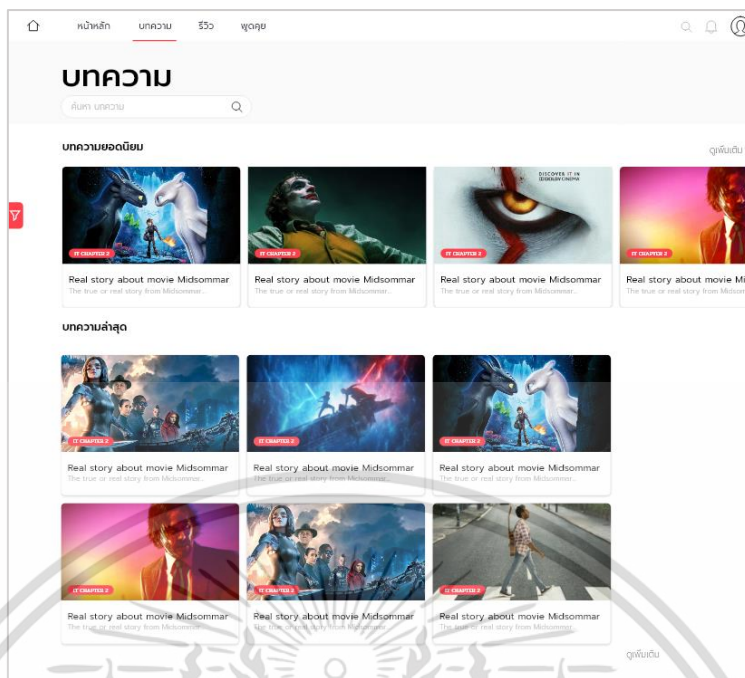


รูปที่ 3.6 หน้าลงทะเบียน

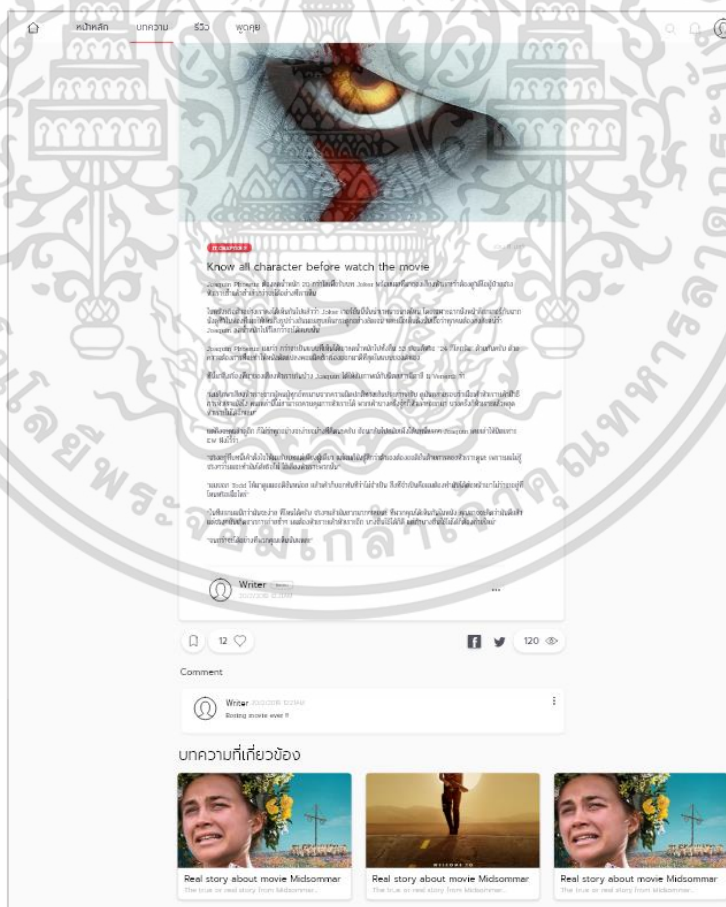


รูปที่ 3.7 หน้าหลัก ซึ่งแสดงรายการหนัง และเมื่อดูที่หนังเรื่องนั้น ๆ

เอกสารนี้เป็นเอกสารที่สวอนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

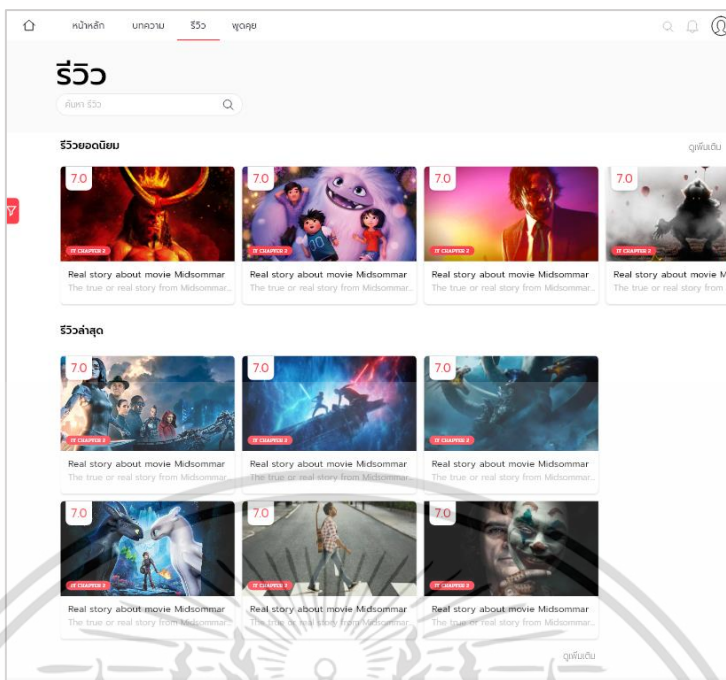


รูปที่ 3.8 หน้าแสดงรายการบทความ

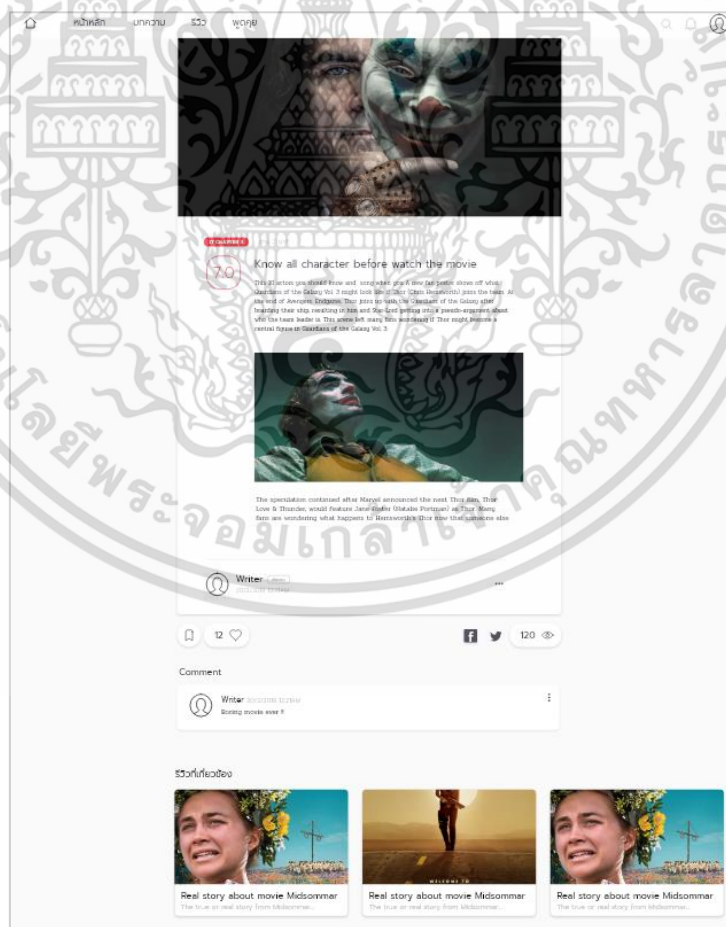


รูปที่ 3.9 หน้าแสดงบทความ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

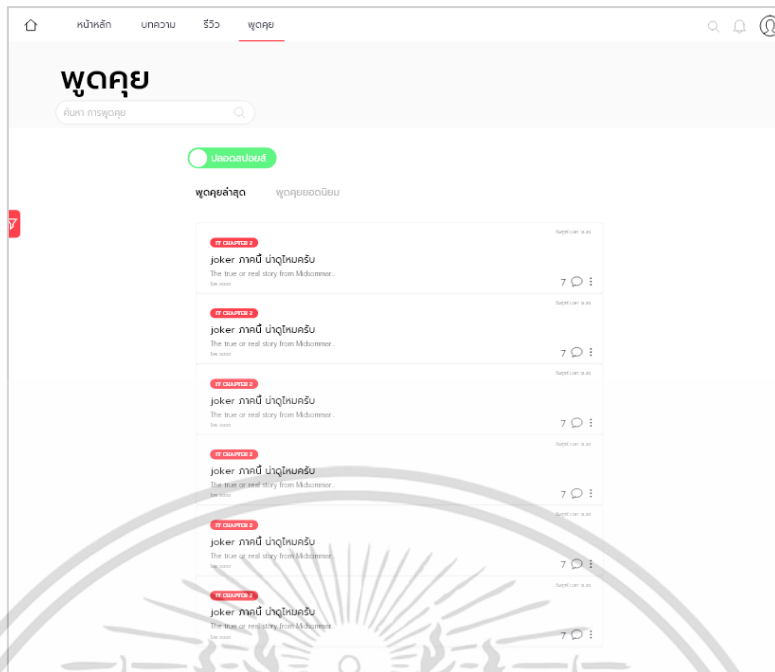


รูปที่ 3.10 หน้าแสดงรายการรีวิว

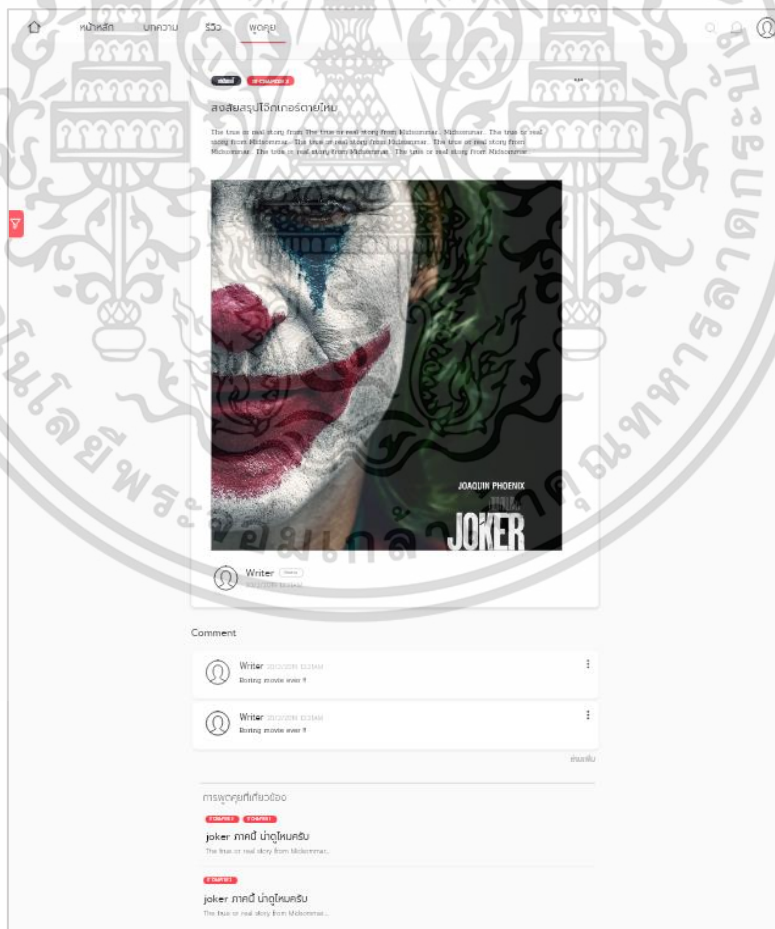


รูปที่ 3.11 หน้าแสดงรีวิว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.12 หน้าแสดงรายการหัวข้อพูดคุย



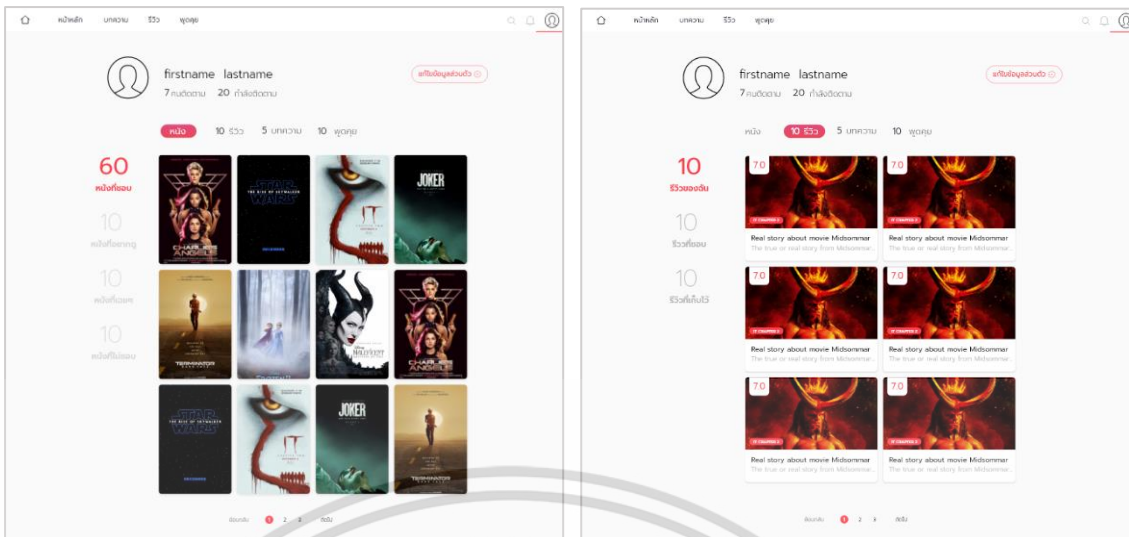
รูปที่ 3.13 หน้าแสดงหัวข้อการพูดคุย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

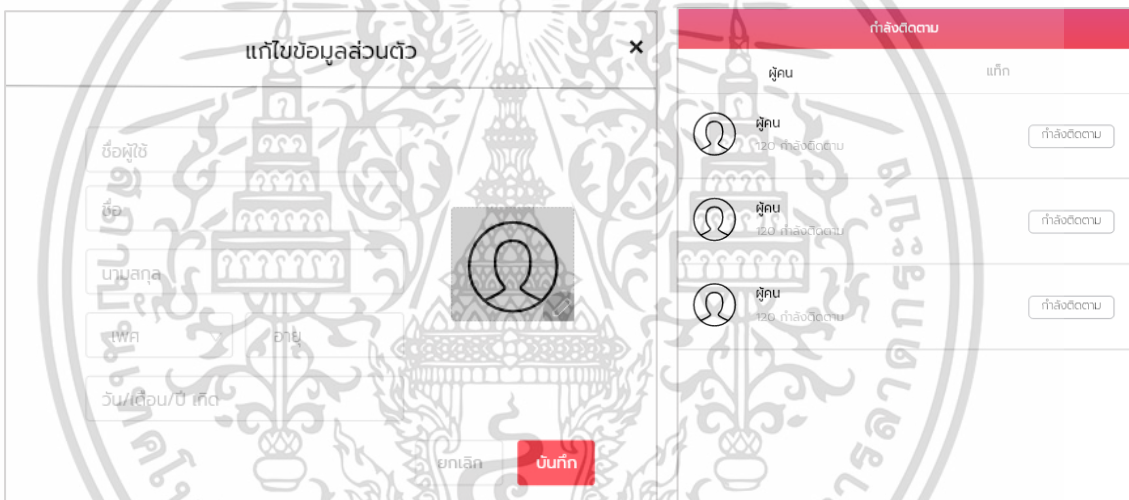


รูปที่ 3.14 หน้าแสดงหนัง (แบบเต็ม)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

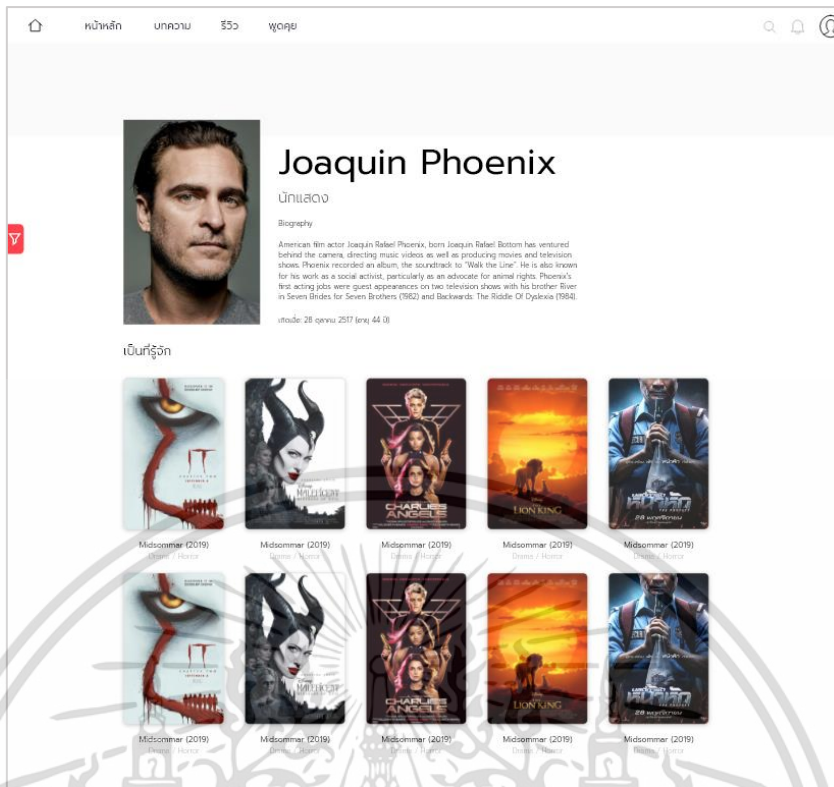


รูปที่ 3.15 หน้าโปรไฟล์ของผู้ใช้งาน โดยสามารถดูหัวข้อที่ทำการเก็บบันทึกไว้ได้

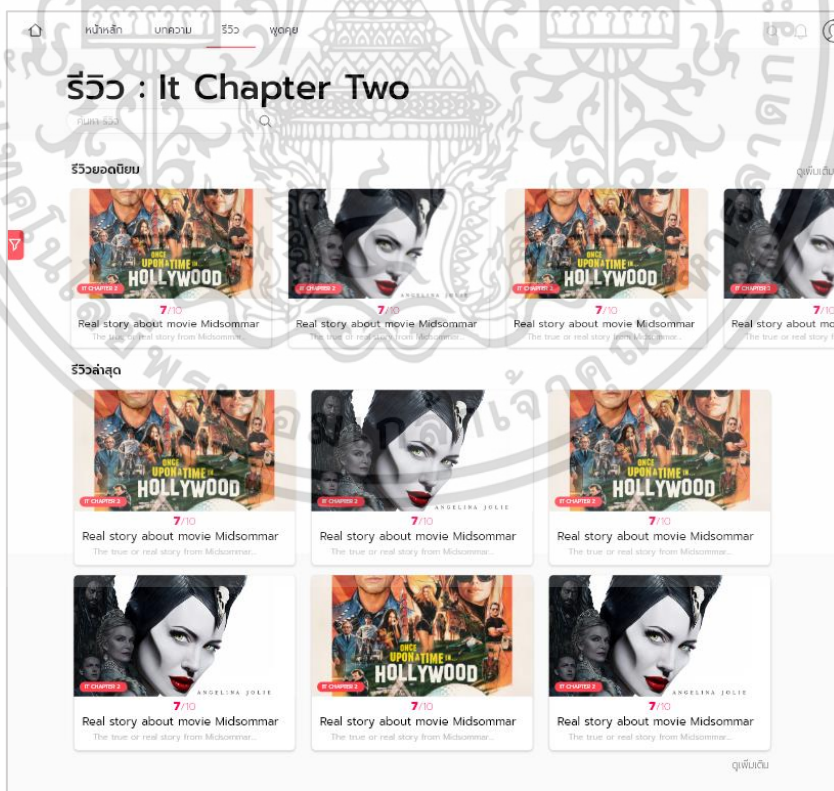


รูปที่ 3.16 ป๊อปอัพสำหรับการแก้ไขโปรไฟล์ และดูการติดตามของผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

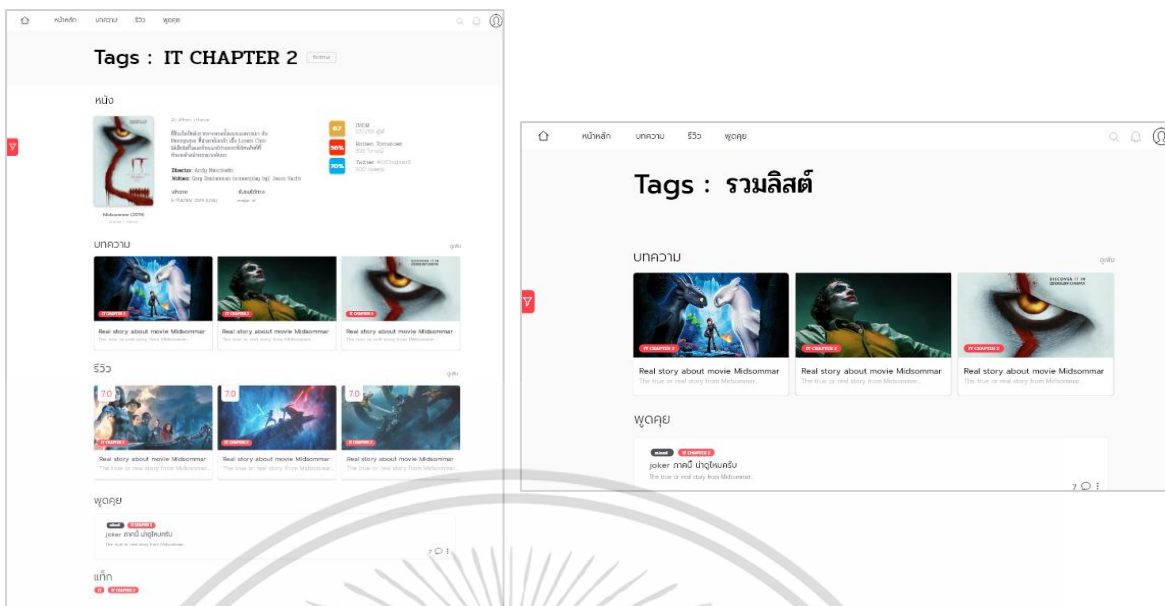


รูปที่ 3.17 เมื่อกดที่นักแสดงหรือบุคคลที่เกี่ยวข้อง จะแสดงหน้าที่มีรายละเอียดของบุคคลนั้น

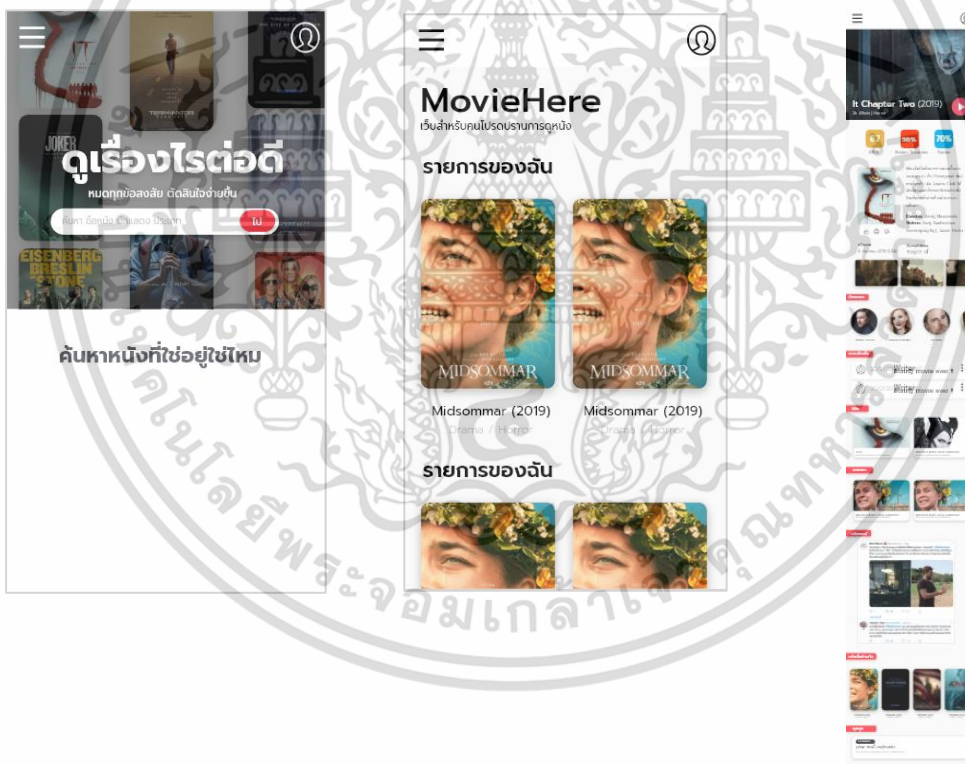


รูปที่ 3.18 เมื่อทำการค้นหาจากหน้ารีวิว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.19 แสดงหน้าการค้นหา เมื่อทำการค้นหาจากแท็ก



รูปที่ 3.20 การออกแบบสำหรับมือถือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การพัฒนา

3.2.1 การเก็บข้อมูล

3.2.1.1 การเก็บข้อมูลทวิตจากทวิตเตอร์

ข้อความที่ใช้สำหรับการเทรน โมเดล เป็นข้อความที่ส่วนมากมาจากทวิตเตอร์ ซึ่ง จะทำการดึงความคิดเห็นของภาพยนตร์เรื่องต่าง ๆ จากแฮชแท็ก โดยใช้บริการของทวิตเตอร์ สำหรับผู้พัฒนา และไลบรารีของไพทอน เช่น Tweepy และ TwitterAPI ที่ช่วยให้การเข้าถึงง่ายมา กยิ่งขึ้น

โปรแกรม 3.1 การดึงทวิตจากทวิตเตอร์ด้วย Tweepy

```
searched_tweets = [status._json for status in tweepy.Cursor(api.search, q= '# โจรปล้นโจร'+
    " - filter:retweets", count=100, tweet_mode='extended', lang='th',
    result_type='mixed') .items(1000)]
```

นอกจากยังสามารถดึงทวิตจากทวิตเตอร์ด้วย TwitterAPI ซึ่งสามารถกำหนดวัน เริ่มต้น และวันสิ้นสุดที่ต้องการดึงได้

โปรแกรม 3.2 การดึงทวิตจากทวิตเตอร์ด้วย TwitterAPI

```
r = api.request('tweets/search/%s/:%s' % ('fullarchive', 'development'),
    {'query': '# โจรปล้นโจร' + ' lang:th', 'fromDate': '201910290000' ,
    'toDate': '201911070000' })
```

3.2.1.2 การเก็บข้อมูลภาพยนตร์

ฐานข้อมูลภาพยนตร์สำหรับเว็บไซต์ รวบรวมได้จากการดึงข้อมูลจาก เว็บไซต์ The Movie Database (TMDb) ซึ่งเป็นชุมชนออนไลน์ที่ช่วยกันสร้างฐานข้อมูลของภาพยนตร์และ รายการโทรทัศน์ตั้งแต่ปี 2008 และยังมี การเก็บข้อมูลที่เป็นภาษาไทยอีกด้วย โดยมีไลบรารีที่ช่วย ในการดึงข้อมูลเช่น tmdbsimple ของไพทอน ซึ่งช่วยให้การดึงข้อมูลสะดวกขึ้น โดยสามารถดึง ข้อมูลได้หลายส่วน เช่น การค้นหาภาพยนตร์ ข้อมูลเบื้องต้นของ ประวัติของคนที่ทำงานเกี่ยวข้องกับ ภาพยนตร์

โปรแกรม 3.3 การค้นหาภาพยนตร์ในช่วงเวลาที่ต้องการด้วย tmdbsimple

```
def get_discover_page (i, release_lte = "", release_gte = "") :
    kwargs = { 'page' : i , 'release_date.lte' : release_lte , 'release_date.gte' : release_gte ,
               'language' : 'th-TH' , 'sort_by' : 'release_date.desc' }
    return discover.movie(**kwargs)
```

โปรแกรม 3.4 การดึงข้อมูลเบื้องต้นภาพยนตร์ด้วย tmdbsimple

```
movie = tmdb.Movies(id)
response = movie.info(append_to_response='credits, images, keywords, similar, videos')
```

3.2.2 การเตรียมข้อมูลก่อนการประมวลผล

ในการประมวลผลทางภาษาจะต้องมีการเตรียมข้อมูลให้ดีขึ้นนำไปประมวลจริง เช่น การแก้คำผิด การตัดคำ เพราะจะช่วยลดสิ่งรบกวนที่อยู่ในข้อมูล และช่วยทำให้การประมวลผลมีประสิทธิภาพดีขึ้น

3.2.2.1 การทำความสะอาดข้อความ

ก่อนเข้าสู่กระบวนการตัดคำจะต้องจัดการกับยูอาร์แอล แฮชแท็ก และการลากตัวอักษรต่างๆ เพื่อให้การตัดคำมีความแม่นยำมากที่สุด

โปรแกรม 3.5 การทำความสะอาดข้อความ

```
print ( clean_text (
    ""
    ฝ่ายุทธการ.... ดับแผนอหังการ US มันส์มากกกก 🤪👍 #MOVIE_ABC
    https://t.co/D2M2rYnoBY https://t.co/dNhp8QV38S
    ""
))

# ฝ่ายุทธการ... ดับแผนอหังการ US มันส์มาก 🤪👍
```

3.2.2.2 การตัดคำ

ในการตัดคำโดยใช้ไลบรารี PyThaiNLP มีอัลกอริทึมให้เลือกใช้หลายแบบ โดยในที่นี้จะเลือกทดสอบ 3 แบบได้แก่ Newmm ตัดคำโดยใช้อัลกอริทึมการจับคู่สูงสุด (maximal matching algorithm), Deepcut ใช้โครงข่ายประสาทเทียม (Deep Neural Network) และ AttaCut ตัดคำโดยใช้ Convolutional Neural Network (CNN)

โปรแกรม 3.6 การตัดคำ

```
text = clean_text("#MOVIE_ABC ลุ้นระทึก โคตรมันส์ ไม่ต้องคิสเยอะ อีหยังวะ เป็นหนังที่บันเทิงมาก 8.5/10 #movietwit https://t.co/j8sWQK35tr ")
text1 = word_tokenize(text, engine='newmm')
# ['ลุ้น', 'ระทึก', 'โคตร', 'มันส์', 'ไม่', 'ต้อง', 'คิส', 'เยอะ', 'อี', 'หยัง', 'วะ', 'เป็น', 'หนัง', 'ที่', 'บันเทิง', 'มาก', '8.5', '/', '10']
text2 = word_tokenize(text, engine='deepcut')
# ['ลุ้น', 'ระทึก', 'โคตร', 'มันส์', 'ไม่', 'ต้อง', 'คิสเยอะ', 'อีหยัง', 'วะ', 'เป็น', 'หนัง', 'ที่', 'บันเทิง', 'มาก', '8', '/', '5', '/', '10']
text3 = word_tokenize(text, engine='attacut')
# ['ลุ้น', 'ระทึก', 'โคตรมันส์', 'ไม่', 'ต้อง', 'คิสเยอะ', 'เป็น', 'หนัง', 'อีหยัง', 'วะ', 'ที่', 'บันเทิง', 'มาก', '8', '/', '5', '/', '10']
```

โดยเมื่อใช้กับข้อความที่ขนาดยาว จะใช้เวลาในการประมวลผล 2.99 มิลลิวินาที 161 มิลลิวินาที และ 60.8 มิลลิวินาที ตามลำดับ ซึ่งการตัดคำของ newmm เป็นที่ค่อนข้างน่าพอใจ นอกจากจะใช้น้อยที่สุด ยังสามารถเพิ่มพจนานุกรมได้ จึงเลือกใช้อัลกอริทึมนี้ในการตัดคำ

3.2.2.3 การแก้คำผิด

จากการตัดคำเบื้องต้นจะพบว่าจะพบปัญหาในการตัดคำทับศัพท์ ซึ่งคน คำที่มักเขียนผิด ทำให้การตัดผิดพลาด จึงต้องทำการเพิ่มพจนานุกรมในการตัดคำ และแก้ไขคำให้อยู่ในรูปแบบที่ถูกต้องเพื่อความแม่นยำมากขึ้น

โปรแกรม 3.7 การแก้คำผิด

```

text = process_text(text)
# ['ลื้ัน', 'ระเท็ก', 'โคตร', 'มันส์', 'ไม่ต้อง', 'คิด', 'เยอะ', 'จบ', 'ได้', 'แบบ', 'อึหยังวะ', 'มาก',
'เป็น', 'หนัง', 'ที่', 'บันเทิง', 'มาก', '8.5', '/', '10']

```

3.2.2.4 การทำให้คำอยู่ในรูปปกติ (Normalization)

จากการตัดคำในข้อความจากโลกออนไลน์จะพบคำที่ไม่อยู่ในรูปปกติ และการมีการใช้คำหลากหลาย เช่น

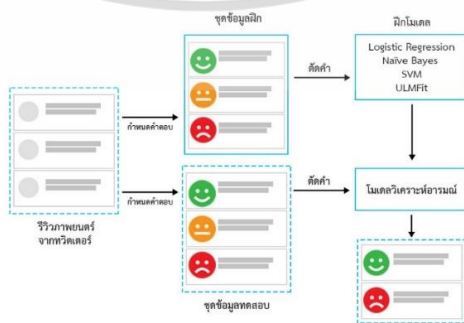
- คำประชด เช่น คีออก แปลงเป็น อีดอก เป็นต้น
- การพิมพ์ซ้ำ เช่น สนุกกกกกก มากกกก เป็นต้น
- คำย่อ เช่น ชท. ฉน. อห. ผกก. ครต. คหสต. ม.ค. เป็นต้น
- คำหยาบ เช่น อี... แปลงเป็น อี... เป็นต้น
- คำศัพท์ภาษาอังกฤษ เช่น review, rom-com, Oscar, move on เป็นต้น

3.2.2.5 การคัดแยกโฆษณา

ในการดึงข้อความจากทวิตเตอร์พบว่ามีข้อความบางส่วนที่เป็นโฆษณาสินค้า โฆษณาภาพยนตร์จากโรงพยาบาลหรือทางค่ายผู้ผลิตภาพยนตร์ นอกจากนี้ยังมีข้อความที่เกาะกระแสของโลกลอนไลน์ซึ่งไม่เกี่ยวข้องกับแฮชแท็กบนข้อความดังกล่าว ก่อนการประมวลผลจึงต้องการกรองคัดแยกทวิตข้อความที่ไม่เกี่ยวข้องกับภาพยนตร์ออกก่อนให้มากที่สุดเท่าที่จะทำได้

3.2.3 การฝึกโมเดล

ในการให้คำตอบด้านอารมณ์ของข้อความจากทวิตเตอร์ ต้องทำการกำหนดตอบ โดยพิจารณาจากแง่มุมส่วนใหญ่ และบริบทในข้อความ หลังจากนั้นจึงนำข้อความเข้าสู่กระบวนการเตรียมข้อมูลก่อนการประมวลผลที่กล่าวไว้ข้างต้น และเข้าสู่กระบวนการฝึก โมเดลโดยได้เลือกอัลกอริทึมในการสร้างโมเดล ได้แก่ Logistic Regression, Naïve Bayes, SVM และ ULMFiT



รูปที่ 3.21 การฝึกโมเดล

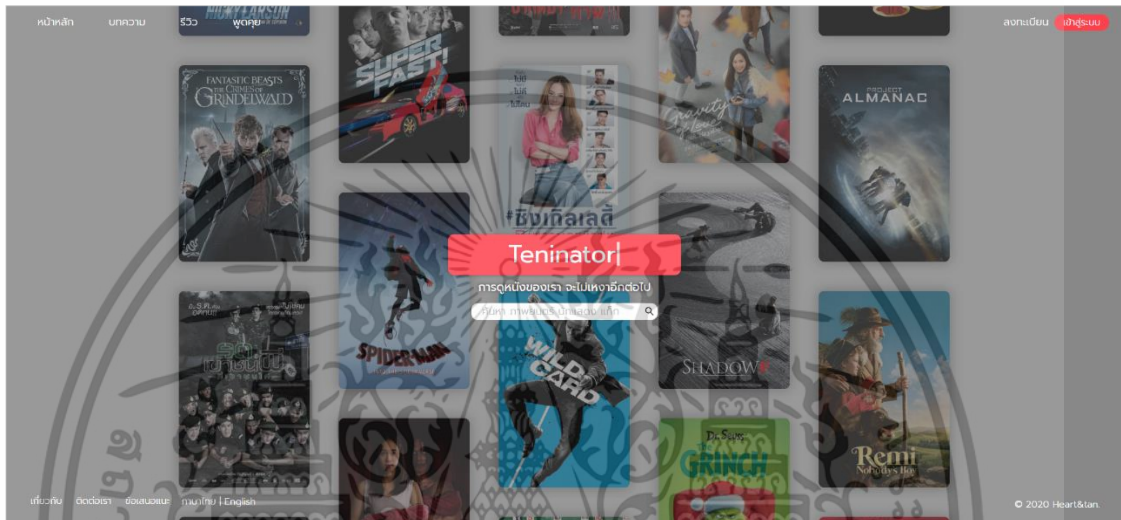
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การใช้งานและการทดลอง

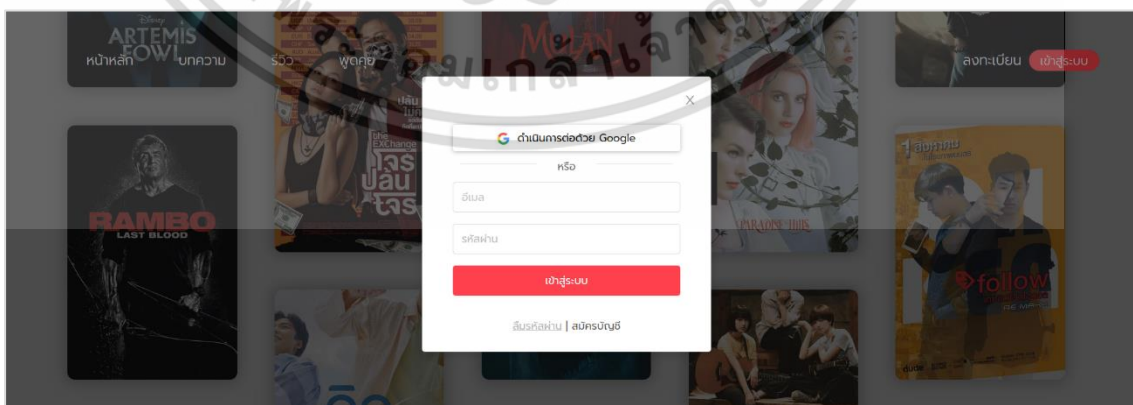
4.1 เว็บไซต์

4.1.1 เว็บไซต์ฝั่งผู้ใช้งาน



รูปที่ 4.1 หน้า Landing page

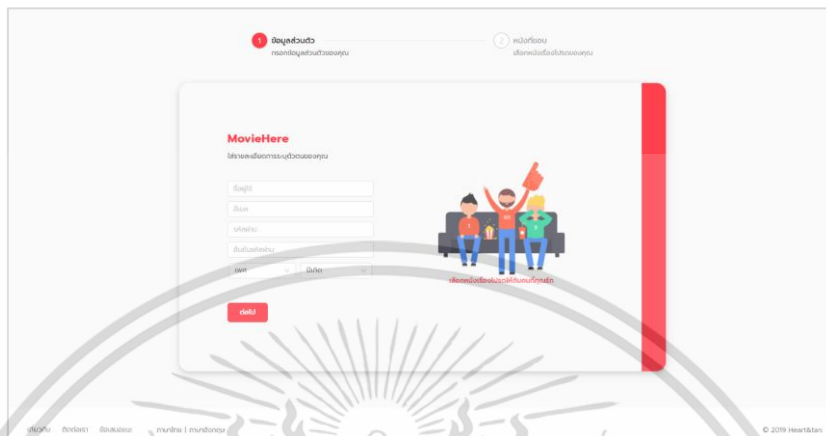
เป็นหน้าเริ่มต้นเมื่อผู้ใช้งานทำการเข้าสู่เว็บไซต์ โดยจะแสดงรายการภาพยนตร์เรื่องต่าง ๆ ที่ถูกสุ่มจากระบบ เมื่อผู้ใช้งานที่โปรสเตอร์ของภาพยนตร์เรื่องใดระบบจะแสดงรายละเอียดของภาพยนตร์นั้น ๆ ดังรูปที่ 4.5 ข)



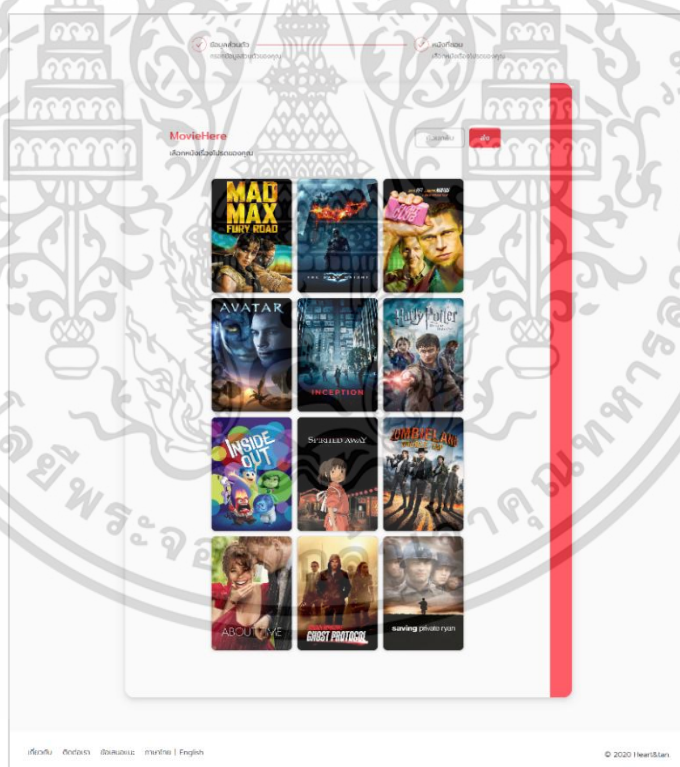
รูปที่ 4.2 หน้าการยืนยันตัวตน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับผู้ใช้ที่ต้องการใช้งานพีเจอาร์ต่าง ๆ ภายในเว็บไซต์จำเป็นต้องทำการเข้าสู่ระบบก่อน โดยจะแสดงเป็นป๊อปอัพ จากนั้นเมื่อผู้ใช้กรอกอีเมลและรหัสผ่านแล้วทำการกดปุ่มเข้าสู่ระบบ ระบบจะทำการตรวจสอบ ถ้ามีตัวตนระบบจะทำการเข้าสู่หน้าหลักต่อไป



ก)



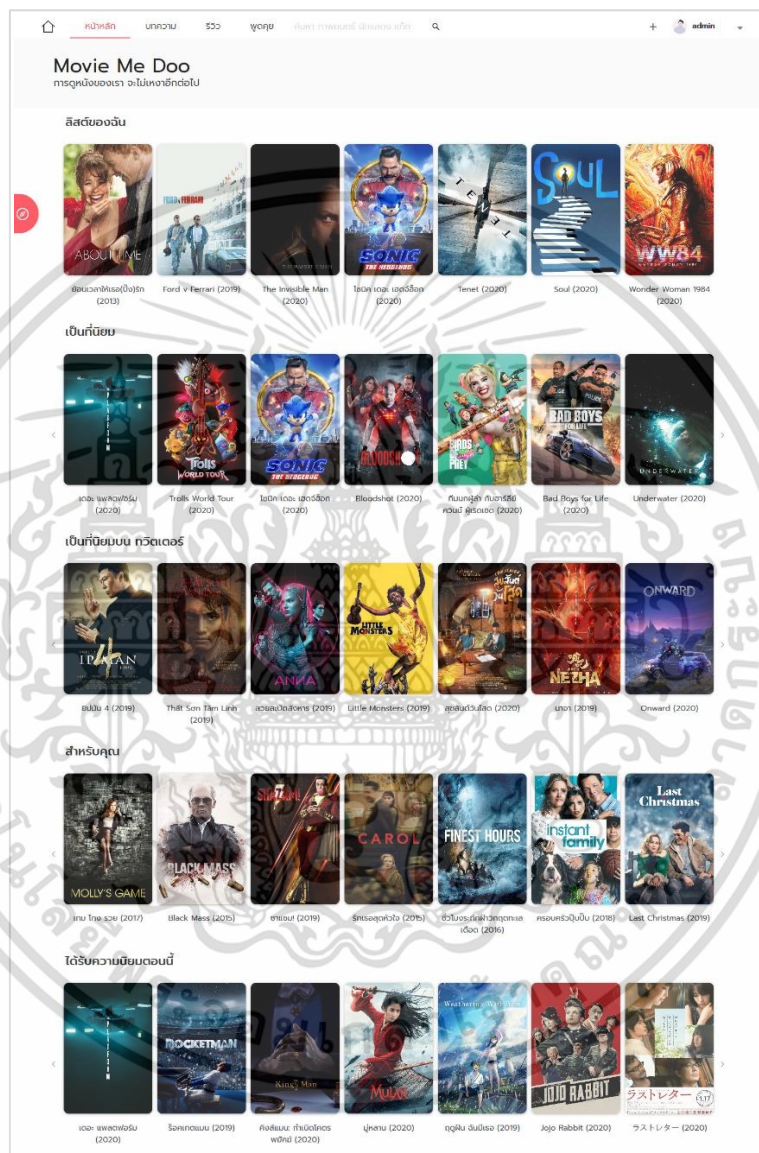
ข)

รูปที่ 4.3 หน้าการลงทะเบียนใช้งาน

- ก) หน้าสำหรับให้ผู้ใช้กรอกรายละเอียดของตนเอง
- ข) หน้าสำหรับให้ผู้ใช้เลือกภาพยนตร์ที่ชอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

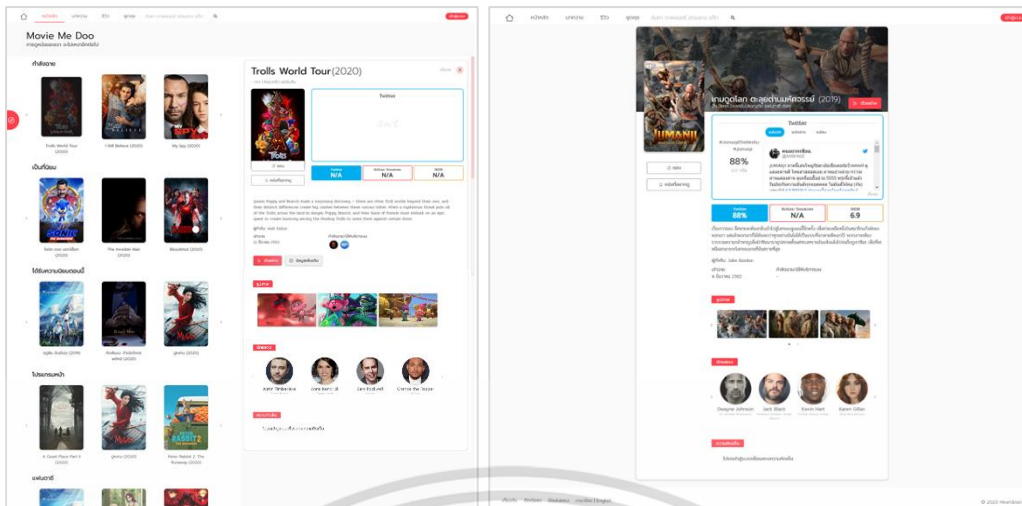
สำหรับผู้ใช้ที่ต้องการลงทะเบียนเข้าใช้งานสามารถปุ่มลงทะเบียนจากหน้าหลักได้จากนั้นผู้ใช้จะต้องกรอกข้อมูลที่ระบบต้องการให้ครบถ้วนคือ ชื่อผู้ใช้ อีเมล รหัสผ่าน เพศ และปีเกิด เมื่อผู้ใช้กรอกข้อมูลครบแล้วกดปุ่มต่อไปจะเป็นหน้าสำหรับให้ผู้ใช้เลือกภาพยนตร์ที่ชอบจากรายการที่ระบบให้มา เมื่อเลือกเสร็จแล้วสามารถกดปุ่มส่ง ถือเป็นอันสิ้นสุดการลงทะเบียน



รูปที่ 4.4 หน้าหลัก โดยแสดงรายการภาพยนตร์

หน้าสำหรับแสดงรายการภาพยนตร์ต่าง ๆ โดยภายในหน้ามีหัวข้อ เช่น รายการของฉัน แนะนำสำหรับคุณ เป็นต้น เมื่อผู้ใช้ทำการกดที่โปสเตอร์ของภาพยนตร์เรื่องใดเรื่องหนึ่ง ระบบจะแสดงรายละเอียดของภาพยนตร์นั้น ๆ โดยแยกเป็นสองส่วนดังรูปที่ 4.5 ก)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ก)

ข)

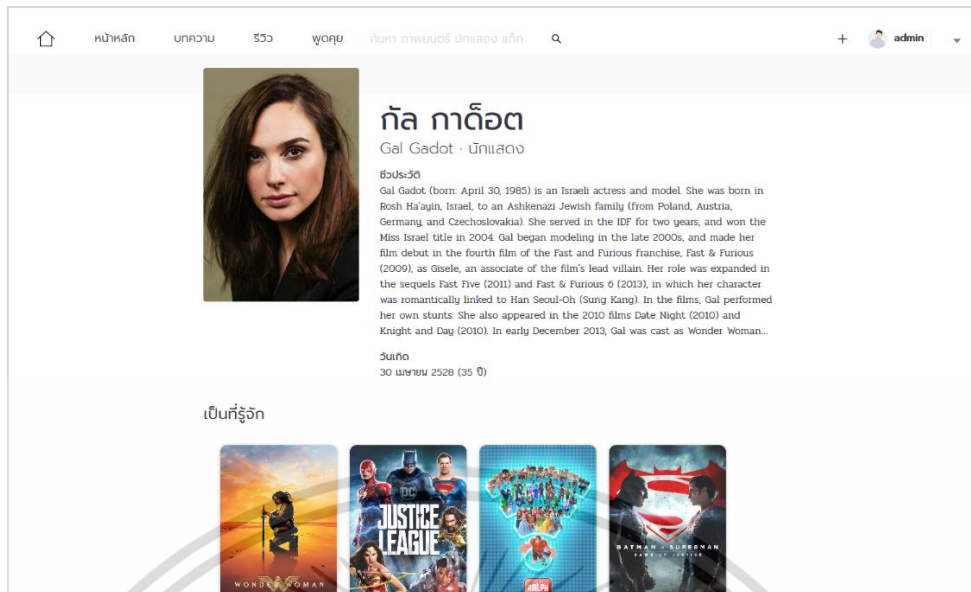
รูปที่ 4.5 หน้าหลัก เมื่อกดที่โปสเตอร์ของภาพยนตร์

ก) หน้าหลัก เมื่อกดที่โปสเตอร์ของภาพยนตร์

ข) หน้าหลัก เมื่อกดปุ่มดูข้อมูลเพิ่มเติม

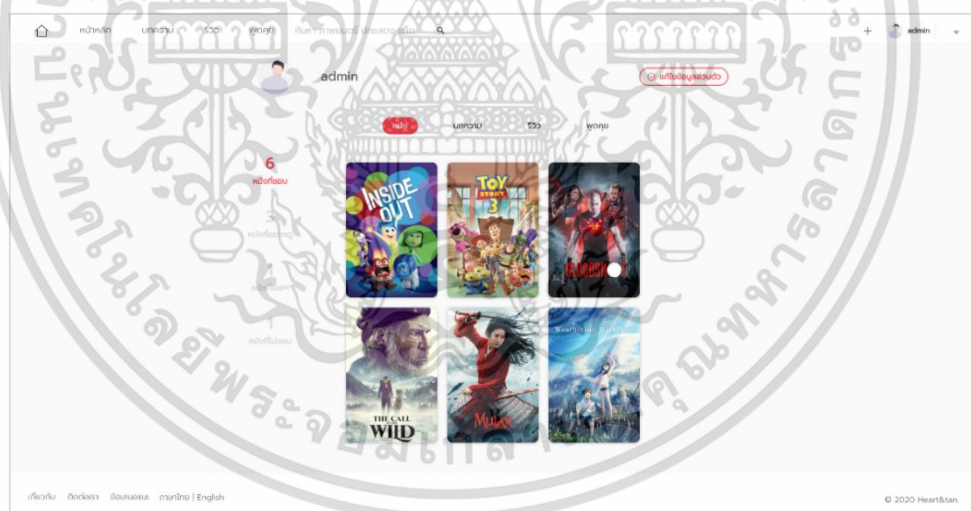
เมื่อผู้ใช้งานกดภาพยนตร์เรื่องใดเรื่องหนึ่งที่หน้าหลักจะแยกส่วนรายละเอียดของภาพยนตร์ไปทางขวาดังรูปที่ 4.5 ก) และเมื่อกดปุ่มดูข้อมูลเพิ่มเติมจะแสดงรายละเอียดต่างๆ ของภาพยนตร์ดังรูปที่ 4.5 ข) ซึ่งจะแสดงรายละเอียดที่เหมือนกันกับรูปที่ 4.5 ก) ประกอบด้วย เรื่องย่อ คะแนนจากเว็บไซต์ IMDB, Rotten tomatoes และ Twitter ซึ่งจะแสดงทวิตที่เกี่ยวข้องกับภาพยนตร์เรื่องนั้น โดยจะแยกแ่งมุมทั้ง 3 แ่งมุม คือ แ่งบน แ่งกลางและแ่งลบ นอกจากนี้ยังแสดง นักแสดง ความคิดเห็นของภาพยนตร์เรื่องนั้น ๆ รีวิวและบทความที่เกี่ยวข้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.6 หน้าแสดงรายละเอียดของนักแสดง

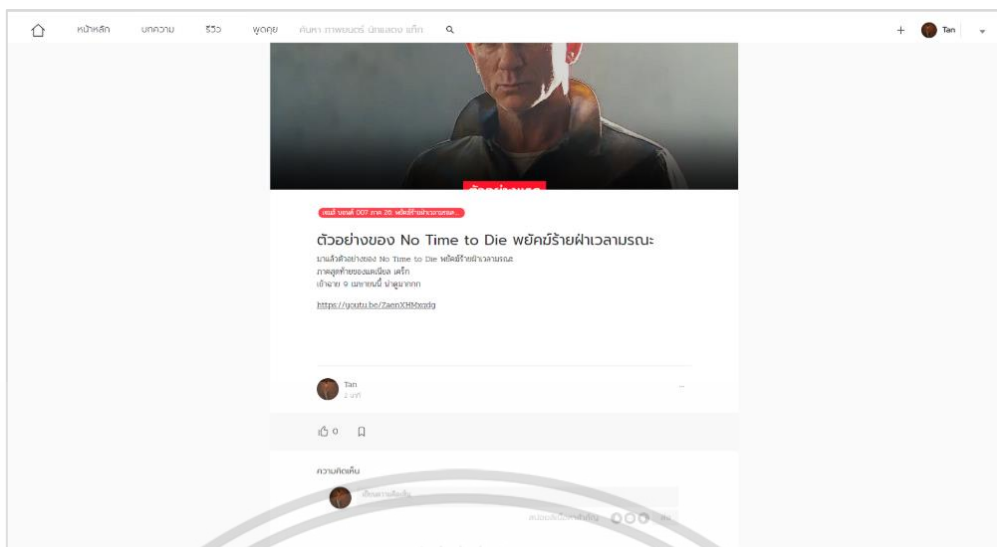
แสดงรายละเอียดของนักแสดง ประกอบด้วย ชื่อ ประวัติโดยย่อ วัน/เดือน/ปีเกิด และรายการภาพยนตร์ที่นักแสดงเคยแสดงมาก่อน



รูปที่ 4.7 หน้าแสดงโปรไฟล์ของผู้ใช้งาน

เป็นหน้าสำหรับผู้ใช้ที่ได้ทำการเข้าสู่ระบบของเว็บไซต์เป็นที่เรียบร้อยแล้ว โดยแสดงรายละเอียดของผู้ใช้ประกอบด้วย ชื่อ และรายการที่ต่าง ๆ คือ หนังสือ บทความ รีวิว และการพูดคุย เช่น รายการภาพยนตร์ที่ชอบ อยากดู เฉยๆ และไม่ชอบ นอกจากนี้ผู้ใช้ยังสามารถแก้ไขรายละเอียดของตนเองได้โดยกดปุ่มแก้ไขข้อมูลส่วนตัว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.8 หน้าบทความ

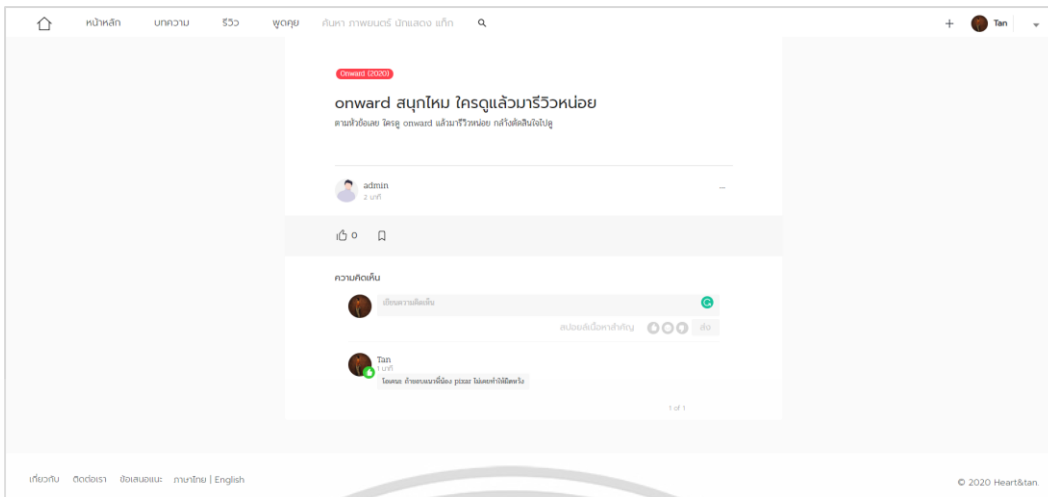
หน้าสำหรับบทความทางภาพยนตร์ โดยสามารถใส่หัวข้อ ข้อความ และรูปภาพ ซึ่งผู้ใช้อื่นสามารถแสดงความคิดเห็นได้



รูปที่ 4.9 หน้ารีวิว

หน้าสำหรับรีวิวภาพยนตร์ เหมือนกับการเขียนบทความ โดยสามารถใส่หัวข้อ ข้อความ รูปภาพ และคะแนนของภาพยนตร์เรื่องนั้น ๆ อีกทั้งผู้ใช้อื่นสามารถแสดงความคิดเห็นได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.10 หน้ากระดานการพูดคุย

หน้าสำหรับให้ผู้ใช้ทำการพูดคุยเพื่อถามคำถามเกี่ยวกับภาพยนตร์ต่าง ๆ โดยผู้ใช้สามารถแสดงความคิดเห็นหรือตั้งคำถามสอบถามเมื่อผู้ใช้ทำการเข้าสู่ระบบของเว็บไซต์แล้ว

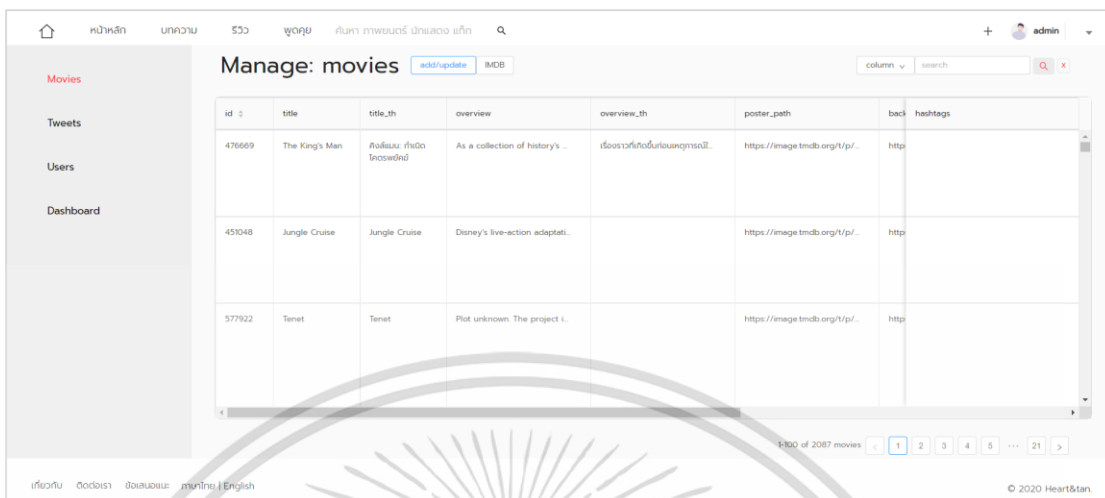


รูปที่ 4.11 ฟังก์ชันการกดชอบ เหยียด หรือไม่ชอบสำหรับภาพยนตร์เรื่องต่างๆ

เป็นฟังก์ชันเมื่อผู้ใช้ทำการเข้าสู่ระบบเข้าสู่ระบบของเว็บไซต์เป็นที่เรียบร้อยแล้ว โดยผู้ใช้สามารถกดปุ่มชอบ เหยียด ไม่ชอบ หรือเก็บไว้ในลิสต์ของภาพยนตร์เรื่องต่างๆ ได้ เมื่อผู้ใช้ทำการกดปุ่มดังกล่าว ภาพยนตร์จะถูกจัดเก็บไปยังหน้าโปรไฟล์ ดังรูปที่ 4.7

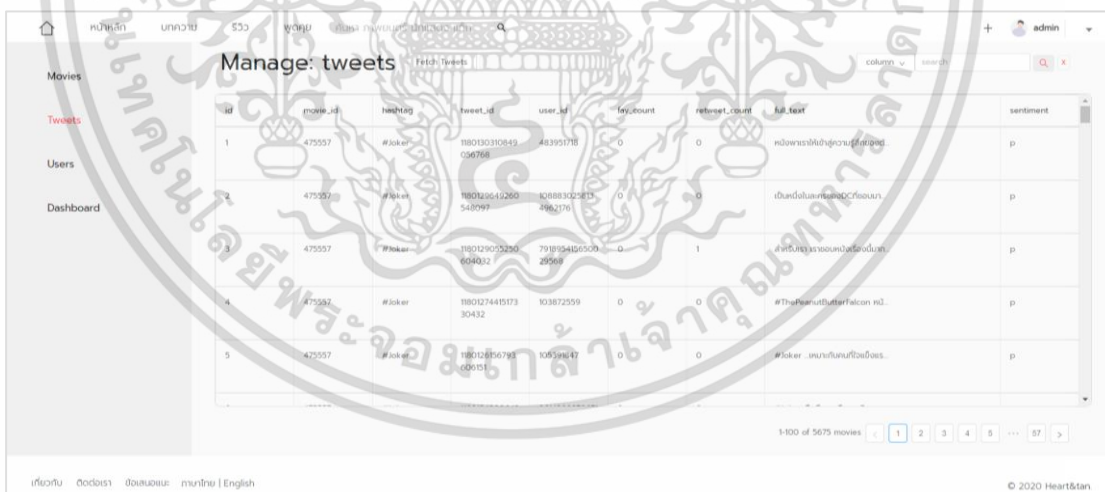
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2 เว็บไซต์ฝั่งผู้ดูแล



รูปที่ 4.12 หน้าสำหรับจัดการภาพยนตร์

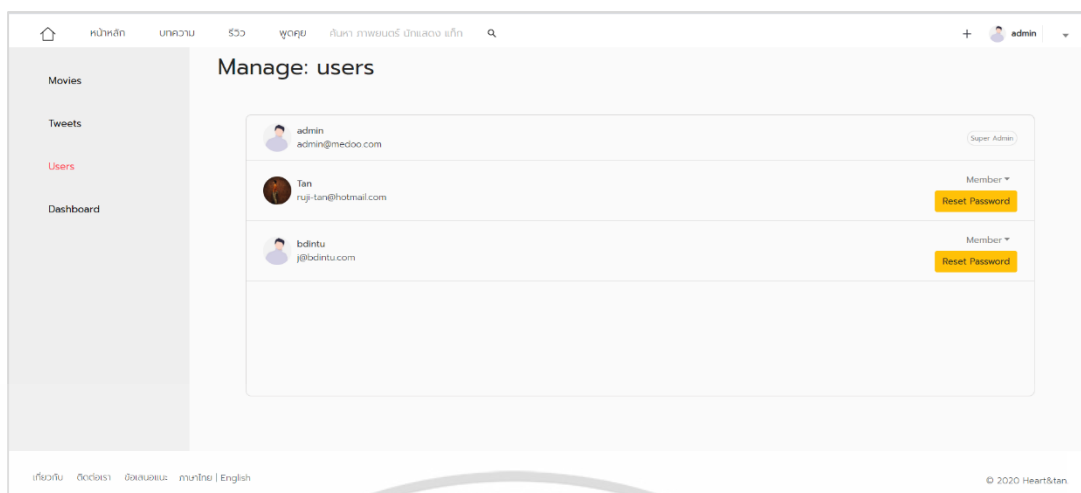
เป็นหน้าสำหรับจัดการภาพยนตร์ โดยผู้ดูแลระบบสามารถดึงภาพยนตร์เรื่องใหม่จาก API ที่กำหนดไว้ หรืออาจปรับเปลี่ยนรายละเอียดของภาพยนตร์เรื่องต่าง ๆ ได้ และสามารถทำการค้นหาภาพยนตร์ได้จาก ID หรือ ชื่อเรื่อง เป็นต้น



รูปที่ 4.13 หน้าสำหรับจัดการข้อความทวิต

เป็นหน้าสำหรับจัดการข้อความทวิต ผู้ดูแลระบบสามารถดึงข้อความทวิตจาก API ที่กำหนด โดยสามารถทำการส่งข้อความทวิตเหล่านั้นไปยัง API ที่ทำการตรวจสอบว่าข้อความทวิตดังกล่าวมีแง่อารมณ์ในด้านไหน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.14 หน้าสำหรับจัดการสมาชิก

เป็นหน้าสำหรับการจัดการสมาชิกของเว็บไซต์ โดยผู้ดูแลระบบสามารถเปลี่ยนแปลงบทบาทของสมาชิก หรือจัดการรหัสผ่านเมื่อสมาชิกลืมรหัสผ่าน เป็นต้น

4.2 ผลการทดสอบการวิเคราะห์อารมณ์จากข้อความ

4.2.1 ข้อมูลที่จัดเก็บทั้งหมด

ตารางที่ 4.1 จำนวนข้อมูลที่จัดเก็บในแต่ละแง่มุม

อารมณ์ของข้อความ	จำนวน(ข้อความ)
แง่บวก	7367
แง่กลาง	746
แง่ลบ	922
ไม่บ่งบอกอารมณ์	9711
โฆษณา	292
รวม	19038

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 ชุดข้อมูลสำหรับฝึกโมเดล

ชุดข้อมูลที่ใช้สำหรับการฝึกโมเดลต่าง ๆ ประกอบด้วยอารมณ์จากข้อความทั้ง 3 แง่มุม คือ แง่บวก (positive) แง่กลาง (neural) และแง่ลบ (negative) เป็นจำนวนดังตารางที่ 4.1

ตารางที่ 4.2 จำนวนข้อมูลที่ใช้ในการฝึกโมเดล

อารมณ์ของข้อความ	จำนวน(ข้อความ)
แง่บวก	705
แง่กลาง	478
แง่ลบ	703
รวม	1886

ตารางที่ 4.3 จำนวนข้อมูลที่ใช้ในการทดสอบโมเดล

อารมณ์ของข้อความ	จำนวน(ข้อความ)
แง่บวก	178
แง่กลาง	93
แง่ลบ	130
รวม	401

ตารางที่ 4.4 ตัวอย่างข้อมูลที่ใช้ในการฝึกโมเดล

ข้อความ	อารมณ์ของข้อความ
#crawl1 สนุกมากกกก คือลุ้นส์ต ตกใจจริงๆ ตกใจจนต้องขำหน้าป๊อปคอร์นลอยมาข้างหลัง เต็มรองเท้าอีกนี่เลยจ้า ตอนแรกนึกว่าไม่เยอะ พอจะออกจากโรงเท่านั้นแหละ เต็มพื้น 555555	แง่บวก
อยากดู #OnceUponATimeInHollywood อีก 99+รอบ	แง่บวก
#Joker วาดินทะเลดูกราฟ ที่เหลือก็อื่นๆ ออกแนว "ก็ดีนะ" #movietwit	แง่กลาง
#AngelHasFallen สนุกตามมาตรฐาน ดูแบบไม่คาดหวัง ดูฆ่าเวลาไปมันส์ ๆ แต่ภาคเก่ามันส์กว่า #movietwit	แง่กลาง
รู้สึกเหมือนหนังตลกมากกว่า และที่ผิดหวังมากที่สุดคือฉากเว็บบๆทำกุปวุดหัว ปวดตาจนไม่เกรนแตกออกจากโรง /กุมขมับ 🙄 #ITChapter2	แง่ลบ
#มิสเตอร์คือกันทำเหรียญทอง 6/10 เคมีไม่เข้ากัน ดูรอดเพราะสายฮา มุกดี แต่หนังเฉื่อยๆ นนกุล กล้ามดีมาก https://t.co/IRLphQdtF1	แง่ลบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3 ผลลัพธ์การฝึกข้อมูล

4.2.3.1 Logistic Regression

การทดสอบความแม่นยำของโมเดล Logistic Regression ในการวิเคราะห์ของ อารมณ์ที่มีต่อภาพยนตร์โดยแบ่งข้อมูลออกเป็น 10 ส่วน ในการฝึกมีอัตราความแม่นยำเฉลี่ยเป็น ร้อยละ 77.68 แล้วเมื่อทำการฝึกโมเดลกับข้อมูลฝึกอีกรอบความแม่นยำร้อยละ 94.86 และทำการ ทดลองกับชุดทดสอบมีอัตราความแม่นยำเป็นร้อยละ 85.79 สามารถสรุปเป็นคอนฟิวชันเมตริกซ์ ดังตารางที่ 4.5 และ ตารางที่ 4.6

ตารางที่ 4.5 คอนฟิวชันเมตริกซ์แสดงผลการฝึกการวิเคราะห์อารมณ์ด้วยโมเดล Logistic Regression กับชุดข้อมูลฝึกสอน

actual \ predicted	predicted		
	แง่บวก	แง่กลาง	แง่ลบ
แง่บวก	685	23	11
แง่กลาง	12	425	13
แง่ลบ	8	30	679

ตารางที่ 4.6 คอนฟิวชันเมตริกซ์แสดงผลการทดสอบการวิเคราะห์อารมณ์ด้วยโมเดล Logistic Regression กับชุดข้อมูลทดสอบ

actual \ predicted	predicted		
	แง่บวก	แง่กลาง	แง่ลบ
แง่บวก	158	9	11
แง่กลาง	11	73	9
แง่ลบ	8	9	113

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 recall, precision และ F1-score ของโมเดล Logistic Regression ของชุดข้อมูล

ทดสอบ

Performance measure \ แม้มุม	แง่บวก	แง่กลาง	แง่ลบ
Precision	0.892	0.785	0.863
Recall	0.887	0.785	0.869
F1-score	0.890	0.785	0.866

จากตารางที่ 4.7 พบว่าการทดสอบโมเดล โดยข้อมูลทดสอบด้วยโมเดล logistic regression มีความถูกต้องของการทำนายแง่บวกมากที่สุด รองลงมาเป็นแง่ลบ และแง่กลาง ตามลำดับ โดยดูได้จาก recall ของชุดข้อมูลทดสอบ ซึ่งแง่บวกจะถูกทำนายเป็นแง่ลบมากที่สุด โดยแง่กลางจะถูกทำนายเป็นแง่บวกเป็นส่วนมาก

4.2.3.2 Naïve Bayes

การทดสอบความแม่นยำของโมเดล Naïve Bayes ในการวิเคราะห์ของอารมณ์ที่มีต่อภาพยนตร์ ฝึกโดยการแบ่งข้อมูลออกเป็น 10 ส่วน ในการฝึกมีอัตราความแม่นยำเฉลี่ยเป็นร้อยละ 63.83 แล้วเมื่อทำการฝึกโมเดลกับข้อมูลฝึกอีกรอบความแม่นยำร้อยละ 70.36 และทำการทดลองกับชุดทดสอบมีอัตราความแม่นยำเป็นร้อยละ 66.83 สามารถสรุปเป็นคอนฟิวชันเมตริกซ์ ดังตารางที่ 4.8 และ ตารางที่ 4.9

ตารางที่ 4.8 คอนฟิวชันเมตริกซ์แสดงผลการฝึกการวิเคราะห์อารมณ์ด้วยโมเดล Naïve Bayes

actual \ predicted	แง่บวก	แง่กลาง	แง่ลบ
แง่บวก	440	88	177
แง่กลาง	33	325	120
แง่ลบ	23	118	562

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.9 คอนฟิวชันเมทริกซ์แสดงผลการทดสอบการวิเคราะห์อารมณ์ด้วยโมเดล Naïve Bayes กับชุดข้อมูลทดสอบ

predicted \ actual	แง่บวก	แง่กลาง	แง่ลบ
แง่บวก	100	39	39
แง่กลาง	7	65	21
แง่ลบ	7	20	103

ตารางที่ 4.10 recall, precision และ F1-score ของโมเดล Naïve Bayes ของชุดข้อมูลทดสอบ

Performance measure \ measure	แง่บวก	แง่กลาง	แง่ลบ
Precision	0.877	0.524	0.632
Recall	0.561	0.699	0.792
F1-score	0.685	0.599	0.703

จากตารางที่ 4.10 พบว่าโมเดลมีความถูกต้องของการทำนายเทียบกับความเป็นจริง (recall) ในแง่ลบถูกมากที่สุด รองลงมาเป็นแง่กลาง และแง่บวก ตามลำดับ โดยแง่บวกทำนายถูกต้องมากที่สุดในแง่ของความแม่นยำว่าเป็นแง่ลบ รองลงมาเป็นแง่ลบ และแง่กลาง นอกจากนี้จะสังเกตได้ว่าข้อความแง่บวกถูกทำนายเป็นแง่ลบ และแง่กลางประมาณกว่าครึ่งหนึ่งของแง่บวกทั้งหมด

4.2.3.3 SVM

การทดสอบความแม่นยำของโมเดล SVM ในการวิเคราะห์ของอารมณ์ที่มีต่อภาพยนตร์ ฝึกโดยการแบ่งข้อมูลออกเป็น 10 ส่วน ในการฝึกมีอัตราความแม่นยำเฉลี่ยเป็นร้อยละ 76.40 แล้วเมื่อทำการฝึกโมเดลกับข้อมูลฝึกอีกรอบความแม่นยำร้อยละ 96.82 และทำการทดลองกับชุดทดสอบมีอัตราความแม่นยำเป็นร้อยละ 86.03 สามารถสรุปเป็นคอนฟิวชันเมตริกซ์ดังตารางที่ 4.11 และ ตารางที่ 4.12

ตารางที่ 4.11 คอนฟิวชันเมตริกซ์แสดงผลการฝึกการวิเคราะห์อารมณ์ด้วยโมเดล SVM

actual \ predicted	predicted		
	แง่บวก	แง่กลาง	แง่ลบ
แง่บวก	693	7	5
แง่กลาง	17	445	16
แง่ลบ	5	10	683

ตารางที่ 4.12 คอนฟิวชันเมตริกซ์แสดงผลการทดสอบการวิเคราะห์อารมณ์ด้วยโมเดล SVM กับชุดข้อมูลทดสอบ

actual \ predicted	predicted		
	แง่บวก	แง่กลาง	แง่ลบ
แง่บวก	156	13	9
แง่กลาง	11	75	7
แง่ลบ	7	9	114

ตารางที่ 4.13 recall, precision และ F1-score ของโมเดล SVM ของชุดข้อมูลทดสอบ

Performance measure	แง่บวก		
	แง่บวก	แง่กลาง	แง่ลบ
Precision	0.90	0.77	0.88
Recall	0.88	0.81	0.88
F1-score	0.89	0.79	0.88

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.13 พบว่าโมเดลมีความถูกต้องของการทำนายเทียบกับความเป็นจริง (recall) แ่งลบมากที่สุด รองลงมาเป็นแ่งบวก และแ่งกลาง ตามลำดับ โดยแ่งบวกทำนายถูกต้องมาก ในแ่งการทำนายว่าเป็นแ่งที่ถูกต้องมากที่สุด รองลงมาเป็นแ่งลบ และแ่งกลาง

4.2.3.4 ULMFiT

การทดสอบความแม่นยำของโมเดล ULMFiT ในการวิเคราะห์ของอารมณ์ที่มีต่อ ภาพยนตร์ โดยแบ่งข้อมูลของในการฝึกมีอัตราความแม่นยำเป็นร้อยละ 88.28 โดยการทดลองกับ ชุดทดสอบมีอัตราความแม่นยำเป็นร้อยละ 75.31 สามารถสรุปเป็นคอนฟิวชันเมตริกซ์ดังตารางที่ 4.14 และ ตารางที่ 4.15

ตารางที่ 4.14 คอนฟิวชันเมตริกซ์แสดงผลการฝึกการวิเคราะห์อารมณ์ด้วยโมเดล ULMFiT กับชุด ข้อมูลฝึกสอน

actual \ predicted	predicted		
	แ่งบวก	แ่งกลาง	แ่งลบ
แ่งบวก	583	65	57
แ่งกลาง	11	411	56
แ่งลบ	4	24	671

ตารางที่ 4.15 คอนฟิวชันเมตริกซ์แสดงผลการทดสอบการวิเคราะห์อารมณ์ด้วย ULMFiT กับชุด ข้อมูลทดสอบ

actual \ predicted	predicted		
	แ่งบวก	แ่งกลาง	แ่งลบ
แ่งบวก	116	43	19
แ่งกลาง	10	73	10
แ่งลบ	5	13	112

ตารางที่ 4.16 recall, precision และ F1-score ของโมเดล ULMFiT ของชุดข้อมูลทดสอบ

Performance measure \ แม้มุม	แฉ่บวค	แฉ่กลลาง	แฉ่ลลบ
Precision	0.885	0.566	0.794
Recall	0.652	0.785	0.866
F1-score	0.751	0.658	0.827

จากตารางที่ 4.16 พบว่าการทดสอบโมเดลโดยข้อมูลทดสอบด้วยโมเดล ULMFiT มีความถูกต้องของการทำนายแฉ่ลลบมากที่สุด รองลงมาเป็นแฉ่กลลาง และแฉ่บวค ตามลำดับ โดยดูได้จาก recall ของชุดข้อมูลทดสอบ ซึ่งแฉ่บวคจะถูกทำนายเป็นแฉ่กลลางมากที่สุด แฉ่ลลบจะถูกทำนายเป็นแฉ่กลลางเป็นส่วนมาก

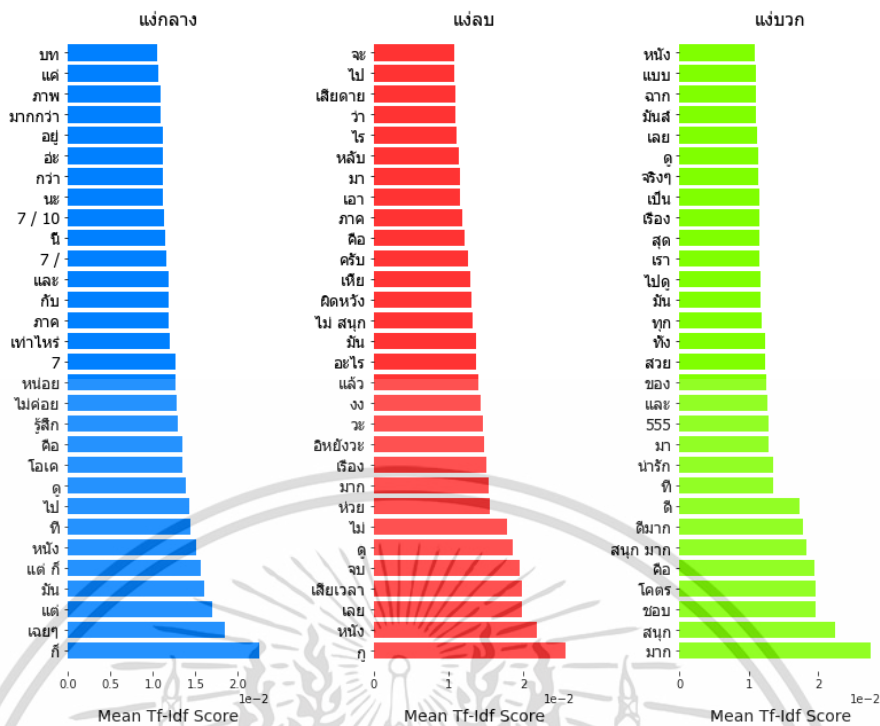
4.2.3.5 TF-IDF

ตารางที่ 4.17 การเปรียบเทียบความแม่นยำในการเลือกใช้ n-gram ในช่วงต่าง ๆ สำหรับการทำให้ TF-IDF กับโมเดล Logistic Regression และ SVM

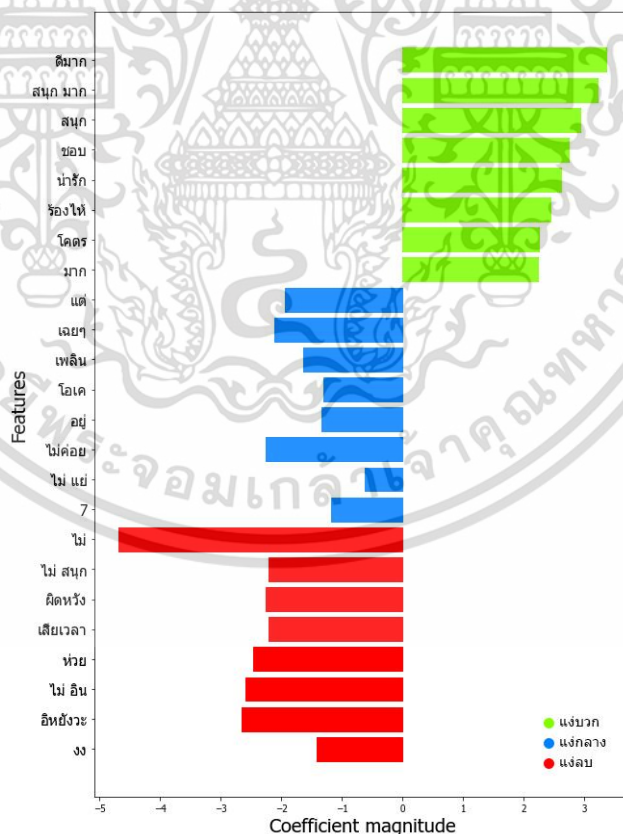
โมเดล	ชุดข้อมูล	ช่วง n-gram								
		(1,2)	(1,3)	(1,4)	(1,5)	(2,3)	(2,4)	(2,5)	(3,4)	(3,5)
Logistic	ชุดฝึก	94.11	94.00	93.96	93.96	88.44	88.23	88.23	65.38	65.43
	ชุดทดสอบ	85.53	85.79	85.79	85.79	74.56	74.31	74.31	62.59	62.84
SVM	ชุดฝึก	96.66	96.82	96.77	96.77	91.89	91.78	91.89	61.45	61.51
	ชุดทดสอบ	86.03	85.79	86.28	86.03	74.06	73.81	73.56	52.87	52.62

จากตารางที่ 4.17 จะเห็นว่า n-gram ในช่วง (1, 4) ของโมเดล Logistic Regression และโมเดล SVM มีความแม่นยำของข้อมูลชุดทดสอบมากที่สุด จึงทำการเลือก n-gram ในช่วง (1, 4) ใช้ในการฝึกสอนโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.15 กราฟแสดงค่าเฉลี่ยของคะแนน TF-IDF ของชุดฝึก



รูปที่ 4.16 กราฟความสัมพันธ์ระหว่าง TF-IDF และสมการ Logistic Function

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.4 ปัญหาที่พบและข้อจำกัดของโมเดล

จากการทดสอบกับชุดข้อมูลกับ โมเดลต่าง ๆ คณะผู้จัดทำพบข้อจำกัดของโมเดลที่ไม่สามารถทำนายข้อความบางประเภทได้ ดังต่อไปนี้

- 1) **ประชดประชัน** เป็นข้อความที่มีข้อความ หรือคำเชิงกระทบกระทั่งแตกแค้น เนื้อหาความหมายตรงกันข้ามกับรูปคำ
- 2) **คำความหมายเชิงลบ** เป็นข้อความที่มีคำที่แสดงถึงความหมายเชิงลบ เช่น ไม่ คำหยาบหรือคำอุทานต่าง ๆ เช่น เหี้ย แม่ง แต่สื่อความหมายในทางที่ดีตามบริบท
- 3) **คำที่ไม่พบในชุดฝึก** เป็นข้อความที่คำที่ไม่เคยพบ หรือพบน้อยในแง่มุมนั้น ๆ
- 4) **ใจความสำคัญน้อย** เป็นข้อความที่ใจความสำคัญที่บ่งบอกอารมณ์ ปรากฏเป็นส่วนน้อยของข้อความทั้งหมด อาจจะมีการยกสถานการณ์ก่อนหน้า โดยอ้างอิงคำพูดของผู้อื่น หรือพรรณนาเรื่องราวก่อน ใจความหลัก
- 5) **ไม่บ่งบอกแง่มุมชัดเจน** เป็นข้อความที่มีข้อความแสดงอารมณ์ แต่ไม่สามารถบ่งบอกแง่มุมได้อย่างชัดเจน
- 6) **ความรู้ทางโลกจริง** เป็นข้อความที่ต้องใช้ความรู้ทางโลกจริงในการเข้าใจแง่มุม

ตารางที่ 4.18 ข้อความที่เป็นปัญหา

ประเภท	ข้อความ
ประชดประชัน	เมื่อคืนไปดูเรื่องนี้มา อย่าถามว่าสนุกมัย ขำจ๊าว เข้าไปหลับสนุกมากก
	เรื่องนี้ที่ประทับใจคงเป็นสีภาพ แก่นี่แหละ
	ภาพ: สวย เนื้อหา:"
คำความหมายเชิงลบ	คุณละร้องเหี้ยทั้งเรื่อง 101010 ไปเรยจะ
	รีวิวกจากที่เอ่ยมาแล้ว! ไม่ใช่อรอมคอม ไม่ขำ ไม่จิกหมอน ไม่สะใจ ไม่ลุ่น ไม่มันส์ ไม่อะไรเลย นี่ไม่ใช่หนังตลาดที่จะมาควาตรายได้ แต่มันจะพาคุณไปเที่ยงกับความเทพของการแสดงระดับออสการ์ขอ ที่ทรงพลังจนแทบจะเปลี่ยนเกย์ให้เป็นชายแท้ได้เลย นางเดินสวยและเอ็กซ์แตกมาก
	ร้องเหี้ยทั้งเรื่อง แม่งเอี้ยยยยยยยย สิบสิบไปเลยยย
	(4.5/5) หนักกะเทยมีงสีอิเกะเทยมาก เพลงประกอบก็คืออยากจะเดินตาม จังหวะเพลง birthday cake ของอิมเมค้ำ หักคะแนนตรงนี้ถือว่า Lizzo จะเด่นกว่านี้
	เหนื่อยมาก ดูจบแล้วอยากจะร้องไห้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.19 ข้อความที่เป็นปัญหา (ต่อ)

ประเภท	ข้อความ
ข้อความขัดแย้ง	หนังสือเสนอมุมมองทางเพศที่แตกต่าง ใครชอบก็คือชอบเลย แต่นี่เฉยๆ ไร
	ไม่ค่อยอินกับจีนหรือเอ็ม ค่อนข้างเฉยๆด้วยซ้ำ แต่พอพูดถึงครอบครัวของจีน ก็ร้องไห้แบบไม่ไหวแล้ววาวว
	ใครไม่อิน กูอิน
	เมื่อวานไปดู เรื่องนี้มาออกมาจากโรง ก็เห็นน้อง ๆ ผู้หญิง สีห้าคนคุยกัน เรื่อง ดาวโคตรเวอร์เลย โคตรไม่สมจริง ไหนจะฟิล์มตอนนางเอกเจออีก เหมือนบอกว่าหนังห่วยไม่สมจริง ... คือมันไม่น่าคิดถึงความจริง ตั้งแต่ ..เป็นหนังที่มีฝีมือแล้วปะวะ แต่หนังดีจริง ชอบ.. รักพลอยซะ"
คำที่ไม่ที่ ไม่พบใน ชุดฝึก	ไม่หือหาว
	ไม่ไหว.. ไม่ไหวมากๆ ดีที่พกทิชชูไป..
	โดยสตรู้สึกว่าจุดหนึ่งที่ทำให้อนิเมะเรื่องนี้เป็นอนิเมะในดวงใจของใครหลายๆคน ก็เพราะการให้เหตุผลนี้แหละนะ
	ไปดูทั้ง 2 ภาค ฮาอยู่นะ จำจริงจิงเลยอะ ส่วนเนื้อเรื่องไม่ปัง มีที่มาที่ไป 8.5/10
	มึง พี่นาคภาคสองคือกูจึ้งมาก รวมทุกอย่างให้เทอแล้ว ตลก น่ากลัว ซึ่งแอบงงนิดๆ
ไม่บ่งบอกแง่มุมชัดเจน	ตลก เสร้า ฟิน อิน กลัว
	ถ้าไม่คิด ไรมากก็ตลกดี นักแสดงดีทุกคน5555555555555555
	เทคนิคการถ่ายภาพดีเยี่ยม งานสร้างสุดยอดเยี่ยมๆ ให้ความรู้ลึกซึ้งนี้มาติดอย่างเปี่ยมล้น เทคนิคเล่นกับแสงเงาคือเจ๋ง ชวนตื่นตาตื่นใจ และยังชวนตื่นเต้นเกือบตลอดทั้งเรื่องอีกด้วย แต่ก็โคดเด่นเกินไปจนทำให้เนื้อเรื่องและการแสดงเป็นตัวรอง ไม่เกิดการตกตะกอนใดๆในส่วนนี้เมื่อดูจบ
	กูไม่รู้จะให้คะแนนยังไงเลย 55555555 ดูจบมาสัปดาห์ละยังไม่รู้เลยว่าตกลงกูชอบป่าว 5555555555555555
	ภาพสวย จบแบบ อึ้งมมม
ความรู้ทางโลกจริง	ความรู้สึกเหมือนดู Green lantern เลย
	รีวิวลายเส้นอนิเมะของค่ายโตเอ เรื่องลายเส้น โดจินเดือนที่กูอ่านยังสวยกว่านี้เลยขอบคุณค่ะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.5 สรุปผลโมเดลที่เลือกใช้

จากการทดลองทั้งสี่ โมเดล ได้แก่ Logistic Regression, Naïve Bayes, SVM และ ULMFiT ซึ่ง Logistic Regression และ SVM มีความแม่นยำในการแบ่งแ่งมุมของข้อมูลทดสอบมากที่สุด โดยมีความแม่นยำของโมเดล SVM อยู่ที่ร้อยละ 86.03 และ โมเดล Logistic Regression มีความแม่นยำอยู่ที่ร้อยละ 85.79 โดยพบว่าโมเดล SVM สามารถทำนายแ่งกลางและแ่งลบได้ดีว่า แต่ทำนายแ่งบวกได้ถูกน้อยกว่าโมเดล Logistic Regression โดยโมเดลที่ได้ความแม่นยำรองลงมาคือ ULMFiT ซึ่งมีความแม่นยำของชุดทดสอบข้อมูลอยู่ที่ร้อยละ 75.31 สาเหตุที่ได้ค่าน้อยกว่าอาจมีสาเหตุจากชุดข้อมูลที่ใช้ในการฝึกสอน โมเดลมีจำนวนค่อนข้างน้อย จนทำให้โมเดลยังไม่สามารถเรียนรู้รูปแบบทางภาษาได้ดีเท่าที่ควร และ โมเดล Naïve Bayes มีความแม่นยำของชุดทดสอบข้อมูล อยู่ที่ร้อยละ 66.83 ซึ่งเป็นความแม่นยำที่ได้น้อยที่สุด

จกตารางด้านล่างผู้จัดทำจึงเลือกใช้ SVM ที่มีความแม่นยำของชุดทดสอบอยู่ที่ร้อยละ 86.03

ตารางที่ 4.20 ความแม่นยำชุดทดสอบของแต่ละโมเดล

โมเดล	ความแม่นยำร้อยละของชุดทดสอบ
SVM	86.03
Logistic Regression	85.79
ULMFiT	75.31
Naïve Bayes	66.83

บทที่ 5

สรุปผลและข้อเสนอแนะ

5.1 บทสรุป

เว็บแอปพลิเคชันภาพยนตร์นี้เกิดขึ้นจากการที่ทางคณะผู้จัดทำสังเกตเห็นว่าในประเทศไทยยังไม่มีเว็บไซต์แหล่งรวมภาพยนตร์สำหรับคนไทยที่รวบรวมบทความ รีวิว หรือกระดานสนทนา ของภาพยนตร์เรื่องต่าง ๆ และใช้การประมวลผลทางภาษาเพื่อช่วยในการตัดสินใจในการรับชม โดยเว็บไซต์นี้ประกอบไปด้วย ส่วนการแนะนำ ซึ่งมีกรนำทวิตของภาพยนตร์เรื่องต่าง ๆ มาตีความหมายเป็นแง่บวก แง่กลาง หรือแง่ลบ แล้วแสดงบนหน้าเว็บไซต์ ประกอบกับการแนะนำจากภาพยนตร์ที่ผู้ใช้กดถูกใจเพื่อช่วยผู้ใช้ในการตัดสินใจเลือกชม และส่วนต่อมาก็คือ ส่วนรีวิวภาพยนตร์ ส่วนบทความภาพยนตร์ และส่วนของกระดานสนทนา

ในโครงการที่ 1 และ 2 ทางผู้จัดทำได้ศึกษาหาความรู้ที่ใช้ในการออกแบบและพัฒนาเว็บไซต์ รวมทั้งทดลองโมเดลในการประมวลผลทางด้านภาษาเพื่อใช้ในการวิเคราะห์อารมณ์บนข้อความ วิเคราะห์ภาพยนตร์จากทวิตเตอร์ ในการพัฒนาโครงการแบ่งเป็นสองส่วนใหญ่ ได้แก่ การพัฒนาเว็บไซต์ และการพัฒนาโมเดลในการประมวลผลทางด้านภาษา

ในส่วนของการพัฒนาเว็บไซต์ทางผู้จัดทำได้ทำการออกแบบและพัฒนาระบบได้แก่ ระบบสมาชิก ระบบการดูรายการหนัง ระบบการให้ความรู้เกี่ยวกับภาพยนตร์ ระบบผู้ดูแล ระบบเก็บภาพยนตร์ที่อยากชม การดูโปรไฟล์ ระบบรีวิว ระบบบทความ และระบบกระดานสนทนา

ในส่วนของการพัฒนาการประมวลผลทางภาษา เนื่องจากผู้จัดได้รวบรวมข้อความวิเคราะห์ภาพยนตร์จากทวิตเตอร์เป็นส่วนใหญ่ และใช้ข้อความจากชุมชนออนไลน์อื่น ๆ บางส่วน จากการเก็บรวบรวมทั้งหมดประมาณ 19000 ข้อความ โดยดึงจากแฮชแท็กภาพยนตร์ พบว่าสามารถแยกประเภทข้อความออกได้เป็นเป็น 5 ส่วน ได้แก่ ข้อความแง่บวก แง่กลาง แง่ลบ โฆษณา ข้อความที่ไม่บ่งบอกอารมณ์ความรู้สึกเกี่ยวกับภาพยนตร์ โดยข้อความเกือบครึ่งหนึ่งของข้อความที่รวบรวมได้เป็นข้อความที่ไม่บ่งบอกอารมณ์ ทำให้ในการรวบรวมและคัดแยกใช้เวลานาน ซึ่งในส่วนของข้อความที่บ่งบอกอารมณ์พบว่า มีข้อความแง่บวกมากกว่าแง่อื่น ๆ หลายเท่าตัว โดยพบแง่ลบ และแง่กลาง รองลงมา เนื่องจากผู้ใช้ทวิตเตอร์ไทยส่วนใหญ่มักจะโพสต์ข้อความในเมื่อรู้สึกชอบภาพยนตร์เรื่องนั้น ๆ มากกว่าอารมณ์ความรู้สึกอื่น ในขณะที่ข้อความแง่ลบผู้ใช้มักจะไม่แสดงออกทางความคิดเห็นมากนัก นอกเสียจากเป็นภาพยนตร์มีชื่อเสียงที่ทำออกมาได้แย่งจะมีข้อความในแง่ลบหรือแง่กลางมากกว่าปกติ นอกจากนี้ในการพัฒนาโมเดล ผู้จัดทำพบว่าข้อความบนทวิตเตอร์ส่วนใหญ่เป็นการใช้คำแสดง คำพิดี และคำทับศัพท์จำนวนมาก จึงทำให้ต้องมีการสร้างและ

ปรับปรุงโมเดลการตัดคำให้เหมาะสม เช่น เพิ่มคำศัพท์ในการตัดคำ การเปลี่ยน การแก้คำทับศัพท์ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาษาอังกฤษให้ถูกต้อง และการแก้คำที่เขียนผิด ในการทดสอบโมเดลทั้งสี่ โมเดลที่ได้ทำการทดสอบกับชุดข้อมูลจากทวิตเตอร์ พบว่าในการทำนายอารมณ์ทั้งสามแง่มีผลการทำนายอารมณ์ของข้อความในแง่บวกและแง่ลบแม่นยำมากที่สุด โดยโมเดลยังไม่สามารถทำนายอารมณ์ของข้อความแง่กลางได้มากเท่าที่ควรเมื่อเทียบกับแง่มุมอื่นอาจเพราะชุดข้อมูลที่ใช้ในการฝึกโมเดลดังกล่าวมีแง่กลางค่อนข้างน้อย จึงต้องทำการเก็บข้อมูลในแง่กลางมากขึ้นต่อไป ซึ่งในท้ายที่สุดทางผู้จัดทำได้เลือกโมเดล SVM เนื่องจากเป็นโมเดลที่ให้ผลลัพธ์ในการทดสอบดีที่สุด

5.2 ความรู้ที่ได้จากทำโครงการ

- 1) ได้รับความรู้ในการออกแบบ และพัฒนาเว็บแอปพลิเคชัน รวมทั้งการออกแบบและการพัฒนาฐานข้อมูล
- 2) ได้รับความรู้ในการออกแบบและพัฒนาระบบประมวลผลทางภาษาไทย
- 3) ได้ฝึกการทำงานการทำงานเป็นทีม โดยใช้การทำงานแบบอไจล์ (Agile)

5.3 ปัญหาอุปสรรคและแนวทางการแก้ไข

- 1) ข้อมูลแง่กลาง และแง่ลบที่ทำการดึงมาจากทวิตเตอร์ มีจำนวนค่อนข้างน้อยเมื่อเทียบกับข้อมูลแง่บวก สามารถแก้ไขได้จากการนำข้อความแง่กลางและแง่ลบจากเว็บไซต์อื่น ๆ เช่น ความคิดเห็นจากโพสทีวีในเฟซบุ๊ก เป็นต้น
- 2) ข้อความที่ทำการดึงมาจากทวิตเตอร์มีคำที่เขียนผิดเป็นจำนวนมาก ซึ่งอาจทำให้ผลการทำนายจากโมเดลมีความแม่นยำน้อยลง จึงได้ทำพจนานุกรมขึ้นมาสำหรับใช้ในการตัดคำ แสดง และคำที่เขียนผิดที่มักพบบ่อย ๆ บนโลกออนไลน์ และทำการแปลงรูปกลับเป็นคำปกติเพื่อลดความหลากหลายของคำศัพท์ ซึ่งต้องทำการปรับปรุงและแก้ไขอยู่เสมอๆ เพื่อให้การนำไปใช้ต่อในโมเดลถัดไปมีประสิทธิภาพยิ่งขึ้น
- 3) โมเดลที่เลือกใช้ยังไม่สามารถรองรับข้อความที่มีเนื้อหาประชดประชัน ข้อความขัดแย้ง เนื้อหานอกประเด็น หรือกำกวมได้ เช่น “คหสต เราชอบภาคแรกมากกว่า” “ไม่ใช่รอมคอมไม่ข่า ไม่จิกหมอน แต่มันจะพาคุณไปทั้งกับความเทพของการแสดงระดับออสการ์” “เหนื่อยมาก ดูจบแล้วอยากจะร้องไห้”

5.4 แนวทางการพัฒนาต่อ

- 1) ปรับเปลี่ยนโมเดลที่ใช้การเรียนรู้เชิงลึก ให้มีความแม่นยำมากขึ้น ที่เรียนรู้โครงสร้างของประโยค และใช้ลำดับของคำในการประมวลผล
- 2) เพิ่มปริมาณของชุดข้อมูล ด้วยการหาเพิ่มข้อมูลที่เป็นข้อความแง่ต่าง ๆ เพิ่มขึ้นทั้งจากทวิตเตอร์และเว็บไซต์ภายนอกให้มากยิ่งขึ้นกว่าเดิม
- 3) พัฒนาเว็บแอปพลิเคชันต่อ ตามที่ได้ออกแบบไว้ เช่น ส่วนของระบบติดตาม เป็นต้น
- 4) พัฒนาปรับปรุงส่วนต่อประสานกับผู้ใช้ (user interface) และประสบการณ์ของผู้ใช้งานต่อการใช้งานของผู้ใช้งาน (user experience) ให้มีความสวยงามและใช้งานง่ายมากยิ่งขึ้น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

- ทรูปลูกปัญญา. 2561. วิชาภาษาไทย ม.ต้น. [Online]. Available : <http://km.ssk.in.th/2561/upload/20180730133937.pdf>
- ปรัชญา บุญขวัญ. 2560. NLP ไทย ไม่ไปไหนจริงหรือ? (TL;DR). [Online]. Available : <https://www.facebook.com/notes/prachya-boonkwan/nlp-ไทย-ไม่ไปไหนจริงหรือ-tldr/10154811091686242/>
- ยุทธ ไถยวรรณ. 2555. หลักการและการใช้การวิเคราะห์การถดถอยโลจิสติกสำหรับการวิจัย. [Online]. Available : <http://rdi.rmutsv.ac.th/rmutsvrj/download/year4-issue1-2555/p1.pdf>
- วิโรจน์ อรุณมานะกุล และธารทอง แจ่มไพบูลย์. 2561. คู่มือการกำกับข้อมูล Thai Dependency Three Bank ตามแนว Universal Dependencies v.2. [Online]. Available : <https://drive.google.com/file/d/15c4eaCwoOh88q2ZZknTGwRJlyoW4fj1e/view>
- สุภาพร คลังเพชร. 2553. ประสิทธิภาพการจัดกลุ่มของวิธีซัพพอร์ตเวกเตอร์แมทซ์และวิธีเคเน็ยเรตเนเบอร์กรณีข้อมูลที่มีการแจกแจงแบบเสถียรที่มีลักษณะหางหนา. [Online]. Available : https://research.psu.ac.th/files/res_che2553/resche_files/983_CUnit2.pdf
- Digital Ventures team. 2018. Natural Language Processing เทคโนโลยีเชื่อมโยงปัญญาประดิษฐ์กับมนุษย์ด้วย “ภาษา”. [Online] Available : <http://dv.co.th/blog-th/get-to-know-natural-language-processing-nlp/>
- Doratong24. 2017. [ML] LSTMs แบบ (เกือบ) ละเอียดยิบ ๆ. [Online]. Available : <https://medium.com/@tongkornkitt/ml-lstms-แบบ-เกือบ-ละเอียดยิบ-ๆ-a3a55cd37883>
- Keng Surapong. 2018. รวมคำศัพท์เกี่ยวกับ Natural Language Processing (NLP). [Online]. Available : <https://www.bualabs.com/archives/119/รวมคำศัพท์เกี่ยวกับ-nlp-natural-language-proce>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Keras team. 2019. **Keras: The Python Deep Learning library**. [Online].

Available : <https://keras.io/>

Oinkina. 2015. **Understanding LSTM Networks**. [Online].

Available : <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

PyThaiNLP team. 2019. **คู่มือการใช้งาน PyThaiNLP**. [Online].

Available : <https://pythainlp.readthedocs.io/en/latest/pythainlp-1-4-thai/>

Tutorialspoint team. 2019. **Artificial Intelligence Tutorial**. [Online].

Available : https://www.tutorialspoint.com/artificial_intelligence



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.

รายการตารางอธิบายยูสเคสบทที่ 3

ตารางที่ ก.1 อธิบายยูสเคส ลงทะเบียน

Use case name	ลงทะเบียน
Actors	ผู้ใช้งาน
Preconditions	-
Termination outcome	- ผู้ใช้สามารถทำการลงทะเบียน โดยเสร็จสมบูรณ์ - ผู้ใช้ไม่สามารถลงทะเบียนได้ เนื่องจากเกิดข้อผิดพลาด
Use case description	ผู้ใช้งานสามารถทำการลงทะเบียน โดยกรอกรายละเอียดต่าง ๆ
Use case associations	ผ่านทาง Facebook

ตารางที่ ก.2 อธิบายยูสเคส เข้าสู่ระบบ

Use case name	เข้าสู่ระบบ
Actors	ผู้ใช้งาน
Preconditions	-
Termination outcome	- ผู้ใช้สามารถทำการเข้าสู่ระบบ ได้สำเร็จ - ผู้ใช้ไม่สามารถเข้าสู่ระบบได้ เนื่องจากเกิดข้อผิดพลาด
Use case description	ผู้ใช้งานสามารถทำการเข้าสู่ระบบ โดยใส่ข้อมูลที่ระบบต้องการ
Use case associations	ผ่านทาง Facebook

ตารางที่ ก.3 อธิบายยูสเคส แก้ไขโปรไฟล์

Use case name	แก้ไขโปรไฟล์
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถแก้ไขโปรไฟล์ของตนเองได้สำเร็จ โดยกดบันทึกข้อมูล - ผู้ใช้ไม่สามารถแก้ไขโปรไฟล์ได้ เนื่องจากกรอกข้อมูลไม่ถูกต้อง
Use case description	ผู้ใช้งานสามารถแก้ไขโปรไฟล์ของตนเองได้
Use case associations	- ดูโปรไฟล์

ตารางที่ ก.4 อธิบายยูสเคส ผ่านทาง Facebook

Use case name	ผ่านทาง Facebook
Actors	ผู้ใช้งาน
Preconditions	-
Termination outcome	- ผู้ใช้สามารถทำการเข้าสู่ระบบผ่านทาง Facebook ได้สำเร็จ - ผู้ใช้สามารถทำการเข้าสู่ระบบผ่านทาง Facebook ไม่สำเร็จ - ผู้ใช้สามารถทำการลงทะเบียนผ่านทาง Facebook ได้สำเร็จ - ผู้ใช้สามารถทำการลงทะเบียนผ่านทาง Facebook ไม่สำเร็จ
Use case description	ผู้ใช้งานสามารถทำการเข้าสู่ระบบ หรือลงทะเบียนผ่านทาง Facebook ได้
Use case associations	- ลงทะเบียน - เข้าสู่ระบบ

ตารางที่ ก.5 อธิบายยูสเคส ดูโปรไฟล์

Use case name	ดูโปรไฟล์
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ (เมื่อต้องการดูโปรไฟล์ของตนเอง)
Termination outcome	- ผู้ใช้ดูโปรไฟล์ของตนเอง หรือของผู้อื่นได้ - ผู้ใช้ไม่สามารถดูโปรไฟล์ของตนเอง เนื่องจากไม่ได้ทำการเข้าสู่ระบบ
Use case description	ผู้ใช้งานสามารถดูโปรไฟล์ของตนเอง และผู้อื่นได้
Use case associations	- ติดตามผู้ใช้งาน - ยกเลิกติดตามผู้ใช้งาน - แก้ไขโปรไฟล์ - ลบรายการหนึ่ง - ลบรายการบทความ - ลบรายการรีวิว - ลบหัวข้อสนทนา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.6 อธิบายยูสเคส ติดตามผู้ใช้งาน

Use case name	ติดตามผู้ใช้งาน
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถกดติดตามผู้ใช้งานอื่นได้สำเร็จ - ผู้ใช้ไม่สามารถกดติดตามผู้ใช้งานอื่นได้ เนื่องจากไม่ได้ทำการเข้าสู่ระบบ
Use case description	ผู้ใช้สามารถกดติดตามผู้ใช้งานอื่นได้
Use case associations	- คู่มือไฟล์

ตารางที่ ก.7 อธิบายยูสเคส ยกเลิกการติดตามผู้ใช้งาน

Use case name	ยกเลิกการติดตามผู้ใช้งาน
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถยกเลิกการติดตามผู้ใช้งานอื่นได้สำเร็จ
Use case description	ผู้ใช้สามารถกดยกเลิกการติดตามผู้ใช้งานอื่นได้
Use case associations	- คู่มือไฟล์

ตารางที่ ก.8 อธิบายยูสเคส ลบรายการหนัง

Use case name	ลบรายการหนัง
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถลบรายการหนังได้สำเร็จ
Use case description	ผู้ใช้สามารถลบรายการหนังที่ผู้ใช้กดถูกใจ เฉย ๆ ไม่ชอบ หรือหนังที่เก็บไว้ได้
Use case associations	- คู่มือไฟล์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.9 อธิบายยูสเคส ลบรายการบทความ

Use case name	ลบรายการบทความ
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถลบรายการบทความได้สำเร็จ
Use case description	ผู้ใช้สามารถลบรายการบทความที่ผู้ใช้สร้าง ถูกใจ หรือหนังสือที่เก็บไว้ได้
Use case associations	- คู่มือไฟล์

ตารางที่ ก.10 อธิบายยูสเคส ลบรายการรีวิว

Use case name	ลบรายการรีวิว
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถลบรายการรีวิวได้สำเร็จ
Use case description	ผู้ใช้สามารถลบรายการรีวิวที่ผู้ใช้สร้าง ถูกใจ หรือหนังสือที่เก็บไว้ได้
Use case associations	- คู่มือไฟล์

ตารางที่ ก.11 อธิบายยูสเคส ลบรายการกระดานสนทนา

Use case name	ลบรายการกระดานสนทนา
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถลบรายการกระดานสนทนาได้สำเร็จ
Use case description	ผู้ใช้สามารถลบรายการกระดานสนทนาที่ผู้ใช้สร้าง ถูกใจ หรือหนังสือที่เก็บไว้ได้
Use case associations	- คู่มือไฟล์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.12 อธิบายยูสเคส ดูรายการหนังสือ

Use case name	ดูรายการหนังสือ
Actors	ผู้ใช้งาน
Preconditions	-
Termination outcome	- ผู้ใช้สามารถดูรายการหนังสือทั้งหมดได้สำเร็จ - ผู้ใช้ดูข้อมูลรายละเอียดของหนังสือเรื่องต่าง ๆ ได้สำเร็จ
Use case description	ผู้ใช้ดูรายการหนังสือ โดยสามารถดูรายละเอียดต่าง ๆ อีกทั้งยังสามารถดูความคิดเห็นเกี่ยวกับหนังสือของหนังสือเรื่องนั้น ๆ หรือค้นหาหนังสือเพื่อดูรายละเอียดได้
Use case associations	- ดูรายการทวิต - เพิ่มรายการหนังสือที่อยากดู - ลบรายการหนังสือที่อยากดู - เพิ่มรายการหนังสือที่ชอบ/ไม่ชอบ/เฉย ๆ - ลบรายการหนังสือที่ชอบ/ไม่ชอบ/เฉย ๆ - ดูความคิดเห็น - ค้นหา

ตารางที่ ก.13 อธิบายยูสเคส ดูรายการทวิต

Use case name	ดูรายการทวิต
Actors	ผู้ใช้งาน
Preconditions	-
Termination outcome	- ผู้ใช้สามารถดูรายการทวิตที่เกี่ยวกับหนังสือเรื่องต่าง ๆ ได้สำเร็จ
Use case description	ผู้ใช้สามารถดูรายการทวิตที่เกี่ยวกับหนังสือเรื่องต่าง ๆ ผ่านทางแฮชแท็กของทวิตเตอร์ได้
Use case associations	- ดูรายการหนังสือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.14 อธิบายยูสเคส เพิ่มรายการหนังสือที่อยากดู

Use case name	เพิ่มรายการหนังสือที่อยากดู
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถเพิ่มรายการหนังสือที่อยากดูได้สำเร็จ
Use case description	ผู้ใช้สามารถเพิ่มรายการหนังสือที่อยากดูได้ โดยการกดที่ปุ่ม “เพิ่มไว้ในลิสต์” เพื่อทำการเก็บหนังสือที่อยากดูไว้
Use case associations	- ดูรายการหนังสือ

ตารางที่ ก.15 อธิบายยูสเคส ลบรายการหนังสือที่อยากดู

Use case name	ลบรายการหนังสือที่อยากดู
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถลบรายการหนังสือที่อยากดูได้สำเร็จ
Use case description	ผู้ใช้สามารถลบรายการหนังสือที่อยากดูได้ โดยการกดที่ปุ่ม “เพิ่มไว้ในลิสต์” อีกครั้งเพื่อเป็นการลบหนังสือที่อยากดู
Use case associations	- ดูรายการหนังสือ

ตารางที่ ก.16 อธิบายยูสเคส เพิ่มรายการหนังสือที่ชอบ/ไม่ชอบ/เฉย ๆ

Use case name	เพิ่มรายการหนังสือที่ชอบ/ไม่ชอบ/เฉย ๆ
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถเพิ่มรายการหนังสือที่ชอบได้สำเร็จ - ผู้ใช้สามารถเพิ่มรายการหนังสือที่ไม่ชอบได้สำเร็จ - ผู้ใช้สามารถเพิ่มรายการหนังสือที่เฉย ๆ ได้สำเร็จ
Use case description	ผู้ใช้สามารถเพิ่มรายการหนังสือที่ชอบ ไม่ชอบ หรือเฉย ๆ ไปยังรายการที่เก็บไว้ได้
Use case associations	- ดูรายการหนังสือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.17 อธิบายยูสเคส ลบรายการหนังสือที่ชอบ/ไม่ชอบ/เลข ๆ

Use case name	ลบรายการหนังสือที่ชอบ/ไม่ชอบ/เลข ๆ
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถลบรายการหนังสือที่ชอบได้สำเร็จ - ผู้ใช้สามารถลบรายการหนังสือที่ไม่ชอบได้สำเร็จ - ผู้ใช้สามารถลบรายการหนังสือเลข ๆ ได้สำเร็จ
Use case description	ผู้ใช้สามารถลบรายการหนังสือที่ชอบ ไม่ชอบ หรือเลข ๆ ออกจากรายการที่เก็บไว้ได้
Use case associations	- คู่มือการหนังสือ

ตารางที่ ก.18 อธิบายยูสเคส คู่มือการบทความ

Use case name	คู่มือการบทความ
Actors	ผู้ใช้งาน
Preconditions	-
Termination outcome	- ผู้ใช้สามารถคู่มือการบทความทั้งหมดได้สำเร็จ - ผู้ใช้ดูข้อมูลของบทความแต่ละบทความได้สำเร็จ
Use case description	ผู้ใช้คู่มือการบทความความ โดยสามารถดูข้อมูลหรือรายละเอียดของบทความ อีกทั้งยังสามารถดูความคิดเห็นเกี่ยวกับบทความนั้น ๆ หรือค้นหาบทความที่ต้องการได้
Use case associations	- ถูกใจบทความ - บัญชีบทความ - บทความ - แก้ไขบทความ - ค้นหา - ดูความคิดเห็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.19 อธิบายยูสเคส ถูกใจบทความ

Use case name	ถูกใจบทความ
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถกดถูกใจบทความได้สำเร็จ
Use case description	ผู้ใช้สามารถกดถูกใจบทความต่าง ๆ ได้ ซึ่งระบบจะทำการจัดเก็บไว้
Use case associations	- คู่มือการใช้งานบทความ

ตารางที่ ก.20 อธิบายยูสเคส บัญชีบทความ

Use case name	บัญชีบทความ
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถกดบัญชีบทความได้สำเร็จ
Use case description	ผู้ใช้สามารถกดบัญชีบทความต่าง ๆ ได้ โดยระบบจะทำการจัดเก็บไว้เพื่อให้สามารถกดเข้าไปอ่านในวันหลังได้
Use case associations	- คู่มือการใช้งานบทความ

ตารางที่ ก.21 อธิบายยูสเคส แก้ไขบทความ

Use case name	แก้ไขบทความ
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถแก้ไขบทความของตนเองได้สำเร็จ
Use case description	ผู้ใช้สามารถแก้ไขบทความที่ตนเองสร้างไว้ได้
Use case associations	- คู่มือการใช้งานบทความ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.22 อธิบายยูสเคส ลบบทความ

Use case name	ลบบทความ
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถลบบทความของตนเองได้สำเร็จ
Use case description	ผู้ใช้สามารถลบบทความที่ตนเองสร้างไว้ได้
Use case associations	- คู่มือการบทความ

ตารางที่ ก.23 อธิบายยูสเคส คู่มือการรีวิว

Use case name	คู่มือการรีวิว
Actors	ผู้ใช้งาน
Preconditions	-
Termination outcome	- ผู้ใช้คู่มือการรีวิวทั้งหมดได้สำเร็จ - ผู้ใช้คู่มือละเอียดแต่ละรีวิวได้สำเร็จ
Use case description	ผู้ใช้คู่มือการรีวิว โดยสามารถดูรายละเอียดของรีวิว หรือสามารถดูความคิดเห็นของรีวิวนั้น ๆ และยังสามารถทำการค้นหา รีวิวที่ต้องการได้
Use case associations	- ฎกรีวิว - บัญชีการรีวิว - บทความรีวิว - แก้ไขรีวิว - ค้นหา - ดูความคิดเห็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.24 อธิบายยูสเคส ถูกใจรีวิว

Use case name	ถูกใจรีวิว
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถกดถูกใจรีวิวได้สำเร็จ
Use case description	ผู้ใช้สามารถกดถูกใจรีวิวต่าง ๆ ได้
Use case associations	- คู่มือการรีวิว

ตารางที่ ก.25 อธิบายยูสเคส บัญชีรีวิว

Use case name	บัญชีรีวิว
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถกดบัญชีรีวิวได้สำเร็จ
Use case description	ผู้ใช้สามารถกดบัญชีรีวิวต่าง ๆ ได้
Use case associations	- คู่มือการรีวิว

ตารางที่ ก.26 อธิบายยูสเคส แก้ไขรีวิว

Use case name	แก้ไขรีวิว
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถแก้ไขรีวิวของตนเองได้สำเร็จ
Use case description	ผู้ใช้สามารถแก้ไขรีวิวที่ตนเองสร้างไว้ได้
Use case associations	- คู่มือการรีวิว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.27 อธิบายยูสเคส ลบรีวิวก

Use case name	ลบรีวิวก
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถลบรีวิวกของตนเองได้สำเร็จ
Use case description	ผู้ใช้งานสามารถลบรีวิวกที่ตนเองสร้างไว้ได้
Use case associations	- คู่มือการรีวิวก

ตารางที่ ก.28 อธิบายยูสเคส คู่มือการกระดานสนทนา

Use case name	คู่มือการกระดานสนทนา
Actors	ผู้ใช้งาน
Preconditions	-
Termination outcome	- ผู้ใช้คู่มือการกระดานสนทนาทั้งหมดได้สำเร็จ - ผู้ใช้คู่มือละเอียดแต่ละกระดานสนทนาได้สำเร็จ
Use case description	ผู้ใช้งานคู่มือการกระดานสนทนาโดยสามารถดูรายละเอียด หรือดูความคิดเห็นของกระดานสนทนานั้น ๆ และยังสามารถทำการค้นหากระดานสนทนาที่ต้องการได้
Use case associations	- ถูกกระดานสนทนา - บัญชีผู้ใช้กระดานสนทนา - ลบกระดานสนทนา - แก้ไขกระดานสนทนา - ค้นหา - ดูความคิดเห็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.29 อธิบายยูสเคส ถูกใจกระดานสนทนา

Use case name	ถูกใจกระดานสนทนา
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้กดถูกใจกระดานสนทนาได้สำเร็จ
Use case description	ผู้ใช้สามารถกดถูกใจกระดานสนทนาในหัวข้อต่าง ๆ ได้
Use case associations	- คู่มือการใช้งานกระดานสนทนา

ตารางที่ ก.30 อธิบายยูสเคส บัญชีกระดานสนทนา

Use case name	บัญชีกระดานสนทนา
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถกดบัญชีกระดานสนทนาได้สำเร็จ
Use case description	ผู้ใช้สามารถกดบัญชีกระดานสนทนาต่าง ๆ ได้
Use case associations	- คู่มือการใช้งานกระดานสนทนา

ตารางที่ ก.31 อธิบายยูสเคส แก้ไขกระดานสนทนา

Use case name	แก้ไขกระดานสนทนา
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถแก้ไขกระดานสนทนาของตนเองได้สำเร็จ
Use case description	ผู้ใช้สามารถแก้ไขกระดานสนทนาที่ตนเองตั้งถามไว้ได้
Use case associations	- คู่มือการใช้งานกระดานสนทนา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.32 อธิบายยูสเคส ลบกระดานสนทนา

Use case name	ลบกระดานสนทนา
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สามารถลบกระดานสนทนาของตนเองได้สำเร็จ
Use case description	ผู้ใช้สามารถลบกระดานสนทนาที่ตนเองตั้งถามไว้ได้
Use case associations	- คู่มือการกระดานสนทนา

ตารางที่ ก.33 อธิบายยูสเคส สร้างบทความ/รีวิว/กระดานสนทนา

Use case name	สร้างบทความ/รีวิว/กระดานสนทนา
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้สร้างบทความได้สำเร็จ - ผู้ใช้สร้างบทความไม่สำเร็จ เนื่องจากข้อความมีความยาวเกินที่กำหนด - ผู้ใช้สร้างรีวิวได้สำเร็จ - ผู้ใช้สร้างรีวิวไม่สำเร็จ เนื่องจากข้อความมีความยาวเกินที่กำหนด - ผู้ใช้สร้างคำถามในกระดานสนทนาได้สำเร็จ - ผู้ใช้สร้างกระดานสนทนาไม่สำเร็จ เนื่องจากข้อความมีความยาวเกินที่กำหนด
Use case description	ผู้ใช้สามารถทำการสร้างบทความ รีวิว หรือตั้งคำถามที่กระดานสนทนาได้
Use case associations	-

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.34 อธิบายยูสเคส ดูความคิดเห็น

Use case name	ดูความคิดเห็น
Actors	ผู้ใช้งาน
Preconditions	-
Termination outcome	- ผู้ใช้ดูความคิดเห็นในหัวข้อต่าง ๆ สำเร็จ
Use case description	ผู้ใช้สามารถดูความคิดเห็น 4 หัวข้อ คือ ความคิดเห็นของรีวิว กระดานสนทนา บทความ หรือหนังสือต่าง ๆ เป็นต้น
Use case associations	<ul style="list-style-type: none"> - ดูรายการกระดานสนทนา - ดูรายการรีวิว - ดูรายการบทความ - ดูรายการหนังสือ - ลบความคิดเห็น - สร้างความคิดเห็น - แก้ไขความคิดเห็น

ตารางที่ ก.35 อธิบายยูสเคส สร้างความคิดเห็น

Use case name	สร้างความคิดเห็น
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้แสดงความคิดเห็นได้สำเร็จ
Use case description	ผู้ใช้สามารถแสดงความคิดเห็นในหัวข้อต่าง ๆ ได้
Use case associations	- ดูความคิดเห็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.36 อธิบายยูสเคส แก้ไขความคิดเห็น

Use case name	แก้ไขความคิดเห็น
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้แก้ไขความคิดเห็นที่ตนเคยสร้างได้สำเร็จ - ผู้ใช้แก้ไขความคิดเห็นที่ตนเคยสร้างไม่สำเร็จ เนื่องจากข้อความมีความยาวเกินที่กำหนด
Use case description	ผู้ใช้สามารถแก้ไขความคิดเห็นที่ตนเองเคยสร้างไว้ได้
Use case associations	- ดูความคิดเห็น

ตารางที่ ก.37 อธิบายยูสเคส ลบความคิดเห็น

Use case name	ลบความคิดเห็น
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้ลบความคิดเห็นที่ตนเคยสร้างได้สำเร็จ
Use case description	ผู้ใช้สามารถแก้ไขความคิดเห็นที่ตนเองเคยสร้างไว้ได้
Use case associations	- ดูความคิดเห็น

ตารางที่ ก.38 อธิบายยูสเคส ค้นหา

Use case name	ค้นหา
Actors	ผู้ใช้งาน
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ใช้ค้นหาหัวข้อต่าง ๆ จากคีย์เวิร์คได้สำเร็จ
Use case description	ผู้ใช้สามารถค้นหาเรื่องในหัวข้อต่าง ๆ โดยใช้คีย์เวิร์คในการค้นหา
Use case associations	- รายการกระดานสนทนา - รายการรีวิว - รายการบทความ - รายการหนังสือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.39 อธิบายยูสเคส จัดการสมาชิก

Use case name	จัดการสมาชิก
Actors	ผู้ดูแล
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ดูแลจัดการผู้ใช้งานแต่ละคนได้สำเร็จ
Use case description	ผู้ดูแลสามารถจัดการผู้ใช้งานแต่ละคนได้
Use case associations	- สร้างผู้ใช้งาน - แก้ไข - ลบผู้ใช้งาน

ตารางที่ ก.40 อธิบายยูสเคส ลบผู้ใช้งาน

Use case name	ลบผู้ใช้งาน
Actors	ผู้ดูแล
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ดูแลลบผู้ใช้งานได้สำเร็จ
Use case description	ผู้ดูแลสามารถลบผู้ใช้งาน โดยเลือกระบบบุคคลได้
Use case associations	- จัดการสมาชิก

ตารางที่ ก.41 อธิบายยูสเคส แก้ไขผู้ใช้งาน

Use case name	แก้ไขผู้ใช้งาน
Actors	ผู้ดูแล
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ดูแลแก้ไขผู้ใช้งานได้สำเร็จ
Use case description	ผู้ดูแลสามารถแก้ไขผู้ใช้งาน โดยเลือกระบบบุคคลได้
Use case associations	- จัดการสมาชิก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.42 อธิบายยูสเคส จัดการข้อความทวิต

Use case name	จัดการข้อความทวิต
Actors	ผู้ดูแล
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ดูแลจัดการข้อความทวิตได้สำเร็จ
Use case description	ผู้ดูแลจัดการข้อความทวิตได้ โดยเลือกจัดการแต่ละทวิตได้ไม่ว่าจะเป็นการเพิ่ม หรือลบข้อความทวิต
Use case associations	- เพิ่มข้อความทวิต - ลบข้อความทวิต

ตารางที่ ก.43 อธิบายยูสเคส เพิ่มข้อความทวิต

Use case name	เพิ่มข้อความทวิต
Actors	ผู้ดูแล
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ดูแลเพิ่มข้อความทวิตได้สำเร็จ
Use case description	ผู้ดูแลสามารถเพิ่มข้อความทวิตได้ โดยทำการดึง hashtag จาก API ที่กำหนดไว้
Use case associations	- จัดการข้อความทวิต

ตารางที่ ก.44 อธิบายยูสเคส แก้ไขข้อความทวิต

Use case name	แก้ไขข้อความทวิต
Actors	ผู้ดูแล
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ดูแลแก้ไขข้อความทวิตได้สำเร็จ
Use case description	ผู้ดูแลสามารถแก้ไขข้อความทวิตได้ (hashtag, movie_id)
Use case associations	- จัดการข้อความทวิต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.45 อธิบายยูสเคส จัดการรายการหนังสือ/รีวิว/บทความ/กระดานสนทนา

Use case name	จัดการรายการหนังสือ/รีวิว/บทความ/กระดานสนทนา
Actors	ผู้ดูแล
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ดูแลจัดการรายการหนังสือได้สำเร็จ - ผู้ดูแลจัดการรายการบทความได้สำเร็จ - ผู้ดูแลจัดการรายการรีวิวได้สำเร็จ - ผู้ดูแลจัดการรายการกระดานสนทนาได้สำเร็จ
Use case description	ผู้ดูแลสามารถจัดการเพิ่ม ลบ หรือแก้ไข หัวข้อทั้ง 4 หัวข้อ คือ หนังสือ รีวิว บทความ หรือกระดานสนทนา
Use case associations	- เพิ่มรายการ - ลบรายการ - แก้ไขรายการ

ตารางที่ ก.46 อธิบายยูสเคส เพิ่มรายการ

Use case name	เพิ่มรายการ
Actors	ผู้ดูแล
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ดูแลเพิ่มรายการหนังสือได้สำเร็จ - ผู้ดูแลเพิ่มรายการบทความได้สำเร็จ - ผู้ดูแลเพิ่มรายการรีวิวได้สำเร็จ - ผู้ดูแลเพิ่มรายการกระดานสนทนาได้สำเร็จ
Use case description	ผู้ดูแลสามารถเพิ่มรายการหนังสือ บทความ รีวิว หรือ กระดานสนทนาได้
Use case associations	- จัดการรายการหนังสือ/รีวิว/บทความ/กระดานสนทนา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก.47 อธิบายยูสเคส แก้ไขรายการ

Use case name	แก้ไขรายการ
Actors	ผู้ดูแล
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ดูแลแก้ไขรายการหนึ่งได้สำเร็จ - ผู้ดูแลแก้ไขรายการบทความได้สำเร็จ - ผู้ดูแลแก้ไขรายการรีวิวได้สำเร็จ - ผู้ดูแลแก้ไขรายการกระดานสนทนาได้สำเร็จ
Use case description	ผู้ดูแลสามารถแก้ไขรายการหนึ่ง บทความ รีวิว หรือ กระดานสนทนาได้
Use case associations	- จัดการรายการหนึ่ง/รีวิว/บทความ/กระดานสนทนา

ตารางที่ ก.48 อธิบายยูสเคส ลบรายการ

Use case name	ลบรายการ
Actors	ผู้ดูแล
Preconditions	- เข้าสู่ระบบ
Termination outcome	- ผู้ดูแลลบรายการหนึ่งได้สำเร็จ - ผู้ดูแลลบรายการบทความได้สำเร็จ - ผู้ดูแลลบรายการรีวิวได้สำเร็จ - ผู้ดูแลลบรายการกระดานสนทนาได้สำเร็จ
Use case description	ผู้ดูแลสามารถลบรายการหนึ่ง บทความ รีวิว หรือ กระดานสนทนาได้
Use case associations	- จัดการรายการหนึ่ง/รีวิว/บทความ/กระดานสนทนา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข.

รายการตารางรายละเอียดฐานข้อมูลบทที่ 3

ตารางที่ ข.1 รายละเอียดตารางผู้ใช้งาน (User)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK	id	INT	ไอดีผู้ใช้งาน
UK	username	VARCHAR (100)	ชื่อผู้ใช้งาน
UK	email	TEXT	อีเมลผู้ใช้งาน
FK	follow_id	INT	ไอดีของการติดตาม
	hash_pw	VARCHAR (100)	รหัสผ่านที่เข้ารหัส
	gender	INT	เพศ
	year_of_birth	SMALLINT	ปีเกิดผู้ใช้งาน
	profile_url	TEXT	URL ของรูปภาพ โปรไฟล์
	is_delete	BOOLEAN	สถานะการลบ
	created_at	TIMESTAMP	เวลาที่สร้างผู้ใช้งาน
	modified_at	TIMESTAMP	เวลาที่มีการแก้ไขล่าสุด

ตารางที่ ข.2 รายละเอียดตารางการอนุญาต (Permission)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK, FK	user_id	INT	ไอดีของผู้ใช้งาน
	role_name	VARCHAR (50)	บทบาทของผู้ใช้งาน
	created_at	TIMESTAMP	เวลาที่สร้าง
	modified_at	TIMESTAMP	เวลาที่มีการแก้ไขล่าสุด

ตารางที่ ข.3 รายละเอียดตารางเอนทิตี (Entity)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK	id	INT	ไอดีของเอนทิตี
	type	VARCHAR (100)	ชนิดของเอนทิตี
	is_delete	BOOLEAN	สถานะการลบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.4 รายละเอียดตารางความชอบผู้ใช้ (User_Entity)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK, FK	user_id	INT	ไอดีของผู้ใช้งาน
PK, FK	entity_id	VARCHAR (100)	ไอดีของเอนทิตี
	user_mood	INT	ความรู้สึกของผู้ใช้งาน
	created_at	TIMESTAMP	เวลาที่สร้างบันทึก
	modified_at	TIMESTAMP	เวลาที่มีการแก้ไขล่าสุด

ตารางที่ ข.5 รายละเอียดตารางการเก็บเอนทิตี (User_Keep_Entity)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK, FK	user_id	INT	ไอดีของผู้ใช้งาน
PK, FK	entity_id	INT	ไอดีของเอนทิตี
	created_at	TIMESTAMP	เวลาที่สร้างบันทึก
	modified_at	TIMESTAMP	เวลาที่มีการแก้ไขล่าสุด

ตารางที่ ข.6 รายละเอียดตารางทวิตเตอร์ (Tweets)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK	id	INT	ไอดีของผู้ใช้งาน
FK	movie_id	INT	ไอดีของหนัง
	hashtag	VARCHAR (30)	แฮชแท็กที่ดึงข้อมูล
UK	tweet_id	VARCHAR (25)	ไอดีของข้อความทวิตเตอร์
	user_id	VARCHAR (25)	ไอดีของผู้ใช้ทวิตเตอร์
	full_text	TEXT	ข้อความจากทวิตเตอร์
	fav_count	INT	จำนวนการกดชอบ
	retweet_count	INT	จำนวนการรีทวีต
	sentiment	VARCHAR (1)	การตีความหมาย
	created_at	TIMESTAMP	เวลาที่สร้างบันทึก
	modified_at	TIMESTAMP	เวลาที่มีการแก้ไขล่าสุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.7 รายละเอียดตารางหนัง (Movies)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK	id	INT	ไอดีของหนัง
FK	entity_id	INT	ไอดีของเอนทิตี
UK	imdb_id	VARCHAR (20)	ไอดีหนังของ IMDB
	title	TEXT	ชื่อหนัง
	title_th	TEXT	ชื่อหนังภาษาไทย
	original_title	TEXT	ชื่อหนังดั้งเดิม
	original_language	TEXT	ภาษาดั้งเดิม
	overview	TEXT	เรื่องย่อหนัง
	overview_th	TEXT	เรื่องย่อหนังภาษาไทย
	budget	BIGINT	งบประมาณการสร้าง
	revenue	BIGINT	รายได้
	runtime	INT	ความยาวหนัง (นาที)
	poster_path	TEXT	URL ของโปสเตอร์หนัง
	backdrop_path	TEXT	URL ของรูปภาพพื้นหลัง
	images	TEXT []	URL ของรูปภาพ
	videos	JSON []	ข้อมูลวิดีโอ
	popularity	REAL	ความนิยม
	homepage	TEXT	เว็บไซต์หลัก
	production_companies	JSON[]	บริษัทที่ผลิต
	release_date	TIMESTAMP	วันที่เข้าฉาย
	similar_movies	JSON[]	หนังที่คล้ายคลึง
	netflix_id	INT	ไอดีของหนังจาก Netflix
	imdb_score	JSON	คะแนนจาก IMDB
	rotten_tomatoes	JSON	คะแนนหนังจากเว็บมะเขือเทศเน่า
	twitter_score	DOUBLE PRECISION	คะแนนจากทวิตเตอร์
	hashtags	TEXT []	แฮชแท็กของหนัง
	created_at	TIMESTAMP	เวลาที่สร้าง
	modified_at	TIMESTAMP	เวลาที่มีการแก้ไขล่าสุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.8 รายละเอียดตารางประเภทหนัง (Genres)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK	id	INT	ไอดีของประเภทหนัง
UK	name_en	VARCHAR (100)	ชื่อประเภทหนังภาษาอังกฤษ
UK	name_th	VARCHAR (100)	ชื่อประเภทหนังภาษาไทย

ตารางที่ ข.9 รายละเอียดตารางประเภทของหนัง (Movie_Genres)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK, FK	movie_id	INT	ไอดีของหนัง
PK, FK	genre_id	VARCHAR (100)	ไอดีของประเภทหนัง

ตารางที่ ข.10 รายละเอียดตารางคำสำคัญ (Plot_Keywords)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK	id	INT	ไอดีของคำสำคัญ
UK	name_en	VARCHAR (100)	ชื่อคำสำคัญภาษาอังกฤษ
UK	name_th	VARCHAR (100)	ชื่อคำสำคัญภาษาไทย

ตารางที่ ข.11 รายละเอียดตารางคำสำคัญของหนัง (Movie_Plot_Keys)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK, FK	movie_id	INT	ไอดีของหนัง
PK, FK	plot_key_id	INT	ไอดีของคำสำคัญ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.12 รายละเอียดตารางทีมงาน (People)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK	id	INT	ไอดีของคน
UK	imdb_id	VARCHAR (20)	ไอดีของ IMDB
	name	VARCHAR (100)	ชื่อ
	birthday	TIMESTAMP	วันเกิด
	deathday	TIMESTAMP	วันเสียชีวิต
	gender	INT	เพศ
	known_for_department	VARCHAR (20)	บทบาทที่เป็นที่รู้จัก
	also_known_as	TEXT []	ชื่อในภาษาอื่น
	biography	TEXT	ชีวประวัติ
	profile_path	TEXT	URL รูปภาพโปรไฟล์
	movie_credits	JSON[]	หนังที่ได้รับเครดิต
	created_at	TIMESTAMP	เวลาที่สร้าง
	modified_at	TIMESTAMP	เวลาที่มีการแก้ไขล่าสุด

ตารางที่ ข.13 รายละเอียดตารางเครดิต (Credits)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK	id	TEXT	ไอดีของเครดิต
FK	people_id	INT	ไอดีของคน
FK	movie_id	INT	ไอดีของหนัง
	job	TEXT	หน้าที่
	order	INT	ลำดับของเครดิต
	character	TEXT	บทบาทที่แสดง
	created_at	TIMESTAMP	เวลาที่สร้าง
	modified_at	TIMESTAMP	เวลาที่มีการแก้ไขล่าสุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.14 รายละเอียดตารางรีวิว (Reviews)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK, FK	entity_id	INT	ไอดีของเอนทิตี
FK	user_id	INT	ไอดีของผู้ใช้งาน
	title	VARCHAR (120)	หัวข้อ
	content	VARCHAR (10000)	เนื้อหา
	is_spoil	BOOLEAN	การสปอยล์เนื้อหาสำคัญ
	rating	SMALLINT	คะแนนรีวิว
	view	INT	ยอดผู้เข้าชม
	created_at	TIMESTAMP	เวลาที่สร้างรีวิว
	modified_at	TIMESTAMP	เวลาที่มีการแก้ไขล่าสุด

ตารางที่ ข.15 รายละเอียดตารางบทความ (Articles)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK, FK	entity_id	INT	ไอดีของเอนทิตี
FK	user_id	INT	ไอดีของผู้ใช้งาน
	title	VARCHAR (120)	หัวข้อ
	content	VARCHAR (10000)	เนื้อหา
	is_spoil	BOOLEAN	การสปอยล์เนื้อหาสำคัญ
	view	INT	ยอดผู้เข้าชม
	created_at	TIMESTAMP	เวลาที่สร้างบทความ
	modified_at	TIMESTAMP	เวลาที่มีการแก้ไขล่าสุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.16 รายละเอียดตารางการสนทนา (Discussions)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK, FK	entity_id	INT	ไอดีของเอนทิตี
FK	user_id	INT	ไอดีของผู้ใช้งาน
	title	VARCHAR (120)	หัวข้อ
	content	VARCHAR (10000)	เนื้อหา
	is_spoil	BOOLEAN	การสปอยล์เนื้อหาสำคัญ
	view	INT	ยอดผู้เข้าชม
	created_at	TIMESTAMP	เวลาที่สร้างบทสนทนา
	modified_at	TIMESTAMP	เวลาที่มีการแก้ไขล่าสุด

ตารางที่ ข.17 รายละเอียดตารางความคิดเห็น (Comments)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK	id	INT	ไอดีของความคิดเห็น
FK	entity_id	INT	ไอดีของเอนทิตี
FK	user_id	INT	ไอดีของผู้ใช้งาน
	text	VARCHAR (1000)	ข้อความ
	user_mood	INT	ความรู้สึกของผู้ใช้งาน
	is_spoil	BOOLEAN	การสปอยล์เนื้อหาสำคัญ
	created_at	TIMESTAMP	เวลาที่สร้างความคิดเห็น
	modified_at	TIMESTAMP	เวลาที่มีการแก้ไขล่าสุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ข.18 รายละเอียดตารางแท็ก (Tags)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK	Id	INT	ไอดีของแท็ก
FK	follow_id	INT	ไอดีของการติดตาม
UK	name_en	VARCHAR (100)	ชื่อแท็กภาษาอังกฤษ
	name_th	VARCHAR (100)	ชื่อแท็กภาษาไทย
FK	movie_id	INT	ไอดีของหนัง

ตารางที่ ข.19 รายละเอียดตารางการแท็ก (Tag_Entity)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK, FK	tag_id	INT	ไอดีของแท็ก
PK, FK	entity_id	INT	ไอดีของเอนทิตี

ตารางที่ ข.20 รายละเอียดตารางการติดตาม (Followable)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK,	id	INT	ไอดีของการติดตาม
	type	INT	ชนิดของสิ่งที่ติดตาม

ตารางที่ ข.21 รายละเอียดตารางผู้ติดตาม (Follower)

คีย์	ชื่อแอตทริบิวต์	ชนิดตัวแปร	คำอธิบาย
PK, FK	user_id	INT	ไอดีของผู้ใช้งาน
PK, FK	follow_id	INT	ไอดีของการติดตาม
	follow_at	TIMESTAMP	เวลาที่เริ่มการติดตาม
	unfollow_at	TIMESTAMP	เวลาขกเลิกการติดตาม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้