การจำแนกอารมณ์จากเสียงพูดภาษาไทยโดยใช้การเรียนรู้เชิงลึก

CLASSIFYING EMOTION OF THAI-HUMAN VOICE BY USING DEEP LEARNING

ปิยพัทธ์ บุญแถม

วรนิพิฐ ชัยพินิจนรชาติ

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2565

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจำแนกอารมณ์จากเสียงพูดภาษาไทยโดยใช้การเรียนรู้เชิงลึก

CLASSIFYING EMOTION OF THAI-HUMAN VOICE BY USING DEEP LEARNING

ปิยพัทธ์ บุญแถม

วรนิพิฐ ชัยพินิจนรชาติ

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2565

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปี การศึกษา 2565 ภาควิชาวิศวกรรมคอมพิวเตอร์คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การจำแนกอารมณ์จากเสียงพูดภาษาไทยโดยใช้การเรียนรู้เชิงลึก

CLASSIFYING EMOTION OF THAI-HUMAN VOICE BY USING DEEP LEARNING

ผู้จัดทำ

3.1 นายปิยพัทธ์ บุญแถม รหัสนักศึกษา 62010558

3.2 นายวรนิพิฐ ชัยพินิจนรชาติ รหัสนักศึกษา 62010794

_____อาจารย์ที่ปรึกษา

(รศ.ดร.เกียรติกูล เจียรนัยธนะกิจ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจำแนกอารมณ์จากเสียงพูดภาษาไทยโดยใช้การเรียนรู้เชิงลึก

นายปิยพัทธ์ บุญแถม 62010558
นายวรนิพิฐ ชัยพินิจนรชาติ 62010794
รศ.ดร.เกียรติกูล เจียรนัยธนะกิจ อาจารย์ที่ปรึกษา
ปีการศึกษา 2565

บทคัดย่อ

ปริญญานิพนธ์ฉบับนี้ นำเสนอแอปพลิเคชันบนเว็บไซต์สำหรับจำแนกอารมณ์จากเสียงพูดภาษาไทยโดยใช้การเรียนรู้เชิงลึกเพื่อหาวิธีการที่จะแยกแยะอารมณ์ของลูกค้า หรือ ผู้ใช้เพื่อนำไปใช้ในการตัดสินใจต่างๆหรือไปใช้ประโยชน์ตามความต้องการอื่นๆได้ โดยแอปพลิเคชันจะทำให้แยกแยะหรือจำแนกอารมณ์โดยที่ผู้ใช้อัปโหลดเสียงหรือบันทึกเสียงผ่านเว็บไซต์ นำเสียงที่ได้ไปประมวลผลด้วยโมเดลการเรียนรู้เชิงลึกที่ได้มีการออกแบบสถาปัตยกรรมโมเดลและฝึกฝนโมเดลให้สามารถจดจำและจำแนกอารมณ์ต่างๆได้ทั้ง 5 อารมณ์ ได้แก่ ปกติ โกรธ สุข เศร้า และหงุดหงิด ได้ แล้งจึงแสดงผลว่าช่วงเวลาของเสียงที่ผู้ใช้อัปโหลดมาเป็นอารมณ์ใดบ้าง เพื่อนำไปใช้ประโยชน์ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

CLASSIFYING EMOTION OF THAI-HUMAN VOICE BY USING DEEP LEARNING

Mr. Piyaphat	Boontham	62010558
Mr. Voranipit	Chaipinitnorachart	62010794
Assoc. Prof. Dr.Kietikul	Jearanaitanakij	Advisor

Academic Year 2022

Abstract

This thesis presents a web-based application for classifying emotions from Thai speech using deep learning to find ways to differentiate customers' or users' emotions for use in various decision-making or other purposes. The application will classify or differentiate emotions by allowing users to upload or record sounds through the website. The sound will be processed using deep learning models that have been designed and trained to memorize and differentiate various emotions including normal, anger, happiness, sadness, and frustration. Thus, the application will show which emotions the uploaded sound represents and can be used for various purposes.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปริญญานิพนธ์เล่มนี้สำเร็จได้ด้วยความกรุณาจากอาจารย์ รศ.ดร.เกียรติกูล เจียรนัยธนะกิจ ที่ให้ความช่วยเหลือ ให้คำชี้แนะ ช่วยแก้ปัญหา ตลอดจนให้ความรู้และประสบการณ์ที่ดีแก่คณะผู้จัดทำ

ขอขอบคุณ อ.สรยุทธ กลมกล่อม และ ผศ.ธนา หงษ์สุวรรณ อาจารย์ผู้สอนวิชาการเตรียมโครงงานวิศวกรรมคอมพิวเตอร์ วิชาโครงการและวิชาการพัฒนาทักษะทางวิชาชีพวิศวกรรมคอมพิวเตอร์ ที่ได้กรุณาให้คำแนะนำ ตลอดจนข้อชี้แนะ จนในที่สุดทำให้ปริญญานิพนธ์ฉบับนี้สำเร็จลงได้

สุดท้ายต้องขอขอบคุณ บริษัท แอดวานซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน) ได้ทำการบริจาคชุดข้อมูลนี้ให้แก่สาธารณะ ซึ่งเป็นแหล่ง Dataset ที่ได้ใช้ในโครงงานนี้

ปิยพัทธ์ บุญแถม
วรนิพิฐ ชัยพินิจนรชาติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

การจำแนกอารมณ์จากเสียงพูดภาษาไทยโดยใช้การเรียนรู้เชิงลึก I

CLASSIFYING EMOTION OF THAI-HUMAN VOICE BY USING DEEP LEARNING..... II

กิตติกรรมประกาศ..... III

สารบัญ IV

สารบัญตาราง..... VI

สารบัญรูป VII

บทที่ 1..... 1

1.1 ความเป็นมาของปัญหา..... 1

1.2 วัตถุประสงค์โครงงาน 3

1.3 ขอบเขตการวิจัย..... 3

1.4 ประโยชน์ที่ได้รับ..... 4

1.5 แผนการดำเนินงาน..... 5

บทที่ 2..... 7

2.1 ทฤษฎีเกี่ยวข้อง..... 7

2.2 งานวิจัยที่เกี่ยวข้อง..... 11

บทที่ 3..... 14

3.1 Project Scope 14

3.2 ภาพรวมของระบบ 17

3.3 การออกแบบ Use case Diagram 18

3.4 Requirements ของระบบ..... 21

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

3.5 การออกแบบ Business Process Model and Notation 23

3.6 การออกแบบฐานข้อมูล..... 24

3.7 การออกแบบส่วนติดต่อผู้ใช้..... 26

3.8 การออกแบบโมเดลด้วยการเรียนรู้เชิงลึก..... 33

บทที่ 4..... 36

4.1 ผลการพัฒนาแอปพลิเคชัน..... 36

4.2 ผลการจำแนกอารมณ์จากเสียงของโมเดล 47

บทที่ 5..... 46

5.1 สรุปผลที่ได้จากการทำโครงงาน 46

5.2 ปัญหา อุปสรรค และแนวทางแก้ไข 46

5.3 แนวทางการพัฒนาต่อ..... 49

บรรณานุกรม..... 50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตาราง 1.1 แผนการดำเนินงานในช่วงเทอม 1 ปีการศึกษา 2565 5

ตาราง 1.2 แผนการดำเนินงานในช่วงเทอม 2 ปีการศึกษา 2565 6

ตาราง 3.1 Project Scope..... 14

ตาราง 3.3 Use Case Register 19

ตาราง 3.4 Use Case Login 19

ตาราง 3.5 Use Case Dashboard..... 19

ตาราง 3.6 Use Case History 19

ตาราง 3.7 Use Case Show Emotion From Sound 20

ตาราง 3.8 Use Case Result Display 20

ตาราง 3.9 Use Case Manual 20

ตาราง 3.10 Use Case Related paper..... 20

ตาราง 3.11 Use Case Tools 20

ตาราง 3.7 Requirements ของระบบ..... 21

ตาราง 3.8 Users Entity 25

ตาราง 3.9 Users Entity 25

ตาราง 3.10 ภาพรวมหน้าหลักการใช้งาน (User Interface Overview) 26

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

หน้า

รูป 2.1 ข้อมูลที่ซ้อนกันหลายชั้น โครงข่ายแสดงผลข้อมูลสิ่งที่เห็นถูกรวบรวมโดยการเรียนรู้เชิงลึก 7

รูป 2.2 การทำงานของโมเดลจากโมเดลที่มาจากการแก้ปัญหาเชิงรูปภาพ..... 11

รูป 2.3 Spectrogram..... 12

รูป 2.4 การทำงานของโมเดล ML 13

รูป 3.1 สถาปัตยกรรมของระบบในรูปแบบ 3-Tire..... 17

รูป 3.2 Use Case Diagram ของ Web Application 18

รูป 3.3 BPMN ของระบบ 23

รูป 3.4 Entity Relationship Diagram ของระบบ..... 24

รูป 3.5 Entity Database Schema ของระบบ..... 24

รูป 3.6 ส่วนหน้าหลัก 27

รูป 3.7 ส่วนหน้าสมัครสมาชิก..... 28

รูป 3.8 ส่วนหน้าลงชื่อเข้าใช้..... 28

รูป 3.9 ส่วนหน้าแดชบอร์ด 29

รูป 3.10 ส่วนหน้าแสดงประวัติการใช้งานของผู้ใช้..... 30

รูป 3.11 ส่วนหน้าจำแนกอารมณ์..... 30

รูป 3.12 ส่วนหน้าผลลัพธ์..... 31

รูป 3.13 ส่วนหน้าวิธีใช้งาน..... 31

รูป 3.14 ส่วนหน้าเอกสารที่เกี่ยวข้อง..... 32

รูป 3.15 ส่วนหน้าเครื่องมือ..... 32

รูป 3.10 EffNetB0 model Architecture..... 33

รูปที่ 4.1 ส่วนหน้าหลัก..... 39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

หน้า

รูปที่ 4.2 ส่วนของ Navbar..... 40
รูปที่ 4.3 ส่วนของ Info..... 40
รูปที่ 4.4 ส่วนของบริการ..... 41
รูปที่ 4.5 ส่วนหน้าสมัครสมาชิก..... 41
รูปที่ 4.6 ส่วนหน้าลงชื่อเข้าใช้..... 42
รูปที่ 4.7 ส่วนหน้าแดชบอร์ด..... 43
รูปที่ 4.8 ส่วนหน้าแสดงประวัติการใช้งาน..... 44
รูปที่ 4.9 ส่วนหน้าจำแนกอารมณ์..... 44
รูปที่ 4.10 ส่วนหน้าผลลัพธ์..... 45
รูปที่ 4.11 ส่วนหน้าวิธีใช้งาน..... 45
รูปที่ 4.12 ส่วนหน้าเอกสารที่เกี่ยวข้อง..... 46
รูปที่ 4.13 ส่วนหน้าเครื่องมือ..... 46
รูปที่ 4.14 Confusion Metrics..... 47

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหา

การทำธุรกิจโดยใช้ Data Driven เป็นปัจจัยสำคัญที่จะบอกได้ว่าธุรกิจนั้นจะประสบความสำเร็จหรือล้มเหลว ดังนั้นข้อมูลที่นำมาใช้จึงมีความสำคัญอย่างมาก

เนื่องจากการตัดสินใจของลูกค้า (Customer) ในหลายๆอย่าง ในมุมของจิตวิทยาเชิงธุรกิจ เกิดจากการใช้อารมณ์ (Emotion) มากกว่าเหตุผล (Logic) แต่อารมณ์เป็นสิ่งที่ไม่สามารถวัดค่าออกมาเป็นตัวเลขได้

ดังนั้นเราจึงจะหาวิธีการที่จะแยกแยะและบันทึกอารมณ์ของ Customer เผื่อนำไปใช้ในการตัดสินใจต่างหรือ ไปใช้ประโยชน์ในรูปแบบอื่นๆต่อไป ซึ่งเป็นที่มาของการทำโครงงานการจำแนกอารมณ์จากเสียง (Speech Emotion Recognition : SER)

1.1.1 ตัวอย่างการใช้งาน

1) การใช้งานในแผนก Call Center ทำให้สามารถรู้ได้ว่าคนที่โทรมามีรู้สึกอย่างไรจากเรื่องที่กำลังสนทนาอยู่ และ Record ไว้ เพื่อไปใช้ต่อไป
2) การบันทึกเสียงของผู้ให้บริการต่างๆ เช่น Receptionist พนักงานเสิร์ฟอาหาร สามารถทำให้เจ้าของธุรกิจ สามารถรับรู้ได้ว่า พนักงานของบริษัท มีการใช้น้ำเสียงอย่างไรต่อลูกค้า และส่งผลอย่างไรต่อยอดขาย ซึ่งสามารถใช้เป็น Key Index Performance (KPI) ในการประเมิณพนักงานได้

1.1.2 การเรียนรู้เชิงลึก

ในปัจจุบันเทคโนโลยีในกลุ่มปัญญาประดิษฐ์ได้กลายเป็นเทคโนโลยีที่สามารถพบเห็นได้ในชีวิตประจำวัน และมีการพัฒนาให้ก้าวหน้าอยู่ตลอดเวลา การเรียนรู้เชิงลึกก็เช่นกัน การเรียนรู้เชิงลึกคือการจำลองการประมวลผลของสมองมนุษย์ที่มีการประมวลผลในรูปแบบของ Network ที่มีการทำงานของเซลล์ประสาทต่าง ๆ ในสมองเชื่อมต่อกันจนเป็นโครงข่าย เมื่อได้รับข้อมูล หรือ input เข้ามาแล้ว โครงข่ายนี้ก็จะแบ่งข้อมูลและแยกข้อมูลออกเป็นรายละเอียดต่าง ๆ แล้วจึงนำมาประมวลผล หาจุดที่โดดเด่นของข้อมูลและจุดที่แตกต่างกันของข้อมูล เหมือนเป็นการกรองข้อมูลทีละชั้น ๆ สุดท้ายจึงสรุปผล output ว่าข้อมูลนั้นเป็นอย่างไร และมันเป็นคำตอบที่ถูกหรือไม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.1.3 โครงงาน “การจำแนกอารมณ์จากเสียงพูดภาษาไทยโดยใช้การเรียนรู้เชิงลึก”

โครงงาน “การจำแนกอารมณ์จากเสียงพูดภาษาไทยโดยใช้การเรียนรู้เชิงลึก (CLASSIFYING EMOTION OF THAI-HUMAN VOICE BY USING DEEP LEARNING)” นี้จัดขึ้นเพื่อให้ผู้ใช้งานสามารถนำโมเดลการเรียนรู้เชิงลึกที่ได้พัฒนาขึ้นมานี้ไปใช้งานได้ง่ายและสะดวกขึ้น โดยผู้ใช้งานสามารถส่งไฟล์เสียงเข้ามาในระบบ หลังจากนั้นระบบจะนำไฟล์เสียงที่ส่งเข้ามาไปวิเคราะห์และแยกแยะอารมณ์เหล่านั้นด้วยการใช้โมเดลการเรียนรู้เชิงลึก และนำผลลัพธ์ที่ได้กลับมาแสดงให้ผู้ใช้งานได้เห็นผลลัพธ์ที่แตกต่างกันไปตามอารมณ์ที่ถูกจัดแบ่งไว้ ได้แก่ ปกติ โกรธ สุข เศร้า และ หงุดหงิด ซึ่งอาจมีประโยชน์ในการพัฒนาและใช้งานในหลายสาขาอุตสาหกรรม เช่น สื่อสาร สังคมออนไลน์ การแปลภาษาเสียง หรือการดูแลสุขภาพในด้าน จิตใจ เป็นต้น โดยจัดทำเป็น Web -Application ให้สะดวกต่อการใช้งาน ใช้ง่าย เข้าถึงง่าย และเข้าใจง่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.2 วัตถุประสงค์โครงงาน

1) เพื่อพัฒนา Web Application สำหรับจำแนกอารมณ์จากเสียงพูด
2) เพื่อพัฒนา Speech Emotion Recognition ให้เหมาะสมกับภาษาไทยมากขึ้น
3) เพื่อทำการ Classification อารมณ์ เพื่อนำข้อมูลไปสร้างมูลค่าเชิงธุรกิจต่อไป
4) เพื่อศึกษากระบวนการการทำวิจัยอย่างเป็นระบบ

1.3 ขอบเขตการวิจัย

ขอบเขตการวิจัยในโครงงานนี้ถูกแบ่งเป็น 3 ด้าน ได้แก่ ด้านเนื้อหา ด้านแหล่งข้อมูล และด้านระยะเวลาดังนี้

1.3.1 ขอบเขตด้านเนื้อหา

โครงงานนี้มีวัตถุประสงค์เพื่อศึกษาการจำแนกอารมณ์จากเสียงดังนั้นเราจึงหาวิธี การที่จะแยกแยะและบันทึกอารมณ์ของลูกค้า เพื่อนำไปใช้ในการตัดสินใจหรือไปใช้ประโยชน์ในรูปแบบอื่น

1) การเรียนรู้เชิงลึก (Deep Learning) หมายถึง วิธีการเรียนรู้แบบอัตโนมัติด้วยการเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ (Neurons) โดยนำระบบโครงข่ายประสาท (Neural Network) มาซ้อนกันหลายชั้น (Layer) และทำการเรียนรู้ข้อมูลตัวอย่าง ซึ่งข้อมูล ดังกล่าวจะถูกนำไปใช้ในการตรวจจับรูปแบบ (Pattern) หรือจัด หมวดหมู่ข้อมูล (Classify the Data)
2) Transfer Learning หมายถึง เทคนิคที่ช่วยลดเวลาการเทรนโมเดล Deep Learning ด้วยการนำบางส่วนของโมเดลที่เทรนเรียบร้อยแล้ว กับงานที่ใกล้เคียงกัน มาใช้เป็นส่วนหนึ่งของโมเดลใหม่ ในทางปฏิบัติมีคนจำนวนน้อยมากที่เทรน Convolutional Neural Network ตั้งแต่ต้น เนื่องจากไม่มีชุดข้อมูล Dataset ที่ใหญ่พอ ดังนั้นคนส่วนใหญ่จึงใช้วิธีนำโมเดล CNN ที่เทรนกับชุดข้อมูล Dataset ขนาดใหญ่ เรียกว่า Pre-trained model นำโมเดลนั้นมาเป็นโมเดลตั้งต้นเพื่อเทรนต่อ กับ Dataset ขนาดเล็กในงานเฉพาะทาง หรือ ใช้สกัด Feature สำหรับงานที่ต้องการออกมา
3) การประมวลผลภาพ (Image Processing) หมายถึง กระบวนการจัดการและวิเคราะห์รูปภาพให้เป็นข้อมูลในแบบดิจิทัล โดยใช้คอมพิวเตอร์การ เพื่อให้ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลที่เราต้องการทั้งในเชิงคุณภาพและปริมาณ (ขนาด รูปร่าง) หลังจากนั้นเราสามารถนำข้อมูลไปวิเคราะห์ และสร้างเป็นระบบต่อไป

1.3.2 ขอบเขตด้านแหล่งข้อมูล

Machine Learning model ในโครงงานครั้งนี้ จะถูก Build และ Trian บน Google Colab และ AWS Sagemaker ในส่วน API และ Application จะทำงานบน Docker

1.4 ประโยชน์ที่ได้รับ

1.4.1 ได้รับความรู้และประสบการณ์การใช้งานโมเดลการเรียนรู้เชิงลึก

1.4.2 ได้รับความรู้และประสบการณ์ในการพัฒนา Web Application

1.4.3 แอปพลิเคชันสามารถเป็นแหล่งศึกษาการประยุกต์ใช้การเรียนรู้เชิงลึกในการแยกแยะเสียงและจำแนกสิ่งต่าง ๆ ได้

1.4.4 การจำแนกอารมณ์จากเสียงทำให้เกิดประโยชน์ต่างๆทางธุรกิจมากมายเช่น

1) ในปัจจุบันไม่มีการเก็บ Feed back เชิงอารมณ์ของ Customer รูปแบบอื่นนอกจากแบบสอบถามซึ่งอาจะทำให้เกิด Bias ดังนั้น โปรเจคของเราสามารถเก็บ Feed back เชิงอารมณ์ของ Customer โดยไม่มี Bias

2) ในปัจจุบัน เจ้าของธุรกิจไม่สามารถประเมิณ KPI ของพนักงานบริการ ในเรื่องของการใช้อารมณ์ คำพูด ในการดูแล Customer ได้ ถ้าเราสามารถประเมิณ KPI จากการใช้น้ำเสียงของพนักงานบริการ เราสามารถดูแล Customer ได้ดียิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5 แผนการดำเนินงาน

แผนการดำเนินงานทั้งสองเทอมดังตารางต่อไปนี้

ตาราง 1.1 แผนการดำเนินงานในช่วงเทอม 1 ปีการศึกษา 2565

ลำดับ	ขั้นตอนการดำเนินงาน	สิงหาคม					กันยายน				ตุลาคม					พฤศจิกายน				
		2	9	16	23	30	6	13	20	27	3	10	17	24	31	1	8	15	22	29
1	รวบรวม data set																			
2	Pre- process: แบ่งเสียงให้เป็นระยะเท่าๆกัน																			
3	Pre – Process :Data augmentation																			
4	Pre-Process: Convert sound to spectrogram																			
5	Testing and Choosing model																			
6	Fine -tuning model																			
7	Front- end website																			

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 1.2 แผนการดำเนินงานในช่วงเทอม 2 ปีการศึกษา 2565

ลำดับ	ขั้นตอนการดำเนินงาน	มกราคม					กุมภาพันธ์				มีนาคม				เมษายน				
		4	11	18	25	31	1	8	15	22	1	8	15	22	5	12	19	26	30
1	Front- End																		
2	ศึกษาเรื่อง docker																		
3	สร้าง data pipeline																		
4	Backend																		
5	Integrate with model																		
6	Unit testing																		
7	Deploy on docker																		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

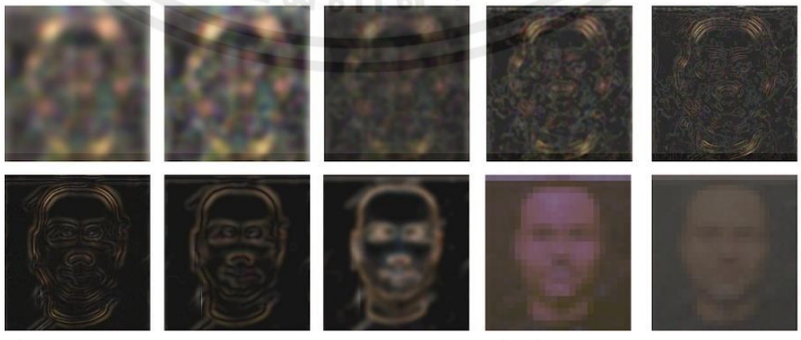
บทที่ 2
ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

ในบทนี้เราจะนำเสนอทฤษฎีของการเรียนรู้เชิงลึก (Deep Learning) หรือ ถ่ายทอดการเรียนรู้ (Transfer learning) และ การประมวลผลภาพ (Image Processing) ตลอดจนถึงงานวิจัยต่างๆ ที่ได้นำ การจำแนกอารมณ์จากเสียงแบบทันทีโดยการใช้ Pre-trained โมเดลที่มาจากการแก้ปัญหาเชิงรูปภาพ : ผลกระทบจากการบีบอัด bandwidth หรือ การจำแนกอารมณ์จากเสียงด้วย CNN และ การจำแนกอารมณ์จากเสียงโดยใช้อัลกอริทึมของ Machine Learning

2.1 ทฤษฎีเกี่ยวข้อง

2.1.1 การเรียนรู้เชิงลึก

การเรียนรู้เชิงลึก (Deep Learning) เป็นการสร้างโมเดลของระบบประมวลผลข้อมูลที่มีความซับซ้อนและลึกลงไปในรายละเอียดของข้อมูล โดยใช้เทคนิคของการเรียนรู้เชิงลึก เช่น การใช้งานโครงข่ายประสาทเทียม (Neural Networks) ที่มีหลายชั้นและมีจำนวนโหนดมากขึ้น เพื่อให้โมเดลสามารถเรียนรู้ลักษณะเฉพาะของข้อมูลและสร้างความเข้าใจที่ลึกซึ้งกับความสัมพันธ์ของข้อมูลในชุดข้อมูล โดยปกติแล้วเราจะใช้เทคนิคการเรียนรู้เชิงลึกในการแก้ไขปัญหาที่มีความซับซ้อน เช่น การจำแนกภาพ การตรวจจับวัตถุ การแปลภาษา และการทำนายผลลัพธ์ต่างๆ ในด้านต่างๆ ของการประมวลผลข้อมูล



รูป 2.1 ข้อมูลที่ซ้อนกันหลายชั้นโครงข่ายแสดงผลข้อมูลสิ่งที่เห็นถูกรวบรวมโดยการเรียนรู้เชิงลึก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 การถ่ายทอดการเรียนรู้

ถ่ายทอดการเรียนรู้ (Transfer Learning) คือการนำโมเดลปัญญาประดิษฐ์ที่เรียนรู้จากข้อมูลใดข้อมูลหนึ่ง มาใช้ในการแก้ไขปัญหาหรืองานที่เกี่ยวข้องกับข้อมูลอื่น ๆ โดยไม่จำเป็นต้องสร้างโมเดลใหม่เต็มรูปแบบทั้งหมด

การใช้โมเดลปัญญาประดิษฐ์ในการถ่ายทอดการเรียนรู้สามารถช่วยลดเวลาและความล่าช้าในการสร้างโมเดลใหม่ โดยโมเดลที่ได้รับการฝึกฝนแล้วสามารถเรียนรู้และสืบทอดความรู้ได้จากข้อมูลที่มีอยู่แล้ว และนำมาใช้กับงานที่มีความซับซ้อนต่างๆ เช่น การจำแนกภาพ การประมวลผลภาษาธรรมชาติ และการแยกแยะเสียง เป็นต้น

โดยทั่วไปแล้ว กระบวนการถ่ายทอดการเรียนรู้จะมีขั้นตอนการทำงานดังนี้

1) เลือกโมเดลปัญญาประดิษฐ์ที่ฝึกฝนแล้วด้วยข้อมูลที่เกี่ยวข้องกับงานที่ต้องการใช้
2) นำโมเดลปัญญาประดิษฐ์ที่ฝึกฝนแล้วมาปรับปรุงหรือฝึกฝนเพิ่มเติมด้วยข้อมูลใหม่ที่เกี่ยวข้องกับงานที่ต้องการใช้
3) นำโมเดลปัญญาประดิษฐ์ที่ฝึกฝนแล้วมาใช้ในงานที่ต้องการใช้

การใช้โมเดลปัญญาประดิษฐ์ที่เคยถูกฝึกฝนด้วยข้อมูลหนึ่ง ๆ มาใช้กับงานที่เกี่ยวข้องกับข้อมูลอื่น ๆ โดยไม่จำเป็นต้องสร้างโมเดลใหม่เต็มรูปแบบทั้งหมด ทำให้เราสามารถประหยัดเวลาและค่าใช้จ่ายในการฝึกฝนโมเดลใหม่ได้ และยังสามารถเพิ่มประสิทธิภาพการทำงานของโมเดลในงานที่ต้องการใช้ได้อีกด้วย

2.1.3 การประมวลผลภาพ

การประมวลผลภาพ (Image Processing) คือ กระบวนการทางคณิตศาสตร์และวิทยาการคอมพิวเตอร์ที่ใช้เทคโนโลยีและอัลกอริทึมต่าง ๆ เพื่อวิเคราะห์และปรับแต่งภาพดิจิตัล เพื่อให้ได้ภาพที่มีคุณภาพสูง หรือทำการแยกแยะวัตถุภายในภาพเพื่อนำไปใช้ในงานต่าง ๆ ได้อย่างมีประสิทธิภาพ

การประมวลผลภาพมีการใช้งานหลากหลายในชีวิตประจำวัน เช่น ในการปรับแต่งภาพถ่าย ใช้ในอุตสาหกรรมการผลิต เช่น การตรวจสอบคุณภาพสินค้า และใช้ในการวิเคราะห์ภาพทางการแพทย์ เช่น การวินิจฉัยโรคด้วยภาพการรักษาทันตกรรม ฯลฯ

2.1.4 Software Development Process

เป็นกระบวนการพัฒนาซอฟต์แวร์เป็นขั้นตอนการสร้างซอฟต์แวร์ที่เป็นระเบียบเรียบร้อยและมีการกำหนดขั้นตอนอย่างชัดเจนเพื่อให้การพัฒนาซอฟต์แวร์เป็นไปได้อย่างมีประสิทธิภาพ ขั้นตอนสำคัญที่เกี่ยวข้องกับการพัฒนาซอฟต์แวร์ประกอบด้วยการกำหนด Project Scope การรวบรวม Requirement การออกแบบ UX/UI การออกแบบ Database การ implement และการทดสอบระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.5 UX/UI Design

ในการออกแบบ UX/UI จะใช้หลักการการจัดสี เพื่อสร้างความสมดุลของสีและเนื้อหาในหน้าจอ เพื่อให้ผู้ใช้งานไม่รู้สึกเกลียดกับสีหรือเนื้อหาที่ซ้ำซ้อน การจัดตำแหน่งและองค์ประกอบของปุ่มต่าง ๆ ก็จะใช้เพื่อให้ผู้ใช้งานสามารถใช้งานได้อย่างสะดวก โดยจะต้องพิจารณาให้ถูกต้องว่าการจัดตำแหน่งและองค์ประกอบของปุ่มนั้นเหมาะสมกับการใช้งานและสื่อสารของหน้าจอนั้น ๆ อีกทั้งยังต้องคำนึงถึงความปลอดภัยของผู้ใช้งานด้วย เช่น ปุ่มสำคัญควรอยู่ในตำแหน่งที่ผู้ใช้งานสามารถเข้าถึงได้ง่าย และไม่เกี่ยวข้องกับการใช้งานที่สามารถเป็นอันตรายได้ ดังนั้นการออกแบบ UX/UI เป็นขั้นตอนที่สำคัญในการพัฒนาซอฟต์แวร์ เพื่อให้ผู้ใช้งานสามารถใช้งานได้อย่างมีประสิทธิภาพและสะดวกสบาย

2.1.6 Front-end Web-App Development

พัฒนาโดยใช้ React โดย JavaScript เป็นภาษาหลักที่ใช้ในการเขียนโปรแกรมฝั่ง Client ในขณะที่ HTML/CSS จะใช้ในการสร้างโครงสร้างและรูปแบบหน้าเว็บไซต์ นอกจากนี้ยังต้องมีความรู้ในการใช้งาน React รวมถึงเครื่องมืออื่น ๆ ที่ใช้ในการพัฒนาเช่น React Router Axios Redux เป็นต้น ทั้งนี้ยังต้องออกแบบ User Interface (UI) และ User Experience (UX) เพื่อให้ผู้ใช้งานสามารถใช้งานแอปพลิเคชันได้อย่างสะดวกและเข้าใจง่าย

2.1.7 Backend development

2.1.7.1 Fast API

Fast API เป็นเฟรมเวิร์กสำหรับการพัฒนา API ด้วย Python ที่มีความเร็วและประสิทธิภาพสูง และได้รับความนิยมอย่างมากในช่วงไม่กี่ปีที่ผ่านมา เพราะว่า Fast API ให้ประสิทธิภาพการทำงานที่ดีกว่าเฟรมเวิร์กอื่น ๆ และมีความสามารถในการสร้าง API ที่ปลอดภัยมากขึ้นด้วยการใช้งาน Pydantic ในการ validate ข้อมูล และการใช้งาน asynchronous ที่ช่วยเพิ่มประสิทธิภาพการทำงานของ API อีกด้วย นอกจากนี้ Fast API ยังมีความยืดหยุ่นในการใช้งานกับภาษา Python และ library ต่าง ๆ ที่เกี่ยวข้อง เช่น SQL Alchemy Alembic OAuth2 JWT เป็นต้น ทำให้ Fast API เป็นตัวเลือกที่ดีสำหรับผู้พัฒนาที่ต้องการสร้าง API ใน Python อย่างเร็วและประสิทธิภาพสูง

2.1.7.2 TensorFlow

TensorFlow เป็นเครื่องมือ Open Source สำหรับ Machine Learning และ Deep Learning ที่พัฒนาโดย Google Brain Team โดย TensorFlow ช่วยให้ผู้ใช้สามารถสร้างและปรับปรุงโมเดล Machine Learning ได้ง่ายขึ้น โดยมีการเชื่อมต่อกับโปรแกรม Python และสนับสนุนการทำงานบนหลายแพลตฟอร์มเช่น Windows Linux macOS Android iOS และอื่น ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TensorFlow มีฟังก์ชั่นการสร้างและฝึกโมเดล Machine Learning ที่สามารถใช้กับหลายๆ งาน เช่น การจำแนกภาพ (image classification) การตรวจจับวัตถุ (object detection) การแปลภาษา (language translation) และอื่นๆ โดย TensorFlow มี API สำหรับ Python C++ Java Go และ Swift เพื่อให้ผู้ใช้งานสามารถเลือกภาษาโปรแกรมที่ต้องการได้ง่ายขึ้น นอกจากนี้ TensorFlow ยังมี TensorFlow Lite สำหรับการนำโมเดล Machine Learning ไปใช้งานบนอุปกรณ์พกพา เช่น สมาร์ทโฟน แท็บเล็ต หรือ IoT อื่นๆ อีกด้วย

2.1.7.3 Py Torch

Py Torch เป็นเครื่องมือสำหรับ Deep Learning ที่มีความยืดหยุ่นสูง และเป็นได้ทั้งในการสร้างและฝึกโมเดล ซึ่งถูกพัฒนาโดย Facebook AI Research (FAIR) โดยมีการออกแบบให้ใช้งานง่ายและเข้าใจได้ง่าย นอกจากนี้ Py Torch ยังมีความสามารถในการคำนวณแบบ GPU ที่เร็วกว่า NumPy และเป็นฟรีและเปิดโค้ด (open-source) ซึ่งช่วยส่งเสริมให้นักพัฒนาและนักวิจัยสามารถนำไปใช้ได้ง่ายและสามารถปรับแต่งได้ตามความต้องการของโครงการหรืองานวิจัยต่างๆ ทั้งในงาน Computer Vision Natural Language Processing Speech Recognition และอื่นๆ อีกด้วย

2.1.8 Database System

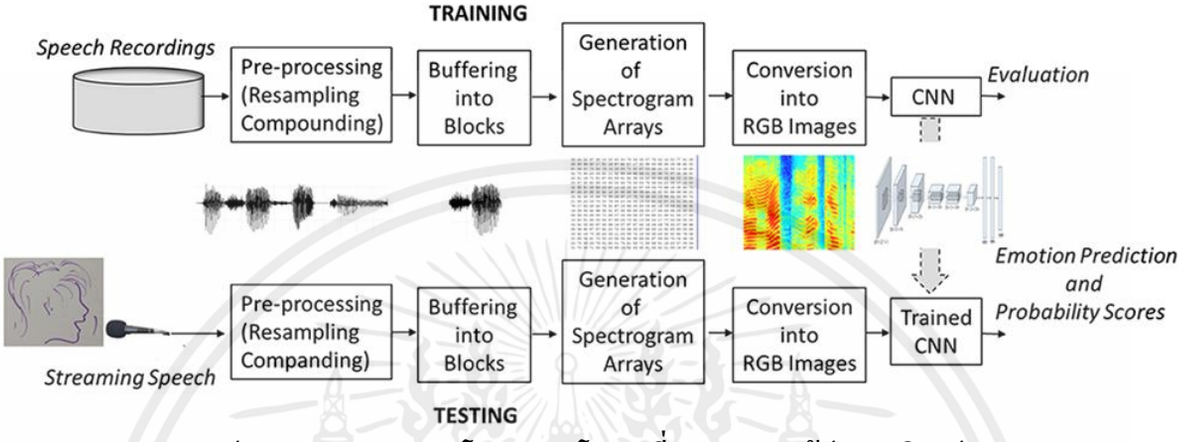
การออกแบบฐานข้อมูล (Database Design) ประกอบด้วยหลายขั้นตอน โดยเริ่มต้นจากการรวบรวมข้อมูลหรือข้อกำหนด (Requirement) ของระบบ จากนั้นจึงเริ่มการออกแบบฐานข้อมูล เช่น การสร้าง Entity Relationship Diagram (ER Diagram) ซึ่งเป็นตัวแสดงแบบกราฟิกของแต่ละ Entity และความสัมพันธ์ระหว่าง Entity ในระบบ จากนั้นจะทำการแปลง ER Diagram เป็น Relational Database Schema หรือโครงสร้างฐานข้อมูลแบบสัมพันธ์ แล้วจึงนำไปสร้างฐานข้อมูลในระบบจัดการฐานข้อมูล (DBMS) เช่น PostgreSQL เป็นต้น

การออกแบบฐานข้อมูลเป็นขั้นตอนสำคัญที่ช่วยให้ระบบฐานข้อมูลสามารถทำงานได้อย่างมีประสิทธิภาพและเป็นระบบได้ ซึ่งต้องพิจารณาปัจจัยต่างๆ เช่น การเลือกใช้โครงสร้างฐานข้อมูลแบบใด การคำนวณความจุของฐานข้อมูล การจัดการความสัมพันธ์ของตาราง และการทำ Index เพื่อเพิ่มประสิทธิภาพในการค้นหาข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 การจำแนกอารมณ์จากเสียงแบบทันทีโดยการใช้ Pre-trained โมเดลที่มาจากการแก้ปัญหาเชิงรูปภาพ : ผลกระทบจากการบีบอัด bandwidth



รูป 2.2 การทำงานของโมเดลจากโมเดลที่มาจากการแก้ปัญหาเชิงรูปภาพ

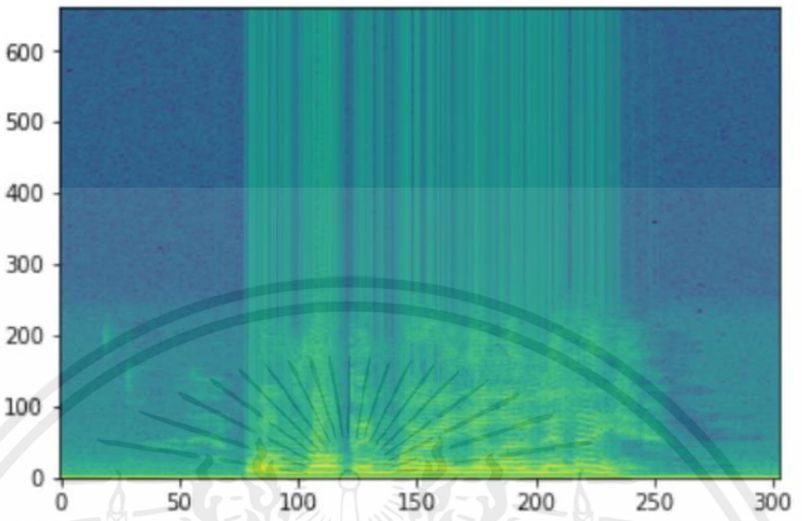
เนื่องจาก Computational Resources มีอยู่อย่างจำกัด และ Database ของตัวอย่างของ SER มีน้อย ดังนั้น เราต้องหาวิธีการการใช้ทรัพยากรณ์ต่างๆ ที่มีอยู่อย่างจำกัด ให้มีประโยชน์สูงสุด เราจึงต้องใช้ Pre-trained model ที่มีอยู่แล้ว แต่ Pre-trained model ที่มีส่วนใหญ่เป็น Image classification ดังนั้น เราจะเปลี่ยนโจทย์ SER จากการใช้เสียงเป็นการใช้ภาพ แทนโดยพยายามแปลงเสียงให้อยู่ในรูปแบบของรูปภาพ (MFCC) จากนั้นเราก็ใช้ pre-trained CNN ของ Image classification มา Fine-tune

การจำแนกอารมณ์แบบทันทีคือการที่ ข้อมูลขาเข้าแบบต่อเนื่อง การประมวลที่รวดเร็ว และ ผลลัพธ์ที่แน่นอน แม่นยำ โดยทำการจำแนกในระยะเวลาที่จำกัดในระดับมิลิเซค

แต่การสอนโมเดลให้เข้าใจใน Training set ไม่จำเป็นต้องทำแบบทันที เมื่อเราทำการสอนโมเดลของเราจนได้โมเดลที่มีผลตามต้องการแล้ว สามารถนำไปใช้ให้โมเดลในการจำแนกแบบทันทีได้ ดังนั้น Training set ของเราสามารถ มีขนาดใหญ่ได้ เพื่อเพิ่มความแม่นยำ แต่ ความเร็วในการจำแนกนั้นขึ้นอยู่กับตัวแปร อื่นๆ อาทิเช่น ความยาวของเสียงที่ต้องการจำแนกอารมณ์ ความแรงของเครื่องประมวลผล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.2 การจำแนกอารมณ์จากเสียงด้วย CNN



รูป 2.3 Spectrogram

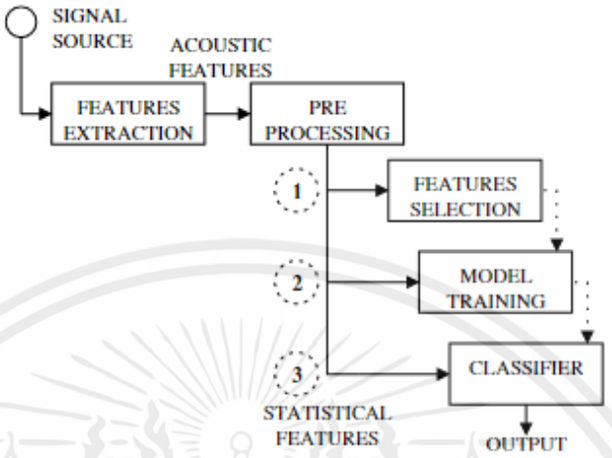
จำแนกอารมณ์จากเสียงโดยใช้ Convolutional Neural Network ที่มี 7 layers เป็น Convolution 6 layers และต่อด้วย Dense layer 1 layers โดยจะใช้ input ของ Model เป็น Spectrogram ที่เกิดจากนำเสียงมาแปลงให้เป็นรูปภาพ เป็นการเปลี่ยนจากปัญหา Audio problems เป็น image classification problem เหมือนกับงานวิจัยที่ 2.1.1

Convolution Layer มีหน้าที่ในการทำ Feature Extraction ทำให้เราสามารถข้ามขั้นตอนการเตรียมข้อมูลหรือ Feature Engineering ทำให้โมเดลสามารถเลือกจุดสำคัญในภาพเองได้โดยที่เราไม่ต้องกำหนดว่าจุดไหนในภาพมีความสำคัญ

Dense Layer มีหน้าที่ในการหา Pattern จากข้อมูลที่มาจาก Spectrogram ที่ผ่านการ Feature Extraction จาก Convolution Layer ว่าจะจำแนกข้อมูลขาเข้าไปเป็นอารมณ์ใด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.3 การจำแนกอารมณ์จากเสียงโดยใช้อัลกอริทึมของ Machine Learning



รูป 2.4 การทำงานของโมเดล ML

งานวิจัยนี้จะต่างจาก2 งานวิจัยแรกเพราะเป็นการใช้ Machine Learning ปกติและ ทำการ Feature Extraction เอาเอง ซึ่ง 2 งานวิจัยแรกนั้นใช้ Deep Learning ทำให้ไม่ต้องทำการ Feature Extraction โดย Feature ที่งานวิจัยนี้ใช้จะมาจากค่าเชิงสถิติต่างของเสียงแต่ละเสียง เช่นความห่างกันของแต่ละแอมพลิจูด และ นำค่าเหล่านี้มา Feature Engineering แล้ว ทดลองใน Machine Learning Algorithms หลายๆ Algorithms แล้วทดสอบว่า Algorithms ไหนให้ผลดีที่สุด

ทำให้การสอนโมเดลแบบนี้ ใช้เวลานานแต่ก็มีข้อดีคือเราสามารถกำหนดจุดที่เราสนใจได้ผ่านการ Feature Engineering แต่ข้อเสียก็คือโมเดลจะมี ความลำเอียงและข้อจำกัดจากมนุษย์ ดังนั้น ทางกลุ่มของเราจึงมองว่าไม่เหมาะสมที่จะเลือกงานวิจัยนี้มาอ้างอิง เป็นอันหลัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การออกแบบและการพัฒนา

ในการออกแบบพัฒนาเว็บไซต์การจำแนกอารมณ์จากเสียงพูดภาษาไทยโดยใช้การเรียนรู้เชิงลึกนี้จะเป็นการพัฒนาเว็บไซต์ขึ้นมาใหม่ให้มีรูปแบบที่ทันสมัย ตอบโจทย์ผู้ใช้งานมากขึ้น เพื่อให้เหมาะกับผู้ใช้ที่ต้องการใช้งานฟังก์ชั่นของตัวเว็บไซต์ให้ได้ประโยชน์มากที่สุด

3.1 Project Scope

Web Application Thai SER นี้ได้ออกแบบขอบเขตการแสดงผลและความสามารถของระบบไว้ดังนี้

ตาราง 3.1 Project Scope

ลำดับ	ขอบเขต	รายละเอียด
1	ส่วนหน้าหลัก	1) มี Navbar มีรายละเอียดดังนี้ - มีแสดงไอคอน Thai SER - สามารถกด หน้าหลัก จำแนกอารมณ์ บริการของเรา เพื่อไป Info นั้นๆ - มีปุ่มให้กดเพื่อเข้าสู่ระบบ และ ลงทะเบียน 2) มี Info เพื่อแสดงรายละเอียดในแต่ละหมวดหมู่ ได้แก่ หน้าหลัก จำแนกอารมณ์ เครื่องมือและ API ในแต่ละ Info จะประกอบด้วย Topline Headline Description Image Button 3) มี Service เพื่อแสดง Card ได้แก่ เครื่องมือ เอกสาร วิธีใช้งาน เมื่อกดเข้าไปที่ Card จะเชื่อมไปยังหน้าครื่องมือ เอกสาร วิธีใช้งาน 4) มี Footer ที่จะมี 2 หมวดหมู่ ได้แก่ - เกี่ยวกับเรา ประกอบด้วย Git Hub CE KMITL - บริการของเรา ประกอบด้วย เครื่องมือ เอกสาร วิธีใช้ สามารถกดแล้วลิงค์ไปยังหน้านั้นได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2	ส่วนหน้าสมัครสมาชิก	1) สามารถกรอกข้อมูลเพื่อลงทะเบียน ได้แก่ - ชื่อผู้ใช้งาน - ชื่อ - อีเมล - อายุ - เพศ - รหัสผ่าน - ยืนยันรหัสผ่าน 2) สามารถกดปุ่มเพื่อลงชื่อเข้าใช้งาน 3) สามารถกด เข้าสู่ระบบเพื่อไปยังหน้าเข้าสู่ระบบ
3	ส่วนหน้าลงชื่อเข้าใช้	1) สามารถกรอกข้อมูลเพื่อเข้าสู่ระบบได้แก่ - ชื่อผู้ใช้งาน - รหัสผ่าน 2) สามารถกดปุ่มเพื่อเข้าสู่ระบบ 3) สามารถกดลงชื่อเข้าใช้งานเพื่อไปยังหน้าเข้าลงชื่อเข้าใช้งาน
4	ส่วนหน้าแดชบอร์ด	1) แสดง Username ของผู้ใช้ 2) แสดง ชื่อ ของผู้ใช้ 3) แสดง อีเมล ของผู้ใช้ 4) แสดง อายุ ของผู้ใช้ 5) แสดง เพศ ของผู้ใช้ 6) สามารถกดปุ่มเพื่อแก้ไขบัญชีได้ 7) สามารถกดปุ่มเพื่อแก้ไขรหัสผ่านได้ 8) สามารถกดปุ่มเพื่อแสดงประวัติการใช้งานของผู้ใช้คนนั้นได้
5	ส่วนหน้าแสดงประวัติการใช้งานของผู้ใช้	1) แสดงประวัติการทำรายการของผู้ใช้ 2) มีปุ่มเพื่อย้อนกลับไปยังหน้าแดชบอร์ด
6	ส่วนหน้าจำแนกอารมณ์	1) สามารถลากหรือกดปุ่มอัปโหลดเสียงได้
7	ส่วนหน้าผลลัพธ์	1) แสดงผลลัพธ์ออกมาเป็นจำนวนอารมณ์ 2) แสดงผลลัพธ์ที่มากที่สุด

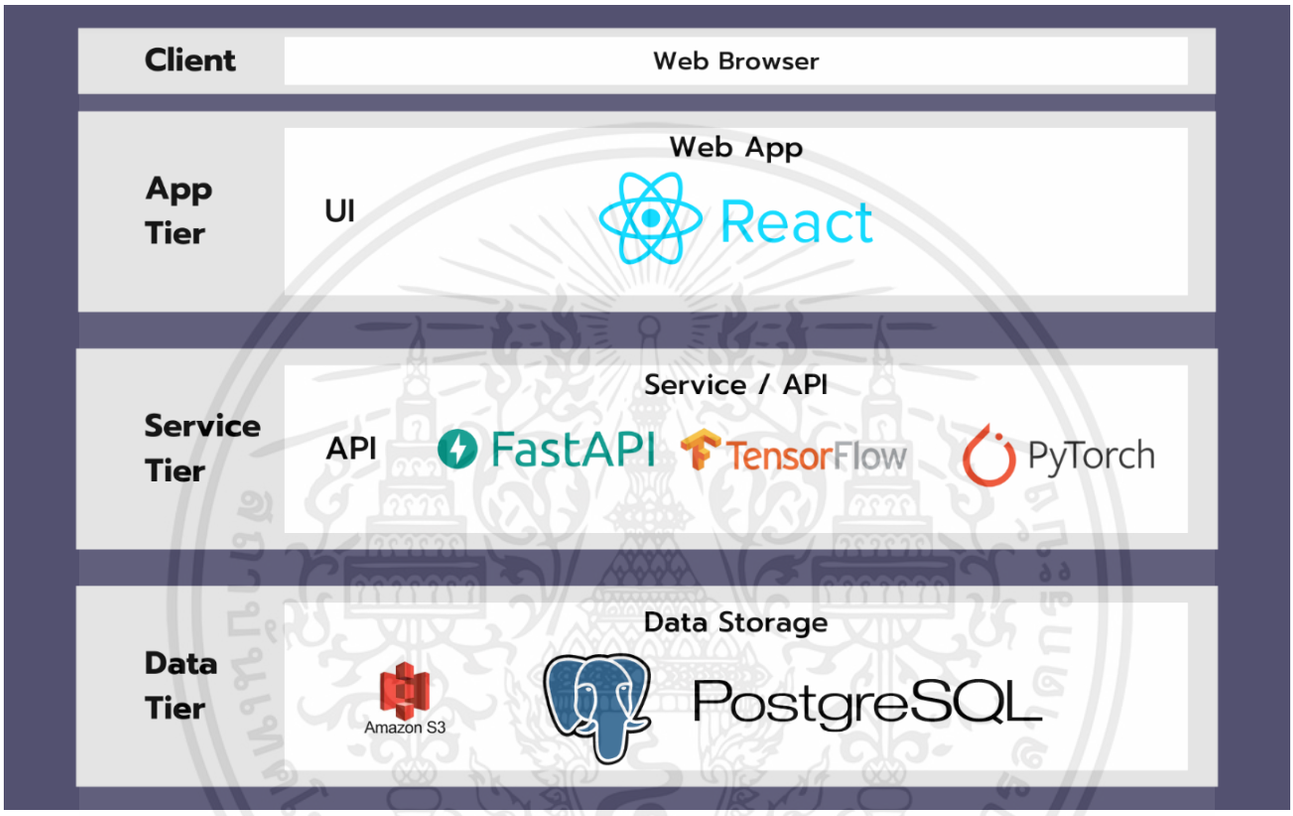
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

		3) สามารถอัปโหลดไฟล์ได้เพิ่มเติมเมื่อกดปุ่ม อัปโหลดเสียง
8	ส่วนหน้าวิธีใช้งาน	1) มี Vertical Navbar สามารถกดปุ่มเพื่อไปยังหมวดต่างๆได้ 2) แสดง Info ในแต่ละ Info จะประกอบด้วย Topline Headline Description Image Button
9	ส่วนหน้าเอกสารที่เกี่ยวข้อง	1) มี Vertical Navbar สามารถกดปุ่มเพื่อไปยังหมวดต่างๆได้ 2) แสดง Info ในแต่ละ Info จะประกอบด้วย Topline Headline Description Image Button
10	ส่วนหน้าเครื่องมือ	1) มี Vertical Navbar สามารถกดปุ่มเพื่อไปยังหมวดต่างๆได้ 2) แสดง Info ในแต่ละ Info จะประกอบด้วย Topline Headline Description Image Button

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 ภาพรวมของระบบ

Web Application จะรับไฟล์ Input จาก User เป็นไฟล์เสียงโดยจะถูก Pre-Process แล้วจะทำการแสดงผล Output เป็นข้อมูลช่วงเวลา และ ผลลัพธ์ของอารมณ์ที่ได้จากการประมวลผลของโมเดลให้ผู้ใช้งานได้เห็น



รูป 3.1 สถาปัตยกรรมของระบบในรูปแบบ 3-Tire

จากรูปที่ 3.1 สถาปัตยกรรมของระบบในรูปแบบ 3-Tier โดย Client จะเป็นส่วนที่สามารถเข้าผ่านทาง Web Browser เท่านั้น

3.1.1 App Tire

เป็นส่วนติดต่อกับผู้ใช้งานหรือส่วนแสดงผล ในส่วนนี้จะใช้ React Framework ในการพัฒนา Web Application

3.1.2 Service Tire

เป็นส่วนที่ทำหน้าที่ติดต่อกับฟังก์ชันประมวลผลการทำงานและประสานงานระ ว่าง App Tier กับ Data Tier ซึ่งส่วน API นั้นจะใช้ Fast API เพื่อเชื่อมต่อระหว่างส่วนของแอปพลิเคชัน

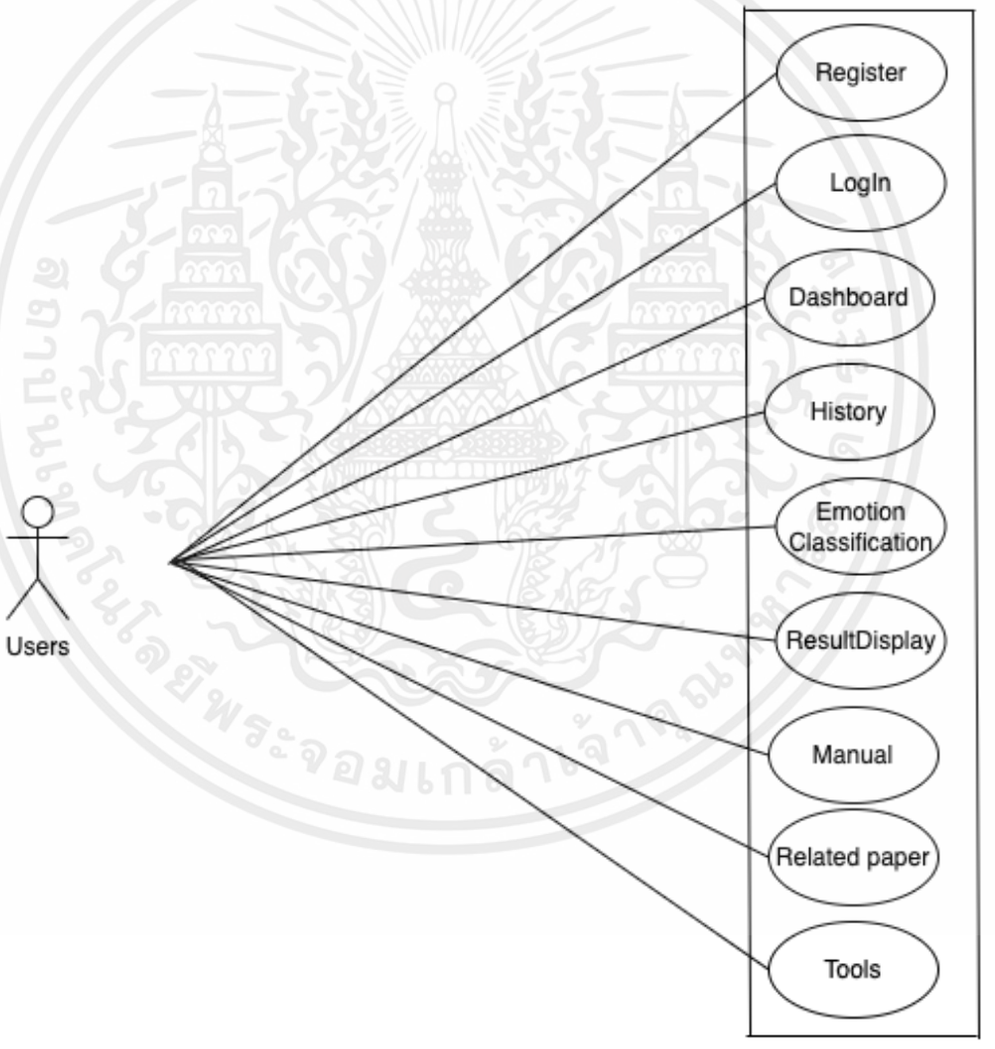
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และส่วนของฐานข้อมูล และส่วน Process จะ Tensor Flow และ Pytorch ในการประมวลผลข้อมูลจาก CNN

3.1.3 Data Tire

เป็นส่วนที่ติดต่อกับฐานข้อมูล จัดเก็บข้อมูล และเข้าถึงข้อมูล โดยในส่วนของ Database จะใช้ PostgreSQL ในการจัดเก็บข้อมูลและ AWS S3 เพื่อใช้ในการเก็บไฟล์เสียง

3.3 การออกแบบ Use case Diagram



รูป 3.2 Use Case Diagram ของ Web Application

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยจะแบ่งฟังก์ชันการทำงานดังนี้

ตาราง 3.3 Use Case Register

Use Case ID	U-1
Use Case	Register
Actor	ผู้ใช้งาน
Main Flow	1) ผู้ใช้งานกรอกข้อมูลเพื่อสมัครสมาชิก 2) สามารถกดเพื่อไปหน้าเข้าสู่ระบบ หากเป็นสมาชิกอยู่แล้ว

ตาราง 3.4 Use Case Login

Use Case ID	U-2
Use Case	Login
Actor	ผู้ใช้งาน
Main Flow	1) กรอกข้อมูลเพื่อเข้าสู่ระบบ 2) สามารถกดไปยังหน้าสมัครสมาชิกได้ หากยังไม่เป็นสมาชิก

ตาราง 3.5 Use Case Dashboard

Use Case ID	U-3
Use Case	Dashboard
Actor	ผู้ใช้งาน
Main Flow	1) ดูรายเอียดของผู้ใช้ได้ 2) สามารถแก้ไขโปรไฟล์ได้ 3) สามารถแก้ไขรหัสผ่านได้ 4) สามารถกดปุ่มเพื่อไปดูประวัติการใช้งานของผู้ใช้ได้

ตาราง 3.6 Use Case History

Use Case ID	U-4
Use Case	History
Actor	ผู้ใช้งาน
Main Flow	1) ดูประวัติการใช้งานของผู้ใช้ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 3.7 Use Case Show Emotion From Sound

Use Case ID	U-5
Use Case	Show Emotion From Sound
Actor	ผู้ใช้งาน
Main Flow	1) แสดงผลลัพธ์

ตาราง 3.8 Use Case Result Display

Use Case ID	U-6
Use Case	Result Display
Actor	ผู้ใช้งาน
Main Flow	1) แสดงผลลัพธ์

ตาราง 3.9 Use Case Manual

Use Case ID	U-7
Use Case	Manual
Actor	ผู้ใช้งาน
Main Flow	1) ผู้ใช้เลื่อนดูแถบวิธีการใช้งาน

ตาราง 3.10 Use Case Related paper

Use Case ID	U-8
Use Case	Related paper
Actor	ผู้ใช้งาน
Main Flow	1) ผู้ใช้เลื่อนดูแถบเอกสารเพื่อเข้าใจเว็บไซต์เพิ่มขึ้น

ตาราง 3.11 Use Case Tools

Use Case ID	U-9
Use Case	Tools
Actor	ผู้ใช้งาน
Main Flow	1) ผู้ใช้เลื่อนดูแถบเพื่อดู API ของเว็บไซต์ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 Requirements ของระบบ

ความต้องการของระบบที่ได้ออกแบบไว้จะแสดงในตารางดังต่อไปนี้

ตาราง 3.7 Requirements ของระบบ

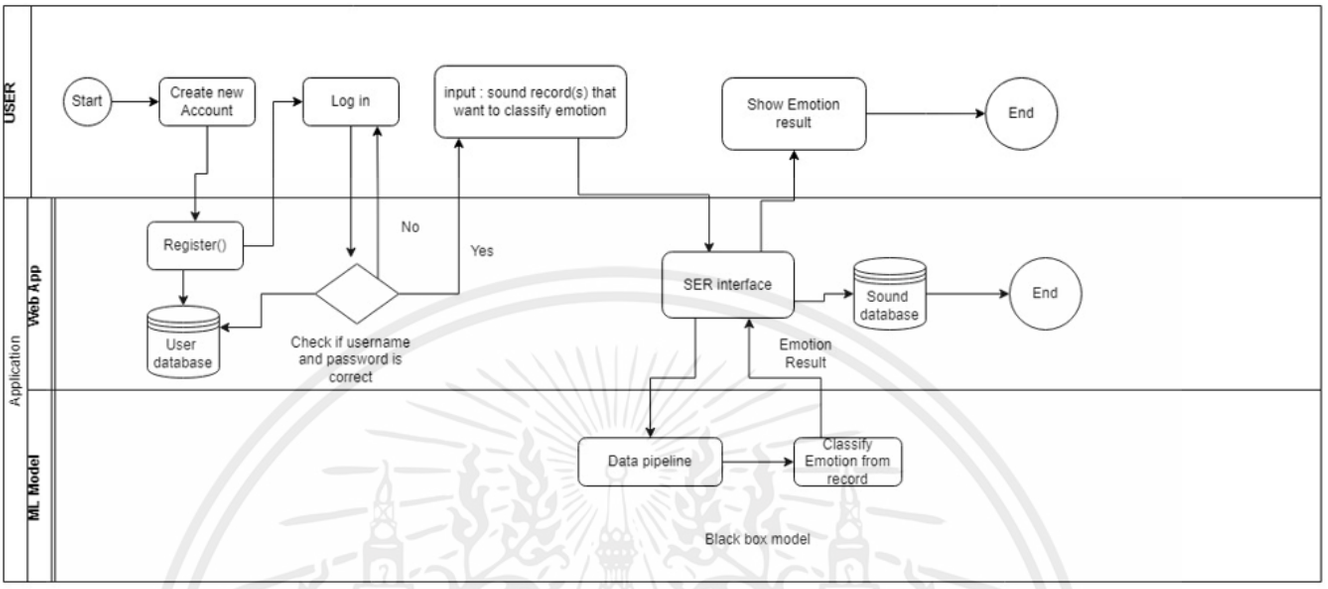
ID	Requirement	Type	Importance
1	**หน้าเข้าสู่ระบบ**	function	Must have
1.1	มีปุ่มให้กดไปหน้ายังสมัครสมาชิก	function	Should have
1.2	มีปุ่มให้กดไปยังหน้าลืมรหัสผ่าน	function	Should have
2	**หน้าลงทะเบียนสมัครสมาชิก**	function	Must have
2.1	มีปุ่มให้กดไปยังหน้าเข้าสู่ระบบ	function	Must have
3	**มีหน้าลืมรหัสผ่าน**	function	Must have
3.1	มีช่องใส่ Email	function	Should have
3.2	มีช่องใช้รหัสยืนยันที่ส่งไปยัง Email	function	Should have
3.3	เมื่อยืนยันสำเร็จจะไปหน้าเปลี่ยนรหัสผ่าน	function	Should have
4.	**มีหน้าเปลี่ยนรหัสผ่าน**	function	Must have
4.1	มีช่องใส่รหัสผ่านใหม่ที่ต้องการ	function	Must have
4.2	กดยืนยันแล้วไปหน้าเข้าสู่ระบบ	function	Must have
5.	**มีหน้าหลัก**	function	Must have
5.1	มีปุ่มเพื่ออัปโหลดเสียง	function	Must have
5.2	มีปุ่มเพื่ออัดเสียงลงไป	function	Must have
6	**มีหน้าแสดงผลลัพธ์**	function	Must have
6.1	มีปุ่มเพื่ออัปโหลดเสียง	function	Must have
6.2	มีปุ่มเพื่ออัดเสียง	function	Must have
6.3	มีแสดงผลลัพธ์	function	Must have
6.4	สามารถตั้งค่าเพิ่มเติมได้เอง	function	Must have

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7	**มีหน้าใช้งานหลัก**	function	Must have
7.1	มีปุ่มเพื่ออัปโหลดเสียง	function	Must have
7.2	มีปุ่มเพื่ออัดเสียงลงไป	function	Must have
8	**มีหน้าสมัครสมาชิกพรีเมียม**	function	Must have
8.1	แสดงราคาต่างๆ	function	Should have
8.2	มีปุ่มสำหรับกดสมัครสมาชิก	function	Should have
9	**มีหน้าแสดงวิธีใช้งาน**	function	Must have
9.1	แสดงวิธีใช้งานเว็บไซต์	function	Must have
10	**หน้าวิธีการใช้งาน API**	function	Must have
10.1	แสดงวิธีการต่างๆในการใช้งาน API	function	Must have
11	**มีหน้าแก้ไขโปรไฟล์**	function	Must have
11.1	แก้ไขรูปภาพ	function	Must have
11.2	แก้ไขรายละเอียดต่างๆ	function	Must have
12	**มีหน้าตั้งค่าข้อมูลการชำระเงิน**	function	Must have
13	**มีหน้าตั้งค่าบัญชีผู้ใช้**	function	Must have
13.1	ปุ่มกดไปยังหน้าแก้ไขโปรไฟล์	function	Must have
14	มี Navbar	function	Must have
14.1	มีปุ่มกดไป Home	function	Must have
14.2	มีปุ่มกดไป How to use	function	Must have
14.3	มีปุ่มกดไป Tool & API	function	Must have
14.4	มีปุ่มกดไป Pricing	function	Must have
14.5	มีปุ่ม Login	function	Must have
14.6	มีปุ่ม Sign Up	function	Must have
14.7	มีปุ่มกดไปหน้าโปรไฟล์	function	Must have
14.8	มีปุ่ม Logout	function	Must have

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5 การออกแบบ Business Process Model and Notation



รูป 3.3 BPMN ของระบบ

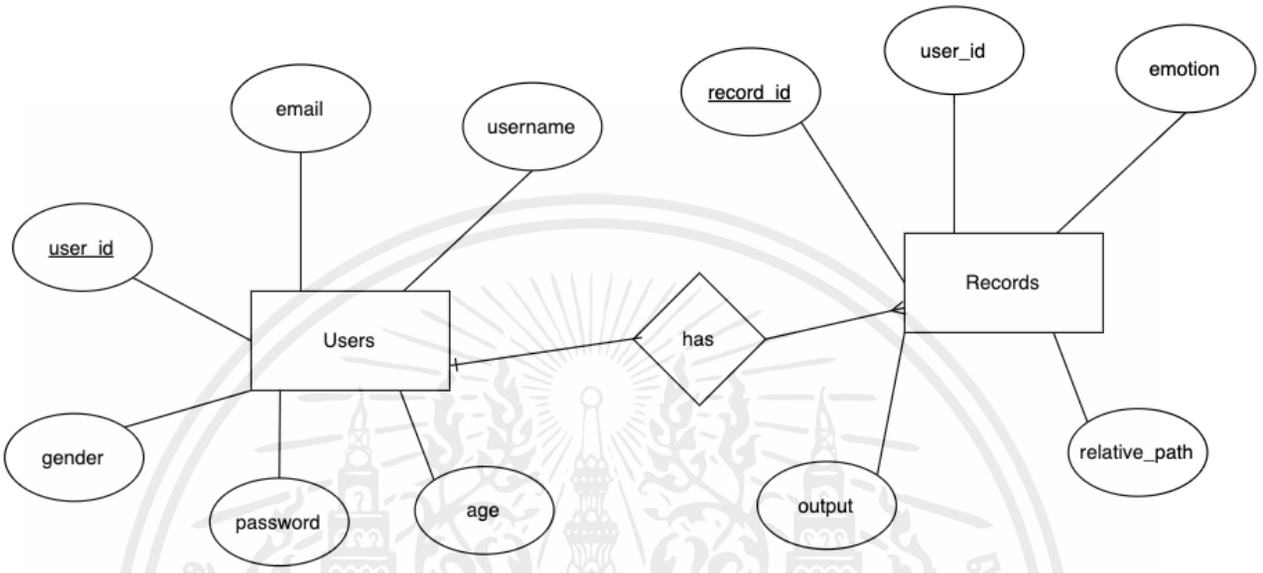
การทำงานจะแบ่งเป็น 3 layers ได้แก่

1) user -> action ของผู้ใช้งาน
2) Web application -> การทำงานของ Application ที่มีการเรียกใช้ DB และ เรียกใช้ ML Model และ แสดงผล
3) ML model -> Data processing pipe line และ Classification

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

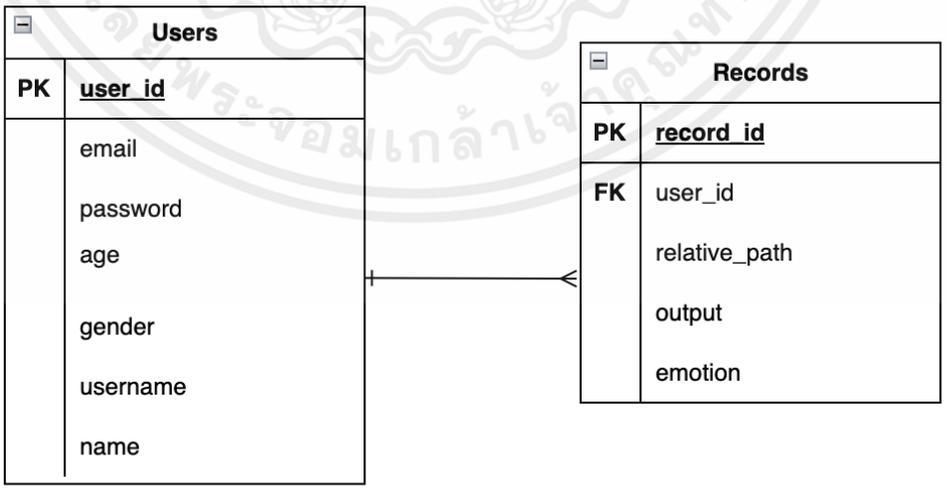
3.6 การออกแบบฐานข้อมูล

3.6.1 การออกแบบ Entity Relationship Diagram



รูป 3.4 Entity Relationship Diagram ของระบบ

3.6.2 การออกแบบ Database Schema



รูป 3.5 Entity Database Schema ของระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.6.2.1 Users Entity

User Entity เป็น Collection สำหรับเก็บข้อมูลของผู้ใช้ดังตารางต่อไปนี้

ตาราง 3.8 Users Entity

Key	Attribute	Type	Detail
PK	user_id	Integer	ไอดีผู้ใช้
	Email	String	อีเมล
	Password	String	รหัสผ่าน
	Age	Integer	อายุ
	Gender	String	เพศ
	Username	String	ชื่อผู้ใช้งาน
	Name	String	ชื่อ

3.6.2.2 Records Entity

เป็น Collection สำหรับเก็บข้อมูลของการจำแนกเสียงออกมาเป็นอารมณ์ดังตาราง ต่อไปนี้

ตาราง 3.9 Users Entity

Key	Attribute	Type	Detail
PK	user_id	Integer	ไอดีผู้ใช้
	Relative_path	String	ชื่อไฟล์
	Output	Integer	จำนวนอารมณ์
	Emotion	String	อารมณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7 การออกแบบส่วนติดต่อผู้ใช้

การออกแบบ User Interface หน้าการใช้งานหลักผ่าน Web Application ดังนี้

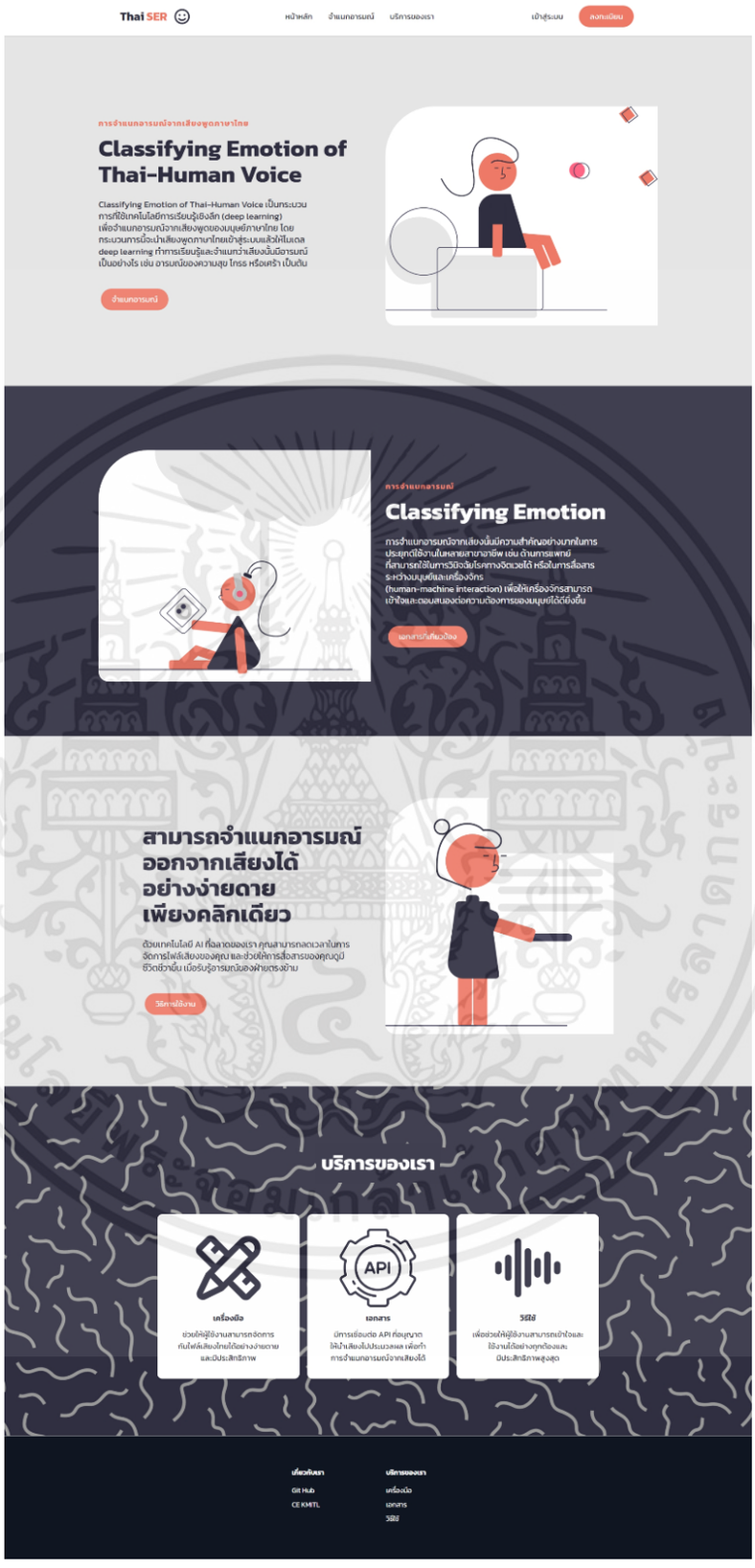
3.7.1 ภาพรวมหน้าหลักการใช้งาน (User Interface Overview)

ตาราง 3.10 ภาพรวมหน้าหลักการใช้งาน (User Interface Overview)

Thai SER	
ส่วนหน้าหลัก	1) Navbar 2) Info Section 3) Services 4) Footer
ส่วนหน้าสมัครสมาชิก	1) กรอกข้อมูล 2) ลงชื่อเข้าใช้งาน
ส่วนหน้าลงชื่อเข้าใช้	1) กรอกข้อมูล 2) เข้าสู่ระบบ
ส่วนหน้าแดชบอร์ด	1) แสดงข้อมูลของผู้ใช้ 2) แก้ไขบัญชี 3) แก้ไขรหัสผ่าน 4) แสดงประวัติการทำรายการของผู้ใช้
ส่วนหน้าแสดงประวัติการใช้งานของผู้ใช้	1) แสดงประวัติการทำรายการของผู้ใช้
ส่วนหน้าจำแนกอารมณ์	1) อัปโหลดเสียง
ส่วนหน้าผลลัพธ์	1) แสดงผลลัพธ์
ส่วนหน้าวิธีใช้งาน	1) แถบวิธีใช้งาน
ส่วนหน้าเอกสารที่เกี่ยวข้อง	1) แถบเอกสารที่เกี่ยวข้อง
ส่วนหน้าเครื่องมือ	1) แถบเครื่องมือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

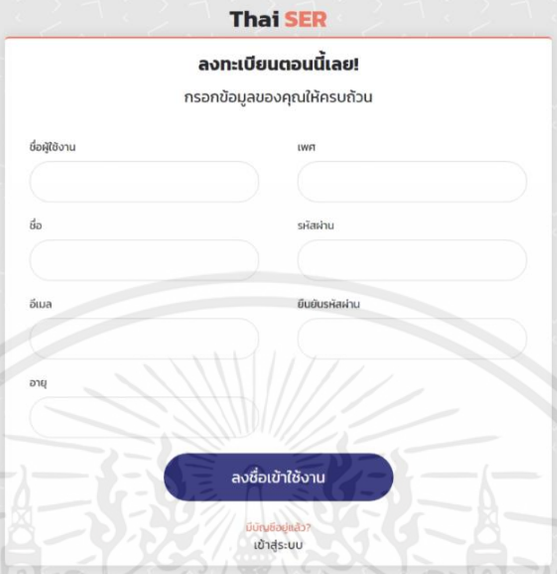
3.7.2 ส่วนหน้าหลัก



รูป 3.6 ส่วนหน้าหลัก

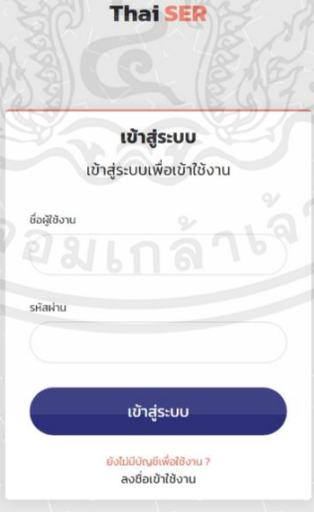
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7.3 ส่วนหน้าสมัครสมาชิก



รูป 3.7 ส่วนหน้าสมัครสมาชิก

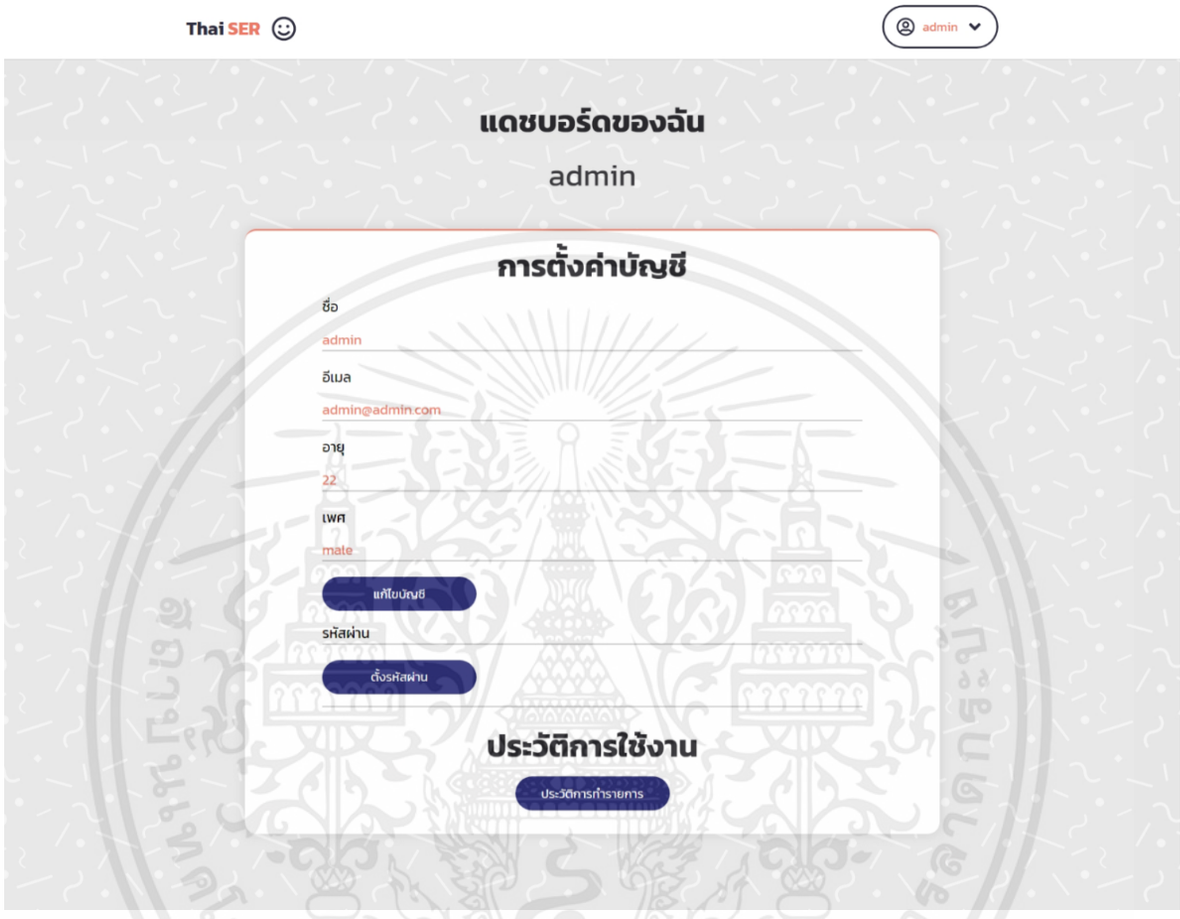
3.7.4 ส่วนหน้าลงชื่อเข้าใช้



รูป 3.8 ส่วนหน้าลงชื่อเข้าใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

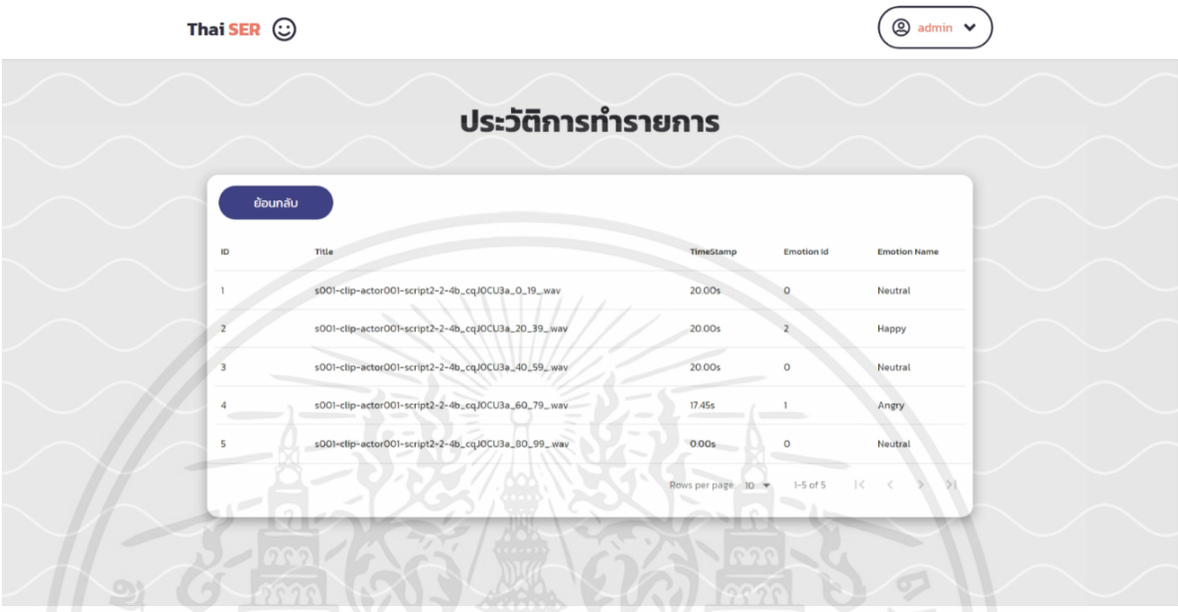
3.7.5 ส่วนหน้าแดชบอร์ด



รูป 3.9 ส่วนหน้าแดชบอร์ด

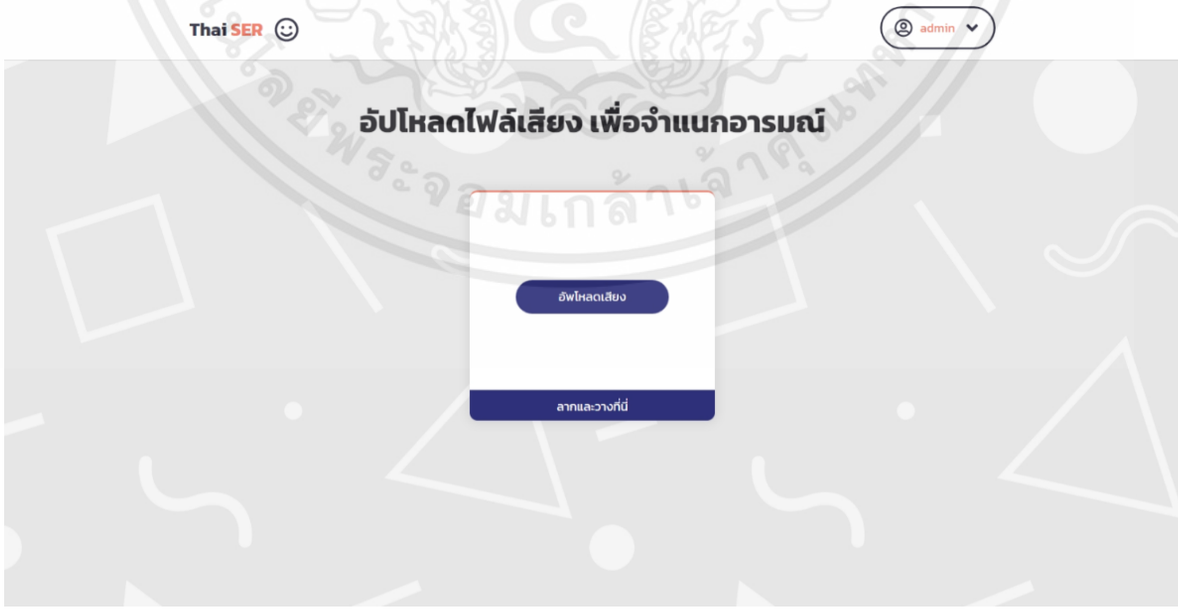
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7.6 ส่วนหน้าแสดงประวัติการใช้งานของผู้ใช้



รูป 3.10 ส่วนหน้าแสดงประวัติการใช้งานของผู้ใช้

3.7.7 ส่วนหน้าจำแนกอารมณ์



รูป 3.11 ส่วนหน้าจำแนกอารมณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7.8 ส่วนหน้าผลลัพธ์

รูป 3.12 ส่วนหน้าผลลัพธ์

3.7.9 ส่วนหน้าวิธีใช้งาน

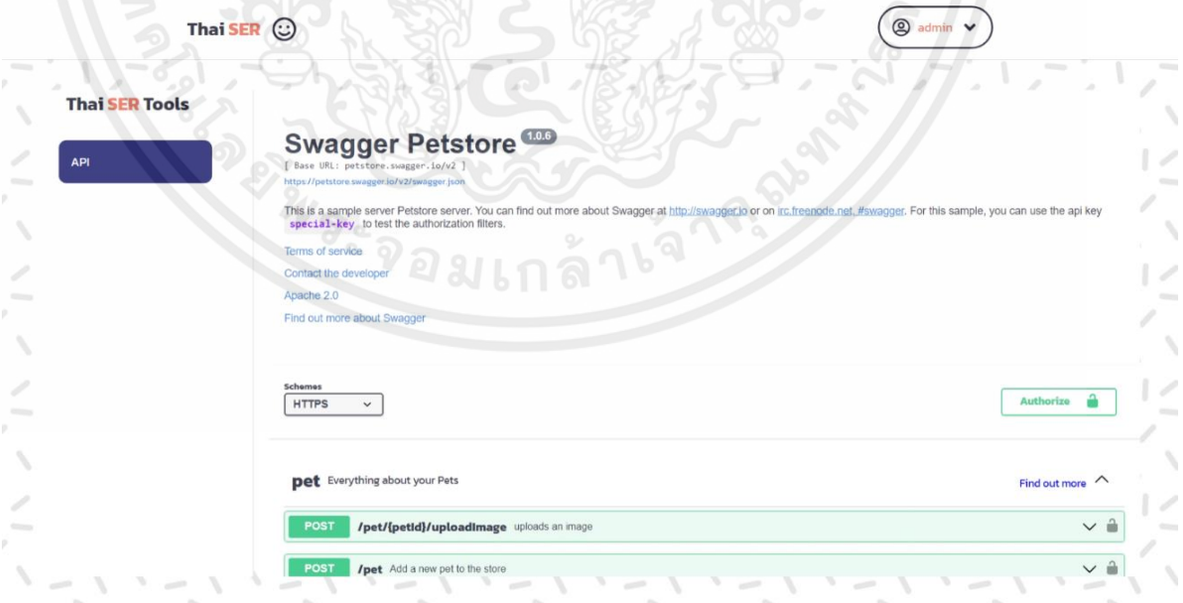
รูป 3.13 ส่วนหน้าวิธีใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7.10 ส่วนหน้าเอกสารที่เกี่ยวข้อง

รูป 3.14 ส่วนหน้าเอกสารที่เกี่ยวข้อง

3.7.11 ส่วนหน้าเครื่องมือ



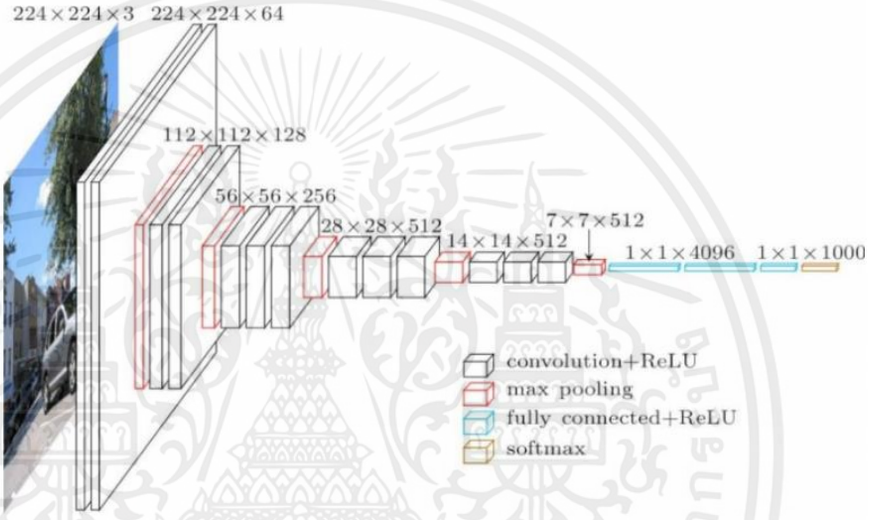
รูป 3.15 ส่วนหน้าเครื่องมือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.8 การออกแบบโมเดลด้วยการเรียนรู้เชิงลึก

3.8.1 Pre-trained model

การออกแบบโมเดลด้วยการเรียนรู้เชิงลึก จะใช้ Pre-trained model มาทำการ Transfer Learning ด้วยเทคนิค Fine-Tuning โดยจะทำการทดลองเพื่อเลือก Pre-trained Model ที่ดีที่สุด สำหรับการทำโมเดลจำแนกอารมณ์จากเสียงโดย Pre-trained model ที่เลือกมาจาก Top 4 for image classification



รูป 3.10 EffNetB0 model Architecture

3.8.2 Data Preparation

3.7.2.1 Convert audio to Image

การแปลงเสียงเป็นรูปเพื่อใช้การแก้ปัญหาด้วย Image Classification ซึ่งมี Pre-trained model จำนวนมากให้เลือก ดังนั้นเราจึงต้องแปลงเสียงเป็นรูปก่อน การจะแปลงเสียงเป็นรูปเราจะใช้การแปลงเป็น spectrogram ซึ่งเป็นกราฟที่แสดงระดับความเข้มเสียง

แต่ในการทำให้โมเดลมีความสามารถในการจำแนกอารมณ์จาก unseen data โดยเราจะใช้หลักการการควบคุม input ให้มีลักษณะเดียวกับ training data มากที่สุด (standardization) แล้วในส่วนที่เหลือเราจะทำการTraining model ให้มีความยืดหยุ่นเพื่อให้รองรับ unseen data

โดยเราจะใช้ data loader 2 ตัวคือ

1) Training data loader

2) Testing data loader

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดย data loader 2 ตัวจะมี pre – processing steps ที่คล้ายกัน(standardization) แต่ใน Training data loader จะมีการ augmentation data ให้ model สามารถรับ input data ได้ยืดหยุ่นมากที่สุดเท่าที่จะเป็นไปได้ แต่ใน Testing data loader จะไม่มีการ augmentation. เพื่อเป็นการสร้างความได้เปรียบในจำแนกอารมณ์

3.7.2.2 Data Augmentation

เทคนิคการปรับแต่งข้อมูล (data) โดยการสร้างข้อมูลเพิ่มขึ้นจากข้อมูลที่มีอยู่เดิมโดยไม่เปลี่ยนแปลงความหมายหรือคุณสมบัติของข้อมูล ซึ่งจะช่วยเพิ่มประสิทธิภาพของโมเดล deep learning ในการฝึกสอน (training) และลดความเสี่ยงของโมเดลในการ overfitting

1) Re – Channel คือ โดยปกติเสียงจะมี แบบ stereo (2 channels) และmono (1 channels) เราจะทำให้มี channel เหมือนกันก่อนเข้า model
2) Resample คือ การเปลี่ยน sample rate ให้มีค่าเท่ากันก่อนเข้า mode
3) Splitting คือ การตัดให้เสียงมีความยาวไม่เกินค่าที่กำหนดไว้
4) Padding คือ ปรับการให้เสียงมีความยาวเท่ากับความยาวที่กำหนดไว้หากเสียงสั้นกว่า ก็จะทำการเติมเสียงเงียบเข้าไป
5) Time Shifting คือ คือการปรับตำแหน่งการเริ่มของเสียง เหตุผลเพื่อให้ model เข้าใจว่าตำแหน่งในช่วงเวลาของเสียงไม่มีผลในการจำแนกอารมณ์

หลังจากแปลงเป็น Spectrogram เราจะ Augment Image โดยการใส่ Masking Image เพื่อให้ยืดหยุ่นต่อระยะเวลาของเสียงที่เป็น Input

3.7.2.3 Convert to Mel- spectrogram

Mel – spectrogram คือ เป็นการแปลงสัญญาณเสียง (audio signal) เป็นภาพ (image) โดยใช้เทคนิค mel-scale ในการแปลงความถี่ (frequency) ของสัญญาณเสียงให้เป็นหน่วย mel ซึ่งเป็นหน่วยที่ใช้วัดความสัมพันธ์ของความถี่ในทางจินตนาการของมนุษย์

ในกระบวนการสร้าง mel-spectrogram จะแบ่งสัญญาณเสียงเป็นช่วงเวลา (time frame) แล้วนำไปทำ Fourier transform เพื่อแปลงสัญญาณเสียงให้เป็นระบบความถี่ (frequency domain) จากนั้นจะนำค่าความถี่ในแต่ละช่วงเวลามาคำนวณด้วย mel-scale เพื่อทำให้เกิดเสียงที่เป็นสัดส่วนอย่างชัดเจนขึ้น และสุดท้ายจะนำผลลัพธ์ที่ได้ไปวาดเป็นภาพ (image)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7.2.4 Image Augmentation

Image masking คือ การปิดรูปบางส่วนให้หายไป เพื่อให้โมเดลได้เรียนรู้เกี่ยวกับเสียงที่ผ่านการ padding มา

3.8.3 Fine – Tuning

นำ Pre-trained ที่เลือกไว้มาปรับค่า Parameter ให้เหมาะสมกับ Data เราเพื่อเพิ่มความแม่นยำ Accuracy

3.8.2.1 ตัวแปรที่จะกำหนดให้เป็นค่าคงตัว

1) ขนาดของรูปภาพ 224 x 224
2) เป็นภาพ RGB
3) ขนาด output layer
4) จำนวน hidden layers

3.8.2.2 ตัวแปรที่ต้องปรับค่าเพื่อให้ความแม่นยำมากขึ้น

1) Dropout rate เป็นพารามิเตอร์สำหรับโมเดล deep learning ที่ใช้ในการลดความเสี่ยงของโมเดลในการ overfitting โดยการสุ่มลบ (dropout) บางโหนด (nodes) ในโมเดลในระหว่างการฝึกสอน (training) ทำให้โมเดลไม่สามารถจำได้ละเอียดเกินไปจากข้อมูล training และทำให้โมเดลมีความสามารถในการทำนายข้อมูลใหม่ที่ไม่เคยเห็นมาก่อนได้ดีขึ้น
2) Loss function เป็นฟังก์ชันที่ใช้ในการวัดความผิดพลาดระหว่างการทำนาย (prediction) และคำตอบที่ถูกต้อง (ground truth) ของโมเดลในการฝึกสอน ฟังก์ชันนี้จะคำนวณค่าความผิดพลาดของโมเดลทุกครั้งที่ทำนาย และใช้ค่านี้ในการปรับพารามิเตอร์ของโมเดลให้มีค่าที่ดีขึ้น โดยเราจะเลือกฟังก์ชันที่เหมาะสมกับงานที่ต้องการจะทำ เช่น mean squared error (MSE) สำหรับงาน regression หรือ categorical cross-entropy สำหรับงาน classification
3) Learning rate เป็นพารามิเตอร์สำหรับโมเดล deep learning ที่กำหนดความเร็วในการปรับพารามิเตอร์ของโมเดลในแต่ละรอบของการฝึกสอน (training) ค่า learning rate นี้จะเป็นตัวกำหนดว่าพารามิเตอร์จะถูกปรับเปลี่ยนในระดับไหน เราจะต้องกำหนดค่า learning rate ให้เหมาะสมเพื่อให้โมเดลสามารถเรียนรู้ได้อย่างมีประสิทธิภาพ
4) Optimizer เป็นอัลกอริทึมที่ใช้ในการปรับพารามิเตอร์ของโมเดล deep learning ให้สอดคล้องกับข้อมูลในแต่ละรอบของการฝึกสอน (training)

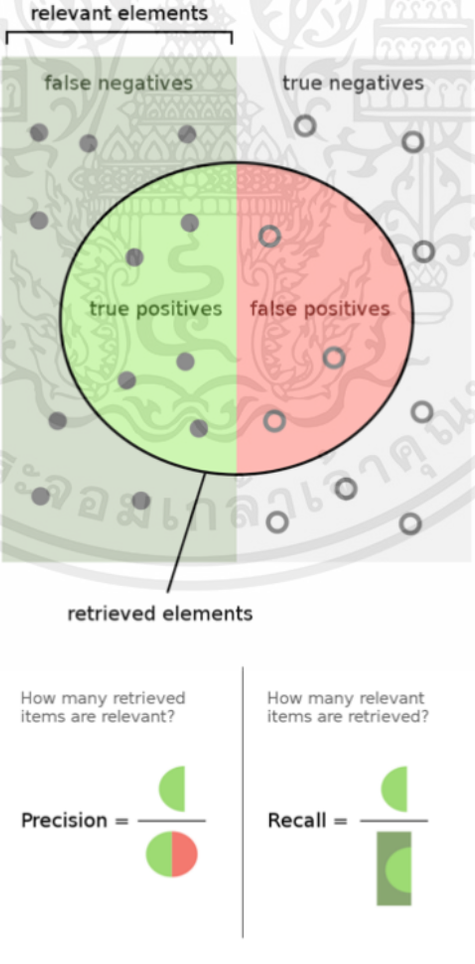
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยจะใช้ Gradient Descent ในการคำนวณค่า gradient และปรับค่าพารามิเตอร์ของโมเดล อย่างไรก็ตาม มี optimizer อื่นๆ เช่น Adam RMSProp ซึ่งมีการปรับปรุงเพิ่มเติมจาก Gradient Descent เพื่อให้สามารถสมรรถนะที่ดีขึ้นในการฝึกสอนโมเดล

5) Epoch and steps เป็นการฝึกสอนโมเดล deep learning โดยการนำข้อมูล training ผ่านโมเดลทั้งหมด 1 รอบ ซึ่งหมายความว่าโมเดลจะมีโอกาสเรียนรู้ข้อมูลจำนวนครั้งตามจำนวน epoch ที่กำหนด ซึ่งสามารถกำหนดได้ตามความเหมาะสมของงาน แต่ที่แนะนำกันมักจะอยู่ที่ 10-100 epoch

3.8.4 การประเมินค่าและวัดผลการทำงานของ Model

ทำการประเมินผลโดยการนาการทดลองที่ 3 มาวัดประสิทธิภาพโมเดล นอกจากจะประเมินด้วยค่า Loss และ Accuracy ทางคณะผู้จัดทำจะทำการประเมินบนพื้นฐาน Precision Recall F-Score และ Confusion matrix



รูป 3.11 Precision Recall

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$(10.1)\quad Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

$$(10.2)\quad Precision = \frac{T_p}{T_p + F_p}$$

$$(10.3)\quad Recall = \frac{T_p}{T_p + T_n}$$

$$(10.4)\quad F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

รูป 3.12 Precision Recall

Confusion matrix เป็นเครื่องมือที่ใช้ในการวัดประสิทธิภาพของโมเดลการจำแนกประเภท (classification) โดยเฉพาะ โดยจะแสดงผลการทำนายของโมเดลเปรียบเทียบกับความจริงว่าโมเดลทำนายถูกหรือผิดแต่ละคลาสแล้วมีจำนวนเท่าไหร่ โดยมีแกนตั้งแสดงถึงคลาสจริง (true class) และแกนนอนแสดงถึงคลาสที่โมเดลทำนาย (predicted class)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการดำเนินงาน

4.1 ผลการพัฒนาแอปพลิเคชัน

หลังจากผู้จัดทำได้พัฒนา Web Application แล้ว ทางผู้จัดทำได้ทำการทดสอบและทดลองใช้งานบนหน้าเว็บไซต์ขนาดมาตรฐาน โดยได้รับผลลัพธ์ที่สมบูรณ์และตรงตามความต้องการ โดยได้รับผลลัพธ์ดังนี้

4.1.1 ผลการทดสอบ Web Application

จัดทำในส่วนติดต่อผู้ใช้ ซึ่งประกอบด้วยส่วนสำหรับตรวจสอบว่าเข้าสู่ระบบหรือยังไม่เข้าสู่ระบบส่วนการสมัครสมาชิกและส่วนการเข้าสู่ระบบหากผู้ใช้ไม่มีบัญชีผู้ใช้ก็จะต้องกดปุ่มลงทะเบียน หรือถ้ามีบัญชีผู้ใช้แล้วแต่ยังไม่ได้เข้าสู่ระบบก็จะต้องกดปุ่มเข้าสู่ระบบใน Navbar ข้างบน โดยผู้ใช้จะต้องกรอกข้อมูลที่กำหนดและกดปุ่มยืนยันเพื่อสมัครสมาชิกหรือเข้าสู่ระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

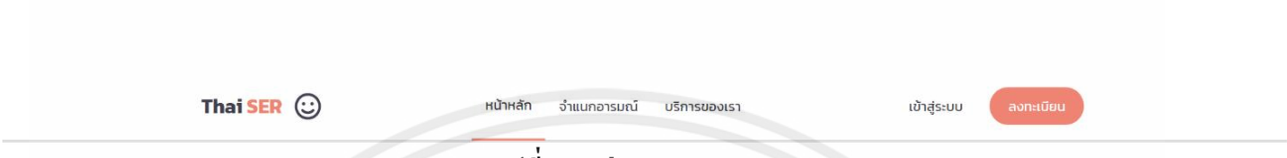
1) ส่วนหน้าหลักประกอบด้วย 4 ส่วนหลักๆได้แก่ Nav Bar Info Service Footer

รูปที่ 4.1 ส่วนหน้าหลัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

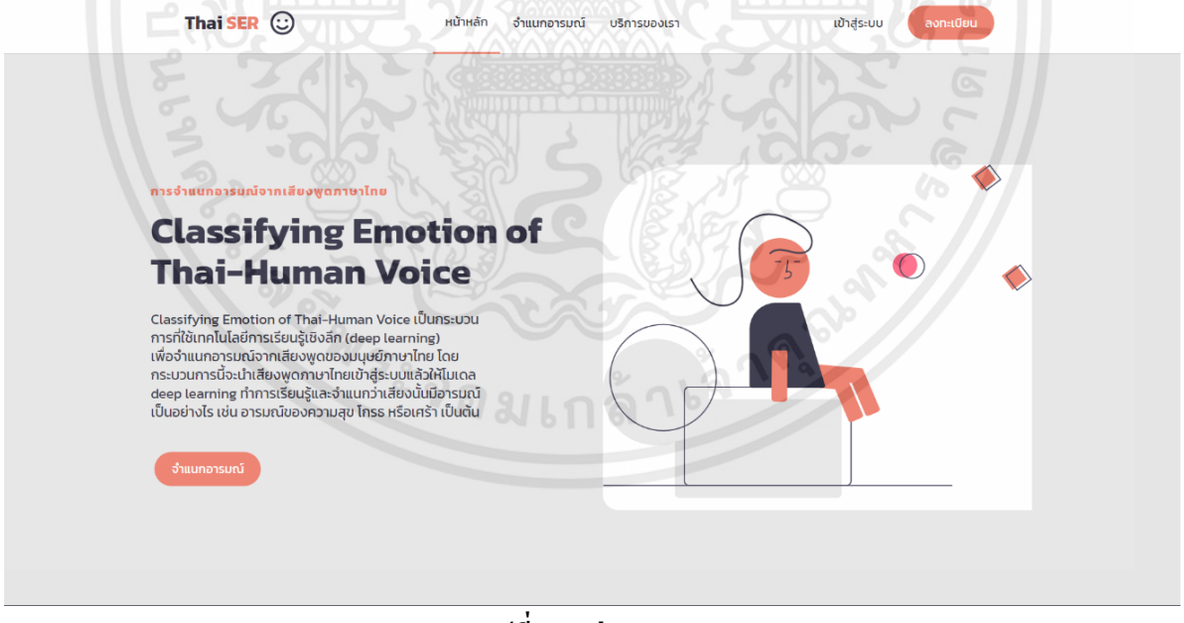
2) ส่วนของ Navbar มีรายละเอียดดังนี้

- มีไอคอน Thai SER เมื่อกดจะไปหน้าแรกได้
- สามารถกด หน้าหลัก จำแนกอารมณ์ บริการของเรา เพื่อไปยัง Info นั้นๆ
- มีปุ่มให้กดเพื่อเข้าสู่ระบบ และ ลงทะเบียน



รูปที่ 4.2 ส่วนของ Navbar

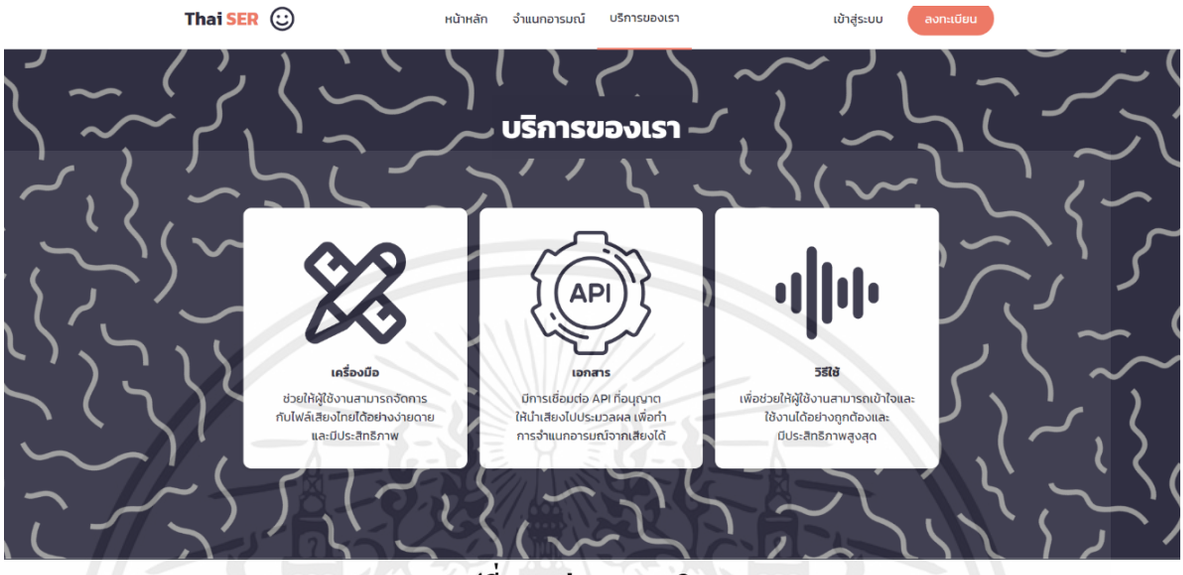
3) ส่วนของ Info มีไว้เพื่อแสดงรายละเอียดในแต่ละหมวดหมู่ ได้แก่ หน้าหลัก จำแนกอารมณ์ เครื่องมือและ API ในแต่ละ Info จะประกอบด้วย Topline Headline Description Image Button



รูปที่ 4.3 ส่วนของ Info

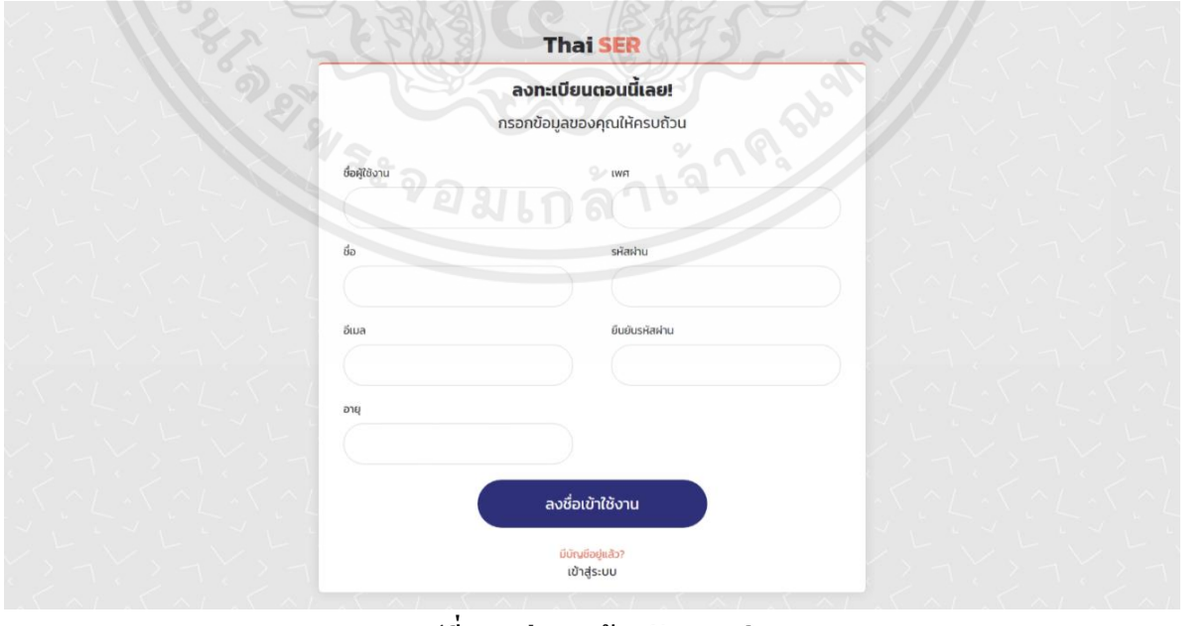
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) ส่วนของบริการจะแสดง Card ได้แก่ เครื่องมือ เอกสาร วิธีใช้ เมื่อกดไปยัง Card ต่างๆจะเข้าสู่หน้าตามที่กำหนดไว้



รูปที่ 4.4 ส่วนของบริการ

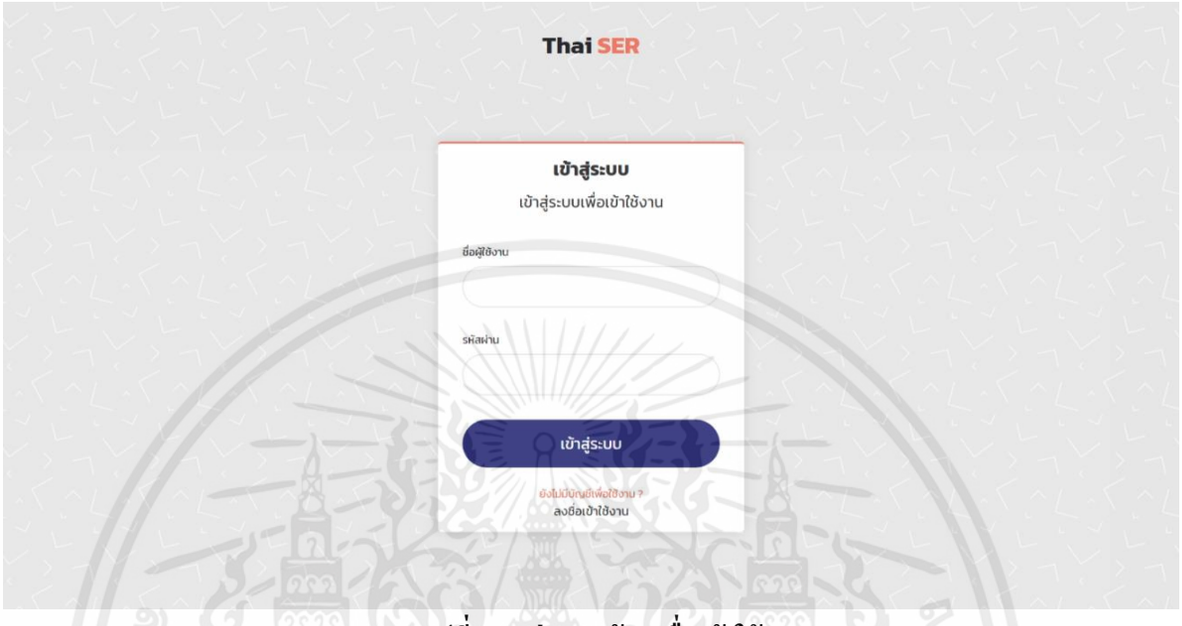
5) ส่วนหน้าสมัครสมาชิก เพื่อที่จะสมัครสมาชิกจะต้องกรอกข้อมูลดังนี้ ชื่อผู้ใช้งาน ชื่อ อีเมล อายุ เพศ รหัสผ่าน ยืนยันรหัสผ่าน เพื่อที่จะสมัครได้



รูปที่ 4.5 ส่วนหน้าสมัครสมาชิก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

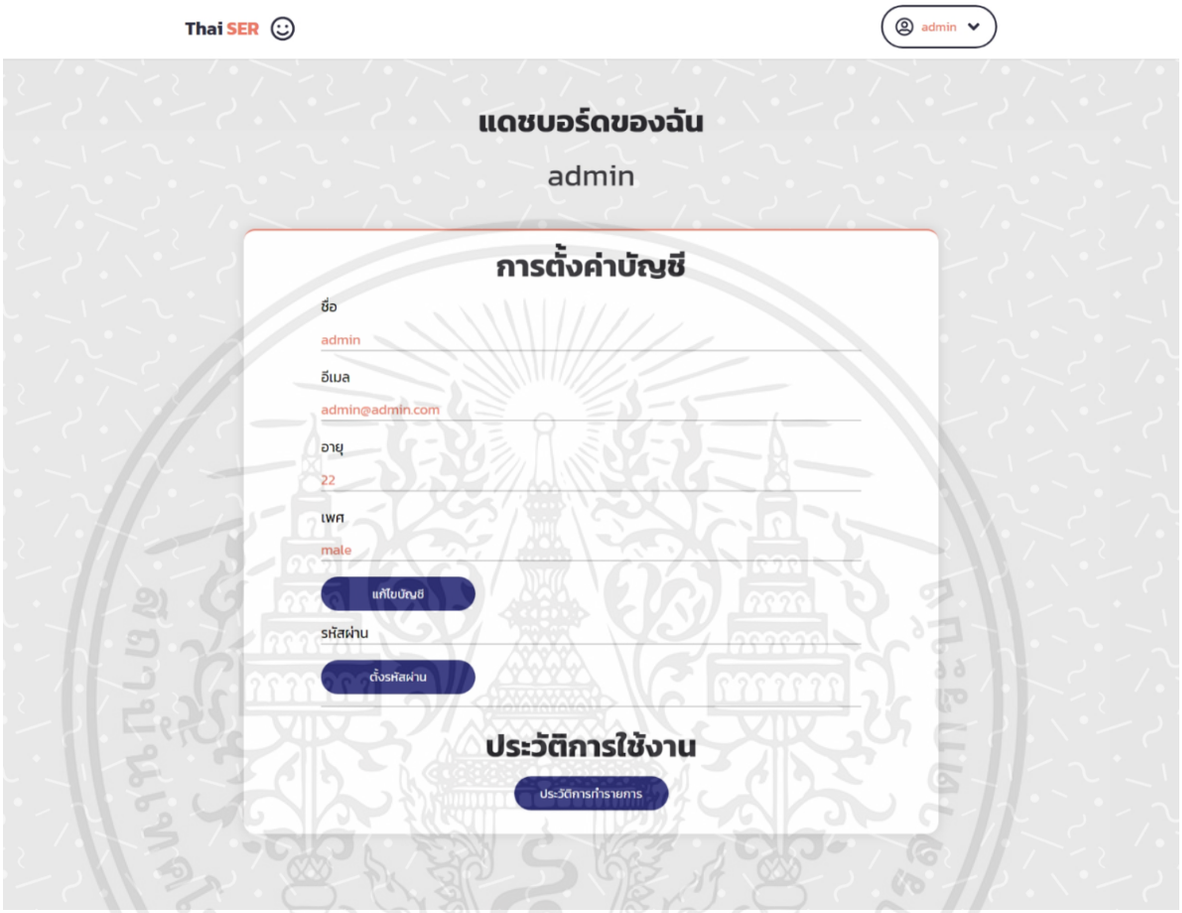
6) ส่วนหน้าลงชื่อเข้าใช้ ต้องกรอกข้อมูล ชื่อผู้ใช้งานและรหัสผ่านเพื่อจะเข้าสู่ระ บบ ได้



รูปที่ 4.6 ส่วนหน้าลงชื่อเข้าใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

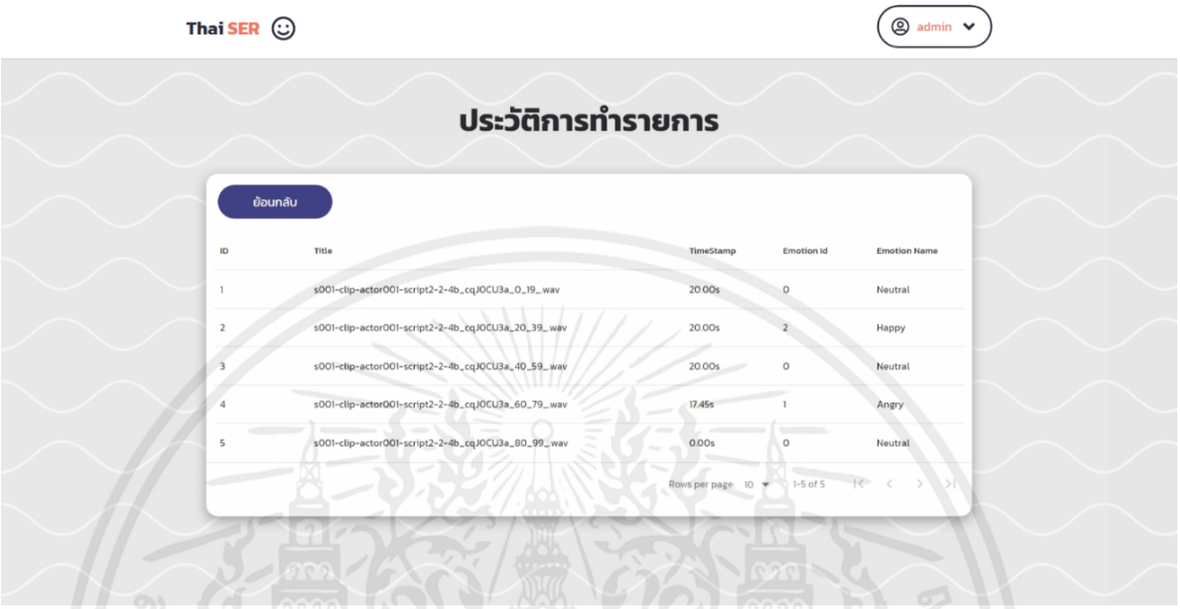
7) ส่วนหน้าแดชบอร์ด แสดงรายละเอียดต่างๆของผู้ใช้ได้แก่ ชื่อผู้ใช้งาน ชื่อ อีเมล อายุ เพศ สามารถแก้ไขบัญชีได้และเปลี่ยนรหัสผ่านได้



รูปที่ 4.7 ส่วนหน้าแดชบอร์ด

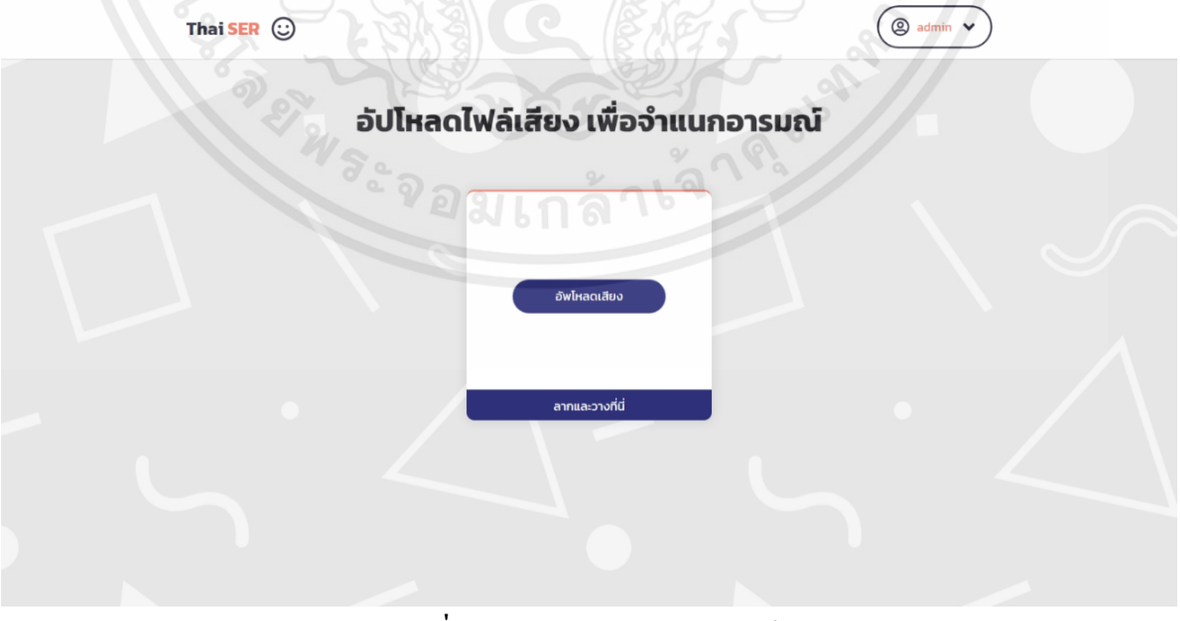
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8) ส่วนหน้าแสดงประวัติการใช้งานของผู้ใช้ แสดงประวัติการทำรายการของผู้ใช้ และมีปุ่มเพื่อย้อนกลับไปยังหน้าแดชบอร์ด



รูปที่ 4.8 ส่วนหน้าแสดงประวัติการใช้งาน

9) ส่วนหน้าจำแนกอารมณ์ สามารถลากหรือกดปุ่มอัปโหลดเสียงได้



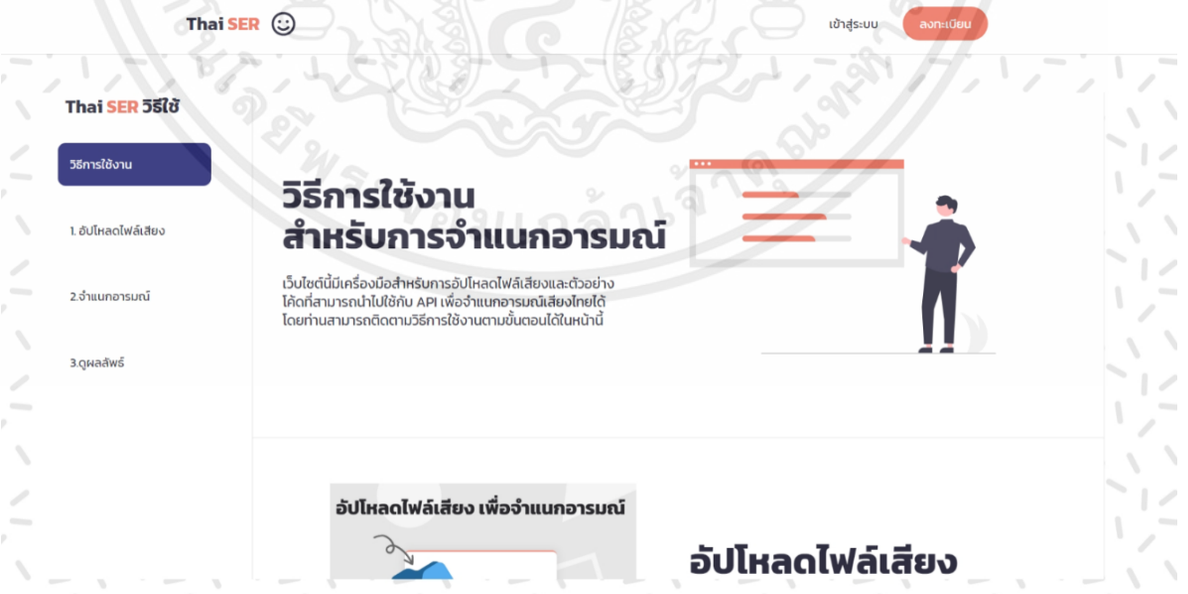
รูปที่ 4.9 ส่วนหน้าจำแนกอารมณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

10) ส่วนหน้าผลลัพธ์ แสดงผลลัพธ์ออกมาเป็นจำนวนอารมณ์และแสดงผลลัพธ์ที่มากที่สุด และสามารถอัปโหลดเสียงเพิ่มได้ในหน้านี้

รูปที่ 4.10 ส่วนหน้าผลลัพธ์

11) ส่วนหน้าวิธีใช้งาน มี Vertical Navbar สามารถกดปุ่มเพื่อไป Info ต่างๆได้



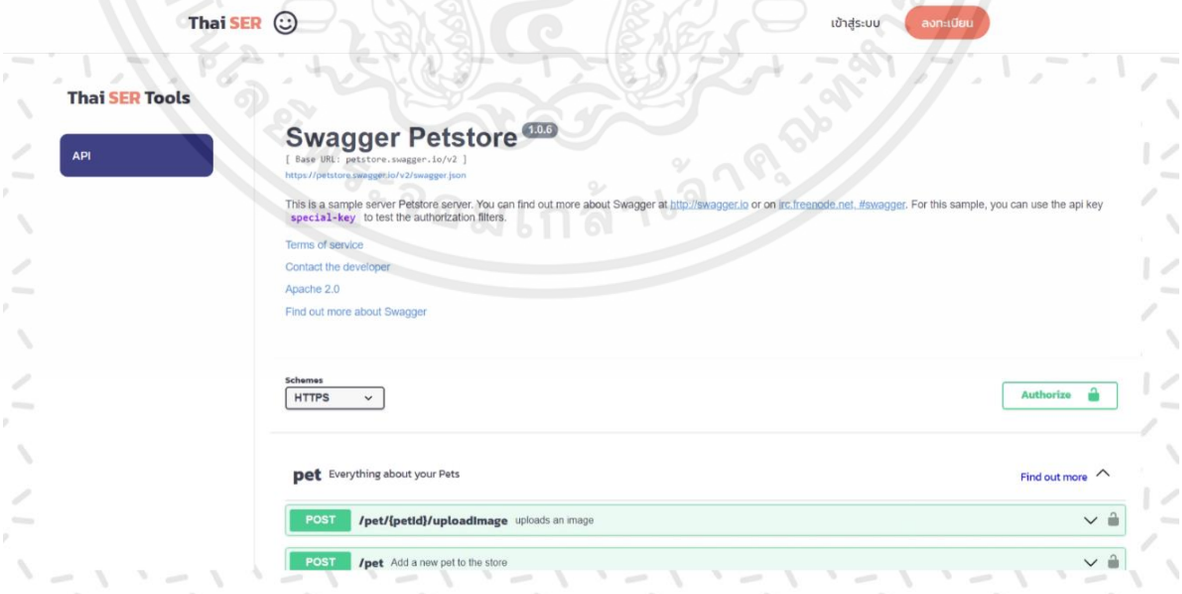
รูปที่ 4.11 ส่วนหน้าวิธีใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

12) ส่วนหน้าเอกสารที่เกี่ยวข้อง มี Vertical Navbar สามารถกดปุ่มเพื่อไป Info ต่างๆ ได้

รูปที่ 4.12 ส่วนหน้าเอกสารที่เกี่ยวข้อง

13) ส่วนหน้าเครื่องมือ มี Vertical Navbar สามารถกดปุ่มเพื่อไป Info ต่างๆ ได้

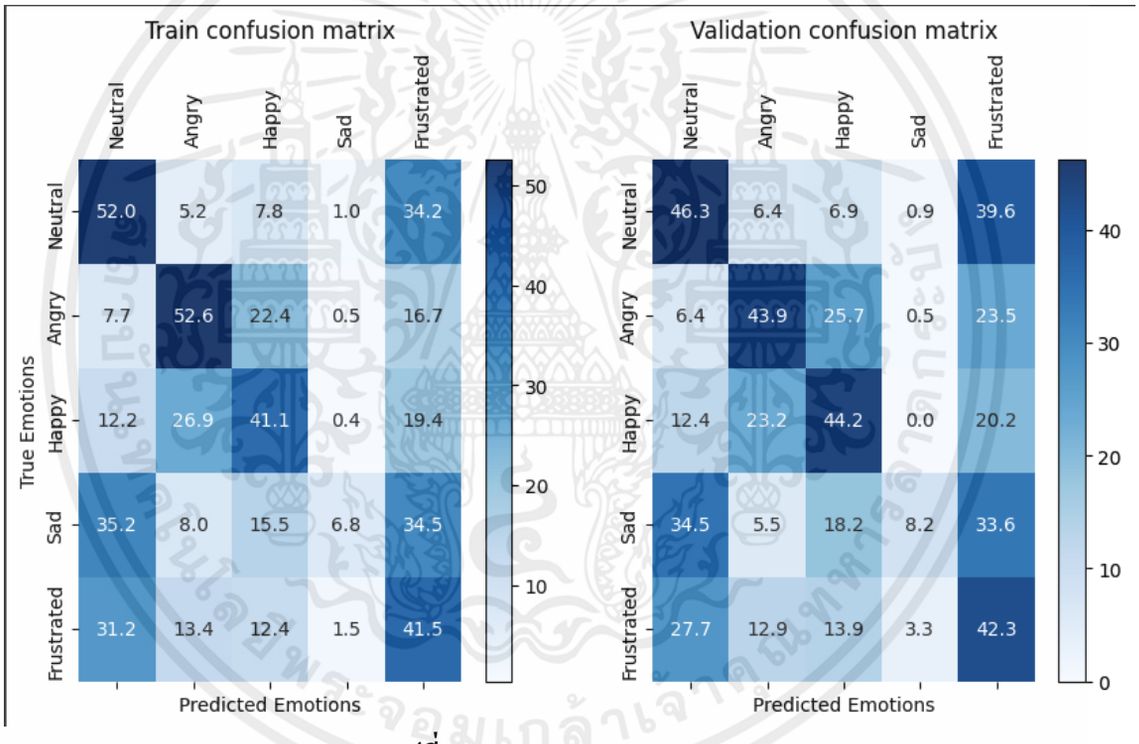


รูปที่ 4.13 ส่วนหน้าเครื่องมือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 ผลการจำแนกอารมณ์จากเสียงของโมเดล

จากผลการจำแนกอารมณ์จากเสียงของโมหลังจากการทดลองปรับ model architecture พบว่าโมเดลที่ดีที่สุดคือ Efficientnetb0 ซึ่งไม่ต้องการการ attach layer เพิ่มเติม แต่เพียงแค่เปลี่ยน output node ใน output layer และ freeze parameters ใน layer ที่เหลืออยู่ ผลลัพธ์ที่ได้คือ Training accuracy ที่มีค่าเท่ากับ 93.89% และ Test accuracy ที่มีค่าเท่ากับ 67.52% จากผลการทดลองจะเห็นได้ว่าโมเดลมีการ overfitting อยู่ แต่ใกล้เคียงกับค่าที่คาดหวังไว้ในตอนแรกที่ต้องการให้ค่า accuracy บนข้อมูลที่ไม่เคยเห็นมีค่าประมาณ 70% นั่นเพราะว่าจากงานวิจัยพบว่ามนุษย์สามารถจำแนกได้ถูกต้องเพียง 63%



รูปที่ 4.14 Confusion Metrics

จาก Confusion matrix จะเห็นว่า Model มีการสับสนระหว่างคู่อารมณ์ Happy – Angry และมีความยากในการ Classify อารมณ์ Frustrated – Neutral – Sad ทำให้คาดการว่าอาจจะต้องหา Feature อื่นเพิ่ม เช่น Context. ของบทสนทนา หรือ อาจจะต้องเปลี่ยนโมเดลเป็น RNN แล้วเพิ่ม data

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปการดำเนินงานของโครงงาน

5.1 สรุปผลที่ได้จากการทำโครงงาน

จากการดำเนินงานของโครงงาน ส่วนโมเดล โมเดลมีความแม่นยำประมาณ 63% ซึ่งถือว่าใกล้เคียงกับความคาดหมาย และ ดีกว่าการจำแนกอารมณ์โดยมนุษย์ (21%) โดยในส่วนของ Web Application สามารถอัปโหลดเสียงไปทำการ Pre-Process ก่อนส่งจะส่งให้โมเดลประมวลผลและแสดงผลลัพธ์หลังจากกระบวนการทั้งหมดเสร็จสิ้นและเก็บประวัติการจำแนกเสียงของผู้ใช้งาน ซึ่งอาจมีประโยชน์ในการพัฒนาและใช้งานในหลายสาขาอุตสาหกรรม เช่น สื่อสาร สังคมออนไลน์ การแปลภาษาเสียง หรือการดูแลสุขภาพในด้านจิตใจ เป็นต้น

5.2 ปัญหา อุปสรรค และแนวทางแก้ไข

5.2.1 ปัญหาจากการพัฒนาโมเดล

1) ปัญหา : Permission denied ใน s3 เผลออัพโหลด credential ไปบน github ทำให้ s3 ทำการ denied permission ในการ delete แล้วถ้าเราไม่รู้สาเหตุทำให้เสียเวลานานในการ debugs

แก้ไขโดย : ไม่ upload permission ลง github หรืออาจจะเลือกใช้ vault secret

2) ปัญหา : เวลาไม่เพียงพอในการ train data. จำนวนมาก ซึ่งปกติจะต้องคำนวณความคุ้มค่าในการลงทุนเวลา

แก้ไขโดย : หากสามารถทำกำไรได้จากโปรเจคนี้แล้วคุ้มค่า เราควรจะ split data ให้เล็กว่านี้

5.2.2 ปัญหาจากการพัฒนาแอปพลิเคชัน

1) ปัญหา : การอัพโหลดไฟล์ไปยัง S3 เนื่องจากต้องขอ Access

แก้ไขโดย : ไปขอ Access เพื่อที่จะได้ Key มาใส่ใน API เพื่อให้อัพโหลดไฟล์ได้

2) **ปัญหา :** การอัพโหลดไฟล์แล้วบันทึกไปยัง Data Base แล้วข้อมูลผิดพลาดไม่สามารถดึง user_id มาใช้งานได้

แก้ไขโดย : ไปลบไฟล์เก่าๆที่ลองเทสไว้เพราะมีไฟล์ซ้ำกันหลายไฟล์เกินไป

3) **ปัญหา :** ไม่สามารถแสดงผลลัพธ์ใน History ได้เพราะยังไม่มี ID ใน Table ของ Records

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แก้ไขโดย : เพิ่ม Colum ID ใน Table Record

5.3 แนวทางการพัฒนาต่อ

หลังจากที่ได้พัฒนาการจำแนกอารมณ์จากเสียงพูดภาษาไทยโดยใช้การเรียนรู้เชิงลึกนั้น คณะผู้จัดทำมีแนวทางในการพัฒนางานวิจัยดังต่อไปนี้

5.3.1 การพัฒนาโมเดล

1) ทำการหาวิธี sound augmentation ที่เหมาะสมมากว่าเพื่อแก้ปัญหาตาม โดยดูจาก confusion metrics ว่าควรทำ augmentation แบบใด และ ควรเพิ่ม dataset เพื่อให้หา underlying pattern ได้ดียิ่งขึ้น
2) ควร Split เสียงให้มี length สั้นกว่านี้ต่อ file จะทำให้โมเดลได้รับข้อมูลเยอะขึ้น แต่กินเวลาในการ Split และ load เพื่อ Train
3) จาก Confusion matrix จะเห็นว่า Model มีการสับสนระหว่างคู่อารมณ์ Happy – Angry และ มีความยากในการ Classify อารมณ์ Frustrated – Neutral – Sad ทำให้คาดการว่าอาจจะต้องหา Feature อื่นเพิ่ม เช่น Context. ของบทสนทนา หรืออาจจะต้องเปลี่ยนโมเดลเป็น RNN แล้วเพิ่ม data

5.3.2 การพัฒนา Web Application

1) เพิ่มการสมัครสมาชิกเพื่อให้มีฟังก์ชั่นเพิ่มเติมขึ้น เช่น สามารถใช้ API ได้ สามารถกำหนดเวลาในการจำแนกอารมณ์ได้ กำหนดจำนวนครั้งสำหรับผู้ใช้ทั่วไป เป็นต้น
2) ทำให้สามารถอัพโหลดไฟล์เสียงได้หลายเสียงภายในครั้งเดียว
3) ทำให้อัพโหลดเสียงได้หลายประเภทเช่น MP3 FLAC AAC เป็นต้น
4) ทำให้หน้าเว็บมีเวอร์ชั่น Mobile เพื่อให้สามารถใช้งานบน Mobile ได้ด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

Divya Sheel. 2020. **Deep Learning คืออะไร?**. [Online]. Available : https://new.abb.com/news/detail/58004/deep-learning

Surapong Kanoktipsatharporn. 2020. **Transfer Learning คืออะไร สอน Transfer Learning จากโมเดล MobileNet JSON ไป Retrain เทรนต่อ ภาพจากกล้อง Webcam ด้วย TensorFlow.js – tfjs ep.10**. [Online]. Available : https://www.bualabs.com/archives/3493/what-is-transfer-learning-build-headless-mobilenet-model-transfer-learning-machine-learning-mobilenet-json-retrain-webcam-tensorflow-js-tfjs-ep-10/

Navapat Jesadapatrakul. 2019. **Image Processing.** [Online]. Available : https://medium.com/tni-university/image-processing-981c65c26289

Designil. 2017. **React คืออะไร? ไขข้อสงสัยสำหรับมือใหม่ + แนวทางการหัด React ตั้งแต่เริ่มต้น**. [Online]. Available : https://www.designil.com/react-%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3/

Margaret Lech1, Melissa Stolar, Christopher Best and Robert Bolia**.** 2020. **Real-Time Speech Emotion Recognition Using a Pre-trained Image Classification Network: Effects of Bandwidth Reduction and Companding**. [Online]. Available : https://www.frontiersin.org/articles/10.3389/fcomp.2020.00014/full#h1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

K HITHEESH RAJ. 2021. **Speech Emotion Recognition with Convolutional Neural Network. 2021**. [Online]. Available : https://www.analyticsvidhya.com/blog/2021/07/speech-emotions-recognition-with-convolutional-neural-networks/

Salvatore Casale, Alessandra Russo,G. Scebba and Salvatore Serrano. 2008. **Speech Emotion Classification Using Machine Learning Algorithms**. [Online]. Available : https://www.researchgate.net/publication/221406028_Speech_Emotion_Classification_Using_Machine_Learning_Algorithms

สถาบันวิจัยปัญญาประดิษฐ์ประเทศไทย (AIResearch) ร่วมมือกับภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ และ ภาควิชาศิลปการละคร คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. 2021. **THAI SER Database by VISTEC.** [Online]. Available. [Online]. Available : https://airesearch.in.th/releases/speech-emotion-dataset/

Dishashree26 Gupta. 2017. **Transfer learning and the art of using Pre-trained Models in Deep Learning**. [Online]. Available : https://www.analyticsvidhya.com/blog/2017/06/transfer-learning-the-art-of-fine-tuning-a-pre-trained-model/

Sebastián Ramírez. 2019. **FastAPI**. [Online]. Available : https://fastapi.tiangolo.com/

TensorFlow I/O. 2020. **TensorFlow Audio Data preparation**. [Online]. Available : https://www.tensorflow.org/io/tutorials/audio

React. **React**. 2017. [Online]. Available : https://reactjs.org/docs/getting-started.html

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้