

การเปรียบเทียบระหว่างอารมณ์ของทำนองเพลง และเนื้อเพลงไทย

Mood Comparison of Melodies and Lyrics in Thai Songs

ชลาติ ผดุงม่วงทอง, ภูวนาส สิงหาราช, ชญานิศ สร้อยคำ, นิชชา ศศิธรรัศมี, ชินกฤต คชภักดี และ สัจจา

ภรณ์ ไวจรรยา, ณัฐโชติ พรหมฤทธิ์

สาขาวิชาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร

บทคัดย่อ

อารมณ์ของผู้คนส่งผลต่อการเลือกเพลงที่ฟัง ผู้คนมักเลือกฟังเพลงที่สะท้อนอารมณ์ในห้วงขณะนั้น ในขณะที่เดียวกับที่เพลงก็มีอิทธิพลส่งผลต่ออารมณ์ของผู้ฟังให้ปรับเปลี่ยนไปตามอารมณ์ที่เพลงถ่ายทอด เช่นเดียวกัน โดยการถ่ายทอดอารมณ์สู่ผู้ฟังของเพลงนั้น มี 2 องค์ประกอบ คือ ทำนองเพลง และเนื้อเพลง บทความนี้มีวัตถุประสงค์เพื่อหาความสัมพันธ์ระหว่างอารมณ์ของทำนองเพลง และเนื้อหา เพื่อให้ผู้ฟังได้รับฟังเพลงที่ถ่ายทอดอารมณ์ตามความต้องการ จากการอ้างอิงโดยโมเดลวงล้ออารมณ์ของรัสเซลล์ (Russell's Circumplex Model of Affect) ในบทความนี้จึงมีการกำหนดอารมณ์เพลงทั้งหมด 4 อารมณ์ ได้แก่ Angry Calm Happy และ Sad เก็บข้อมูลโดยใช้วิธี Web scraping ดึงข้อมูลเนื้อเพลงจากเว็บไซต์ LyricFind อารมณ์ละ 50 เพลง รวมทั้งหมด 200 เพลง และทำนองเพลงจากเว็บไซต์ YouTube อารมณ์ละ 100 เพลง รวมทั้งหมด 400 เพลง โดยทั้งหมดเป็นเพลงภาษาไทย และทำการจำแนกด้วยชุดยูนิตของคณะผู้จัดทำ ทำการทดลอง และเปรียบเทียบประสิทธิภาพเพื่อหาโมเดล และเทคนิคที่เหมาะสมในการจำแนกอารมณ์เนื้อเพลง 3 โมเดล ได้แก่ RNN แบบ LSTM CNN และโมเดลที่ผ่านการ Fine-tune ด้วย WangchanBERTa จำแนกอารมณ์ทำนองเพลงด้วย CNN โดยการสกัด feature 3 เทคนิค ได้แก่ Mel Spectrogram Chromagram และ MFCC พบว่าโมเดล และเทคนิคที่มีความเหมาะสมที่สุด ได้แก่ โมเดลที่ผ่านการ Fine-tune ด้วย WangchanBERTa มีความแม่นยำอยู่ที่ร้อยละ 75 และเทคนิค Mel Spectrogram มีความแม่นยำอยู่ที่ร้อยละ 60 การเปรียบเทียบระหว่างอารมณ์ของทำนองเพลง และเนื้อเพลง 50 เพลงพบว่าสามารถจำแนกเพลงที่มีอารมณ์ของทำนองเพลง และเนื้อเพลงที่ไม่สอดคล้องกันมีความแม่นยำอยู่ที่ร้อยละ 60

คำสำคัญ: การจำแนกข้อความ, การจำแนกเสียง, เพลงไทย, Mel Spectrogram, WangchanBERTa, CNN

ABSTRACT

Sometimes people choose songs to match their mood or for self-awareness at a given time, while songs can also influence and change people's mood. Considering the elements of a song, such as melody and lyrics, its mood is essentially tied to the requirements of certain audiences, which is the focus of this paper. Moreover, based on Russell's Circumplex Model, angry, calm, happy, and sad are labeled as the four classes of moods. A hundred samples of melodies per class were collected from Thai songs by scraping from YouTube, and fifty samples of lyrics per class were similarly scraped from LyricFind, collaboratively classified at the authors' discretion. According to the results of the experiment, Mel Spectrogram is the most suitable technique for classifying melodies compared to Chromagram and MFCC, with all three techniques applied using a CNN model. Meanwhile, the WangchanBERTa fine-tuned model is the most suitable for classifying lyrics compared to RNN-LSTM and CNN. Their accuracies are 60% and 75%, respectively. The comparison between the moods of the melodies and lyrics of 50 songs revealed that these models can identify songs with contrasting moods between melody and lyric with 60% accuracy.

Keywords: Text classification, Audio classification, Thai songs, Mel Spectrogram, WangchanBERTa, CNN

1. บทนำ

ดนตรีมีอิทธิพลอย่างมากต่ออารมณ์ ในปัจจุบันอุตสาหกรรมดนตรีออนไลน์ เช่น Joox, Spotify และแพลตฟอร์มอื่นๆ ทำให้เกิดคลังเพลงขนาดใหญ่รวมถึงเพลงใหม่ที่ถูกสร้างขึ้นอย่างต่อเนื่อง การทำความเข้าใจอารมณ์ของเพลง เป็นส่วนสำคัญในการพัฒนาที่ตอบสนองความต้องการเฉพาะบุคคล เช่น การจัดเพลย์ลิสต์เฉพาะบุคคล เมื่อบุคคลหนึ่งฟังเพลงอารมณ์คล้าย ๆ กันซ้ำๆ จะแนะนำเพลงที่อารมณ์ 1 คล้ายคลึงกันมาให้ บางเพลง เนื้อหาที่มีความแตกต่างกัน เช่น เพลงสนุกเนื้อหาเศร้า, เพลงสบายๆเนื้อหากระตุ้นเร้าใจ ทำให้เกิดความซับซ้อนทางอารมณ์ เพลงซ้ำที่เนื้อหาสนุก อาจทำให้เรามีความสุข เราจึงสร้างโมเดลทำนายอารมณ์ของเพลงใหม่ โดยพิจารณาทั้งสององค์ประกอบหลักได้แก่ เสียงเพลง และเนื้อเพลง เพื่อจำแนกว่าเพลงนั้นทำนองเพลงและเนื้อเพลงอยู่ในอารมณ์ใดดังต่อไปนี้ เพลงที่ให้อารมณ์เศร้า เพลงที่ให้อารมณ์สงบ เพลงที่ให้อารมณ์คุ้นเร้าใจ เพลงที่ให้อารมณ์มีความสุข โดยมีวัตถุประสงค์ดังนี้

1. เพื่อต้องการศึกษาโมเดลและวิธีการทำงานของแต่ละ โมเดล เพื่อเปรียบเทียบประสิทธิภาพการทำงานในการจำแนก
2. วิเคราะห์ความสัมพันธ์ระหว่างอารมณ์ของทำนองเพลงและเนื้อเพลงไทย

3. สร้างโมเดลที่สามารถประเมินอารมณ์เพลงโดยอัตโนมัติจากทำนองและเนื้อเพลงไทย

2. ทฤษฎี และความรู้ที่เกี่ยวข้อง

2.1 Russell's Circumplex Model of Affect หรือ โมเดลวงล้ออารมณ์ของรัสเซลล์

เป็นแบบจำลองทางจิตวิทยาที่อธิบาย และจัดกลุ่มอารมณ์ของมนุษย์ในลักษณะสองมิติ โดยเน้นถึงความสัมพันธ์ระหว่างอารมณ์ต่าง ๆ โมเดลนี้ถูกพัฒนาโดย James A. Russell ในปี 1980 และได้รับความนิยมอย่างมากในด้านจิตวิทยา การวิจัยด้านอารมณ์ และการวิเคราะห์ข้อมูลที่เกี่ยวข้องกับอารมณ์ เช่น คนตรีหรือข้อความ โดยแกน Arousal สื่อถึงระดับความตื่นตัวทางอารมณ์ และแกน Valence สื่อถึงอารมณ์เชิงบวกและเชิงลบ

2.2 Web Scraping

เทคนิคและวิธีในการเข้าถึงข้อมูลจำนวนมากมหาศาลจากเว็บไซต์ต่าง ๆ ที่ต้องการ โดยไม่จำเป็นต้องคัดลอกข้อมูลเหล่านั้นมาจากเว็บไซต์ เทคนิคนี้จะวิเคราะห์ลักษณะของภาษา HTML ในบทความนี้ได้มีการใช้ library ดังต่อไปนี้

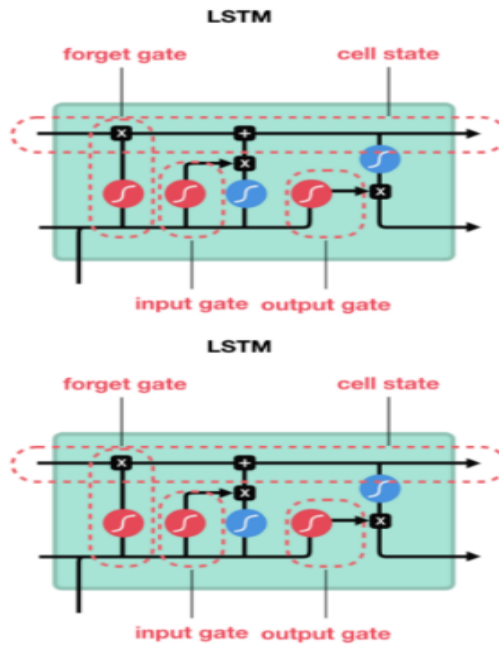
2.2.1 Selenium Selenium เป็นไลบรารีที่ใช้ในการทำ Web Scraping โดยสามารถควบคุมเบราว์เซอร์ และจัดการกับเว็บไซต์ที่มี Dynamic Content (ซึ่งสร้างด้วย JavaScript) ได้ เช่น การคลิก พิมพ์ข้อความ เลื่อนหน้าเว็บ และดึงข้อมูลจากองค์ประกอบ HTML ได้ นอกจากนี้ยังรองรับการใช้งาน Headless Browser เพื่อเพิ่มความเร็วและประหยัดทรัพยากร

2.2.2 ChromeDriverManager ช่วยในการติดตั้งและอัปเดต ChromeDriver โดยอัตโนมัติ ซึ่งเป็นเครื่องมือที่จำเป็นในการใช้ Selenium กับ Google Chrome

2.2.3 TQDM ใช้สำหรับสร้าง Progress Bar ซึ่งช่วยแสดงความคืบหน้าในระหว่างการทำงาน เช่น การโหลดหน้าเว็บหรือการดึงข้อมูลจากหลายหน้าเว็บ

2.3 RNN (Recurrent Neural Network) แบบ LSTM (Long Short-Term Memory)

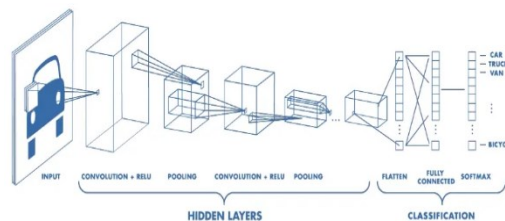
การประมวลผลใน RNN Cell จะมีการส่งผลลัพธ์ออกมาเป็น Vector $h_0, h_1, h_2, \dots, h_t$ ตามลำดับ โดยในการประมวลผลจะมีการนำ Output State ของ RNN Cell ในรอบก่อนหน้ามาเป็น Input State ของ RNN Cell ในรอบถัดไป ภายใน RNN Cell จะมีวงจรสำหรับนำ Input Data x Weight ในรูปแบบต่างๆซึ่งปัจจุบันมี RNN Cell ที่นิยมอยู่ 2 รูปแบบ ได้แก่ LSTM และ GRU โดยบทความนี้เลือกใช้รูปแบบ LSTM ซึ่งเป็นโครงข่ายประสาทเทียมแบบหนึ่ง ที่ถูกออกแบบมาสำหรับการประมวลผลลำดับ (Sequence) LSTM จัดว่าเป็นโครงข่ายประสาทเทียมประเภท Recurrent Neural Network (RNN) ดังภาพที่ 1

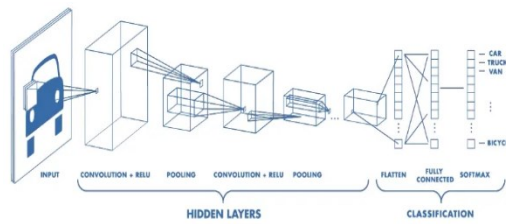


ภาพที่ 1 แบบจำลองพื้นฐานของ LSTM

2.4 Convolutional Neural Network (CNN)

การประมวลผลจะใช้การคำนวณ Convolution Operation ระหว่าง Input Data และ Kernel (หรือ Filter) เพื่อดึงลักษณะสำคัญ (Features) ของข้อมูลออกมา ภายใน CNN Layer จะมีวงจรที่ประกอบด้วยการคูณ Kernel กับ Input Data และตามด้วยการใช้ฟังก์ชัน Activation เช่น ReLU (Rectified Linear Unit) ซึ่งปัจจุบัน CNN มีโครงสร้างสำคัญที่นิยมใช้ในหลายรูปแบบ ได้แก่ AlexNet, VGG, ResNet และ Inception โดยบทความนี้เลือกใช้โครงสร้างพื้นฐานของ CNN ซึ่งเป็นโครงข่ายประสาทเทียมที่ถูกรวบรวมมาสำหรับการประมวลผลข้อมูลที่มีลักษณะเชิงพื้นที่ (Spatial Data) เช่น รูปภาพ หรือวิดีโอ CNN จัดว่าเป็นโครงข่ายประสาทเทียมประเภท Feedforward Neural Network ที่เน้นการประมวลผลข้อมูลแบบเชิงพื้นที่ (Spatial Hierarchy) เพื่อลดมิติข้อมูลและดึงลักษณะสำคัญในลำดับชั้น (Hierarchical Feature Extraction) ดังภาพที่ 2





ภาพที่ 2 โครงสร้างพื้นฐานของ Convolutional Neural Network (CNN)

2.5 WangchanBERTa (Wangchan Bidirectional Encoder Representations from Transformers)

WangchanBERTa ประมวลผลข้อความโดยใช้ Self-Attention Mechanism เพื่อตรวจจับความสัมพันธ์ระหว่างคำและสร้างบริบทที่ลึกซึ้ง (Contextualized Representations) แต่ละคำในประโยคผลลัพธ์ที่ได้คือเวกเตอร์ Embeddings ที่สะท้อนความหมายของคำในบริบทนั้น โมเดลใช้โครงสร้าง BERT ที่ปรับแต่งเฉพาะสำหรับภาษาไทย เช่น การจัดการคำที่ไม่มีการเว้นวรรค และปรับด้วยฟังก์ชัน Activation อย่าง GeLU เพื่อเพิ่มประสิทธิภาพ โมเดลนี้เหมาะสำหรับงานวิเคราะห์ภาษาไทย เช่น การจัดหมวดหมู่ข้อความและการตอบคำถาม

2.6 Mel Spectrogram

เป็นเครื่องมือในงานวิเคราะห์เสียงที่เปลี่ยนสัญญาณเสียง (Audio Signal) ให้อยู่ในรูปแบบภาพ โดยเน้นการแสดงผลที่สอดคล้องกับการได้ยินของมนุษย์ (Human Hearing Perception) ซึ่งมักใช้ในงานวิจัยด้านเสียง การจำแนกเสียง หรือการประมวลผลเสียงในระบบ AI

2.7 MFCC

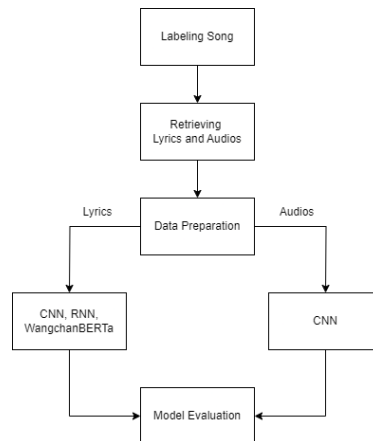
การสกัดลักษณะเด่น เป็นการดึงลักษณะเฉพาะของหน่วยเสียงแต่ละหน่วยที่แตกต่างกันออกมา แล้วให้ระบบทำการรู้จำลักษณะของหน่วยเสียงแต่ละหน่วยเสียงไว้ เมื่อสัญญาณที่เข้ามาภายหลังมีลักษณะเด่นที่เหมือนหรือใกล้เคียงกับลักษณะกับลักษณะเด่นของหน่วยเสียงใด ระบบรู้จำจะสามารถบอกได้ว่าเป็นหน่วยเสียงใดหรือใกล้เคียงกับหน่วยเสียงกลุ่มใดมากที่สุด และสามารถลดจำนวนข้อมูลจำนวนมากจะถูกแปลงเป็นชุดข้อมูลที่มีจำนวนน้อยลงและยังคงคุณสมบัติสำคัญของข้อมูลไว้ได้อย่างถูกต้อง

2.8 Chromagram

เกี่ยวข้องกับ pitch class 12 class ที่แตกต่างกัน คุณลักษณะที่อิงตาม Chroma ซึ่งเรียกอีกอย่างว่า “pitch class profiles” Pitch = ความถี่ของคลื่นเสียง Chromagram เป็นเครื่องมือที่มีประสิทธิภาพสำหรับการวิเคราะห์เพลงซึ่งสามารถจัดหมวดหมู่ระดับเสียงที่มีความหมายได้ (มักเป็นสิบสองประเภท) และการปรับจูนนั้นใกล้เคียงกับระดับอารมณ์ที่เท่ากัน คุณสมบัติหลักประการหนึ่งของคุณสมบัติของโครมาคือ

คุณสมบัติเหล่านี้จับลักษณะฮาร์โมนิกและไพลอะของดนตรี ในขณะที่ยังคงแข็งแกร่งต่อการเปลี่ยนแปลงของเสียงต่ำและเครื่องมือวัด ซึ่งเราสามารถใช้ librosa แยกและแสดง Chromagram ของเสียงได้

3. วิธีการดำเนินงาน



ภาพที่ 3 ภาพรวมกระบวนการ

3.1 การกำหนด และติดป้ายรายชื่อเพลง

ในขั้นตอนแรก ผู้วิจัยได้ทำการแบ่งคลาสของอารมณ์เพลงเป็น 4 คลาส ซึ่งเป็นคลาสเดียวกันทั้งทางด้านเสียงและเนื้อเพลง ได้แก่ Happy, Angry, Sad และ Calm เพื่อใช้ในการเปรียบเทียบในภายหลังว่าทั้งด้านเสียงและเนื้อเพลงให้อารมณ์สอดคล้องกันหรือไม่ ซึ่งยึดตามทฤษฎี Russell's Circumplex Model โดยเป็นรายชื่อเพลงสำหรับสร้างโมเดลจำแนกอารมณ์ของเสียงเพลงคลาสละ 100 เพลง และรายชื่อเพลงสำหรับสร้างโมเดลจำแนกอารมณ์ของเนื้อเพลงคลาสละ 50 เพลง โดยเงื่อนไขในการ label ว่าเพลงใดเป็นคลาสใดทางผู้วิจัยใช้การตัดสินใจร่วมกันภายในทีม โดยใช้ดุลยพินิจของคณะผู้วิจัย

3.2 การดึงข้อมูล

ในด้านเสียงเพลง ทางผู้วิจัยได้ทำการดึงเสียงเพลงจากเว็บไซต์ YouTube ผ่านทาง url และบันทึกเป็นไฟล์ .mp3 แยกตามไฟล์เดอร์ของแต่ละคลาส และในด้านเนื้อเพลง ทางผู้วิจัยได้ทำการดึงเนื้อเพลงจากเว็บไซต์ LyricFind ผ่านทาง url โดยใช้เทคนิค Web Scraping และบันทึกเป็นไฟล์ .txt แยกไฟล์เดอร์ตามคลาส

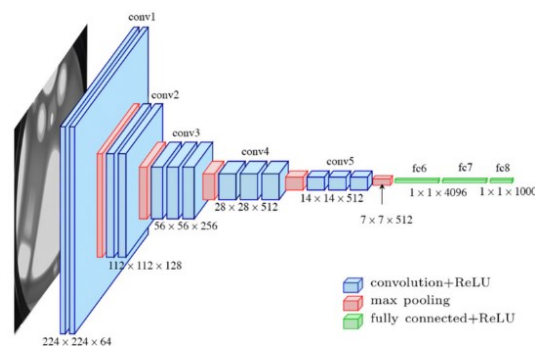
3.3 การเตรียมข้อมูล

ในด้านเสียงเพลง เพื่อให้ได้ข้อมูลที่มีขนาดเท่ากันในขณะที่สร้างโมเดล ผู้ดำเนินงานได้ตัดเสียงจากไฟล์ .mp3 ให้เหลือเพียงเพลงละ 30 วินาที โดยเลือกเฉพาะช่วงที่มีเสียงเพลงตรงตาม label เพื่อไม่ให้เกิด

noise ของข้อมูล และในด้านเนื้อเพลง ผู้วิจัยได้ทำการแบ่งท่อนของแต่ละเพลง โดยกำหนดให้แบ่งทุก ๆ 3 บรรทัด และมีการทำความสะอาดข้อมูล เช่น การตัดคำภาษาอังกฤษออก การลบอักขระพิเศษ เป็นต้น

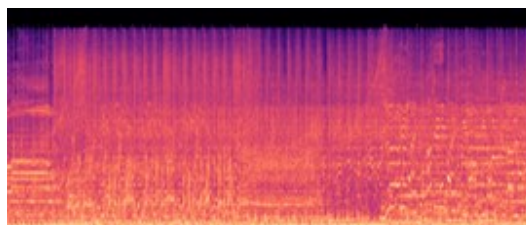
3.4 การสร้าง และคัดเลือกโมเดล

ในด้านเสียงเพลง โมเดลที่เลือกใช้ คือ CNN โดยใช้ vgg16 เป็นเลเยอร์หลักสำหรับดึงฟีเจอร์ มีการเพิ่มเลเยอร์ global_average_pooling2d_4 เพื่อลดมิติข้อมูล จาก 3D เป็น 1D, เลเยอร์ dense (256 Units) สำหรับการประมวลผลเพิ่มเติมโดยใช้ ReLU Activation, และเลเยอร์ dense (4 Unit ตามจำนวนคลาส) เป็นเลเยอร์สุดท้ายโดยใช้ Softmax Activation

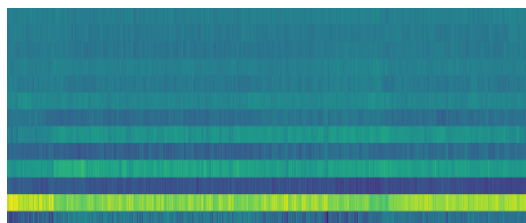


ภาพที่ 4 แสดงโครงสร้างโมเดล CNN (vgg16 base model)

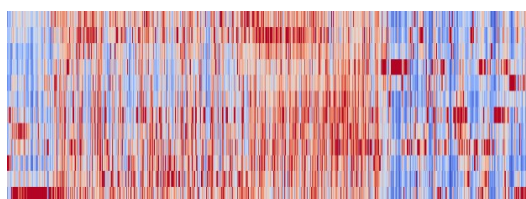
โดยฟีเจอร์ที่นำมาใช้ในการสร้างโมเดล CNN มีดังนี้ Mel Spectrogram, MFCC และ Chromagram ซึ่งเป็นฟีเจอร์ที่อยู่ในรูปแบบรูปภาพ โดยทั้ง 3 ฟีเจอร์มีการกำหนดไฮเปอร์พารามิเตอร์สำคัญเหมือนกัน ได้แก่ epoch เท่ากับ 100, batch size เท่ากับ 16 และใช้ Optimizer คือ Adam ที่ learning rate เท่ากับ 0.0001



ภาพที่ 5 Mel Spectrogram



ภาพที่ 6 MFCC



ภาพที่ 7 Chromagram

ในส่วนของเนื้อเพลง หลังจากทำความเข้าใจเนื้อหาของเนื้อเพลงแล้ว จะนำเนื้อเพลงมาตัดคำ โดยใช้ tokenized ที่ engine='newmm' ดังภาพที่ 8 ตัดคำดังภาพที่ 9 และแปลงให้อยู่ในรูปของตัวเลข หรือเรียกว่า การทำ Encoding ดังภาพที่ 10

	emotion	lyrics
703	happy	แต่ในวันนี้อยากกอดเธอเกินใคร เป็นอะไรที่มันรักจริง
462	happy	และขอบใจเธอตอนอยู่กับฉัน ขอบเวลาที่เราได้รักกัน
139	happy	เพราะคิดว่ารักเธอหมดตัวเธอคงต้องใจอ่อน ถ้างั้น...
337	happy	ยาวนาน ความฝันของฉันหลุดลอยไป ลอยหายในจักรวาล
93	happy	เหมือนฉันได้กลายมาเป็นอีกคนที่ดีกว่าเดิม ชอบดี...
...

ภาพที่ 8 หลังจากทำความเข้าใจเนื้อเพลง

เพลงที่ 1: แต่ ใน วันนี้ อยาก กอด เธอ เกิน ใคร เป็น อะไร ที่ มัน รัก จริง จริง ที่ เรื่องราว ของ คน นิ่ง แต่ ได้ เป็น สิบ เพลง
เพลงที่ 2: และ ขอบ คำ เธอ ตอน อยู่ กับ ฉัน ขอบ เวลา ที่ เรา ได้ รัก กัน ยาว นาน ที่ ฉัน นิ่ง ได้ รัก เธอ อยู่
เพลงที่ 3: เพราะ คิด ว่า รัก เธอ หมด ตัว เธอ คง ต้อง ใจอ่อน ถ้างั้น ฉัน甘 เธอ ดี คง ดี ชื่น ชอบ เข้ามา ได้ ไหม

ภาพที่ 9 ตัดคำ

แต่ ใน วันนี้ อยาก กอด เธอ เกิน ใคร เป็น อะไร ที่ มัน รัก จริง จริง
[22, 20, 104, 30, 185, 1, 248, 23, 14, 43, 4, 12, 1159, 67, 67, 4,

ภาพที่ 10 การทำ Encoding

จากนั้นนำไปทำโมเดลทั้ง 3 โมเดล ได้แก่ Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) แบบ Long short-term memory (LSTM) และ WangchanBERTa และในขั้นตอนสุดท้ายนำผลการทดสอบโมเดลมาประเมินประสิทธิภาพและเปรียบเทียบระหว่างโมเดล

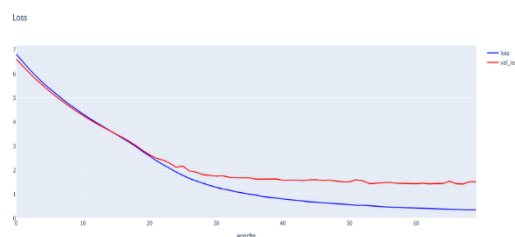
4. ผลการดำเนินงาน

4.1 ผลการทดสอบโมเดลโดยแยกอารมณ์ของเนื้อเพลง

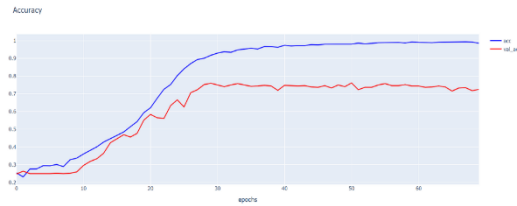
จากการทดสอบโมเดลการจำแนกอารมณ์ของเนื้อเพลงโดยใช้ทดสอบกับชุดฝึก 200 ชุด แบ่งเป็น class ละ 50 เพลง โดยผู้วิจัยมีการทดสอบประสิทธิภาพของโมเดลดังต่อไปนี้

4.1.1 ผลการทดสอบโมเดล Recurrent Neural Network (RNN) แบบ Long short-term memory (LSTM)

กำหนดค่า Epochs เท่ากับ 70, Batch size เท่ากับ 64 และ Learning rate เท่ากับ 0.0001 ในการทดสอบโมเดล สามารถสังเกตค่า Loss ในกราฟของภาพที่ 11 และ Accuracy ในกราฟของภาพที่ 12



ภาพที่ 11 Loss Plot ของโมเดล RNN แบบ LSTM



ภาพที่ 12 Accuracy Plot ของโมเดล RNN แบบ LSTM

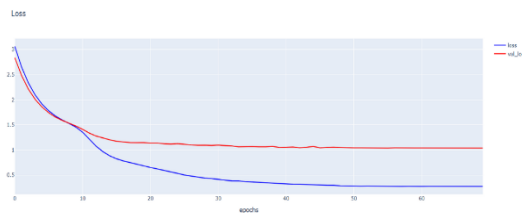
	precision	recall	f1-score	support
sad	0.6267	0.7176	0.6690	131
calm	0.7750	0.7099	0.7410	131
happy	0.7874	0.7634	0.7752	131
angry	0.7717	0.7481	0.7597	131
accuracy			0.7347	524
macro avg	0.7402	0.7347	0.7362	524
weighted avg	0.7402	0.7347	0.7362	524

ตารางที่ 1

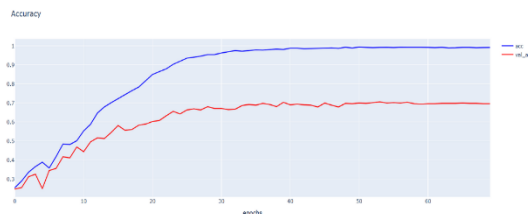
จากตาราง 1 การจำแนกอารมณ์ของเนื้อเพลงด้วยเทคนิค Recurrent Neural Network (RNN) แบบ Long short-term memory (LSTM) ผลทดสอบประเมินประสิทธิภาพโมเดล ได้ 0.7347 หรือประมาณ 73.47%

4.1.2 ผลการทดสอบโมเดล Convolutional Neural Network (CNN)

กำหนดค่า Epochs เท่ากับ 70, Batch size เท่ากับ 64 และ Learning rate เท่ากับ 0.0001 ในการทดสอบโมเดล สามารถสังเกตค่า Loss ในกราฟของภาพที่ 13 และ Accuracy ในกราฟของภาพที่ 14



ภาพที่ 13 Loss Plot ของโมเดล CNN



ภาพที่ 14 Accuracy Plot ของโมเดล CNN

	precision	recall	f1-score	support
angry	0.7414	0.6565	0.6964	131
calm	0.6601	0.7710	0.7113	131
sad	0.6691	0.6947	0.6816	131
happy	0.7479	0.6794	0.7120	131
accuracy		0.7004		524
macro avg	0.7046	0.7004	0.7003	524
weighted avg	0.7046	0.7004	0.7003	524

ตารางที่ 2

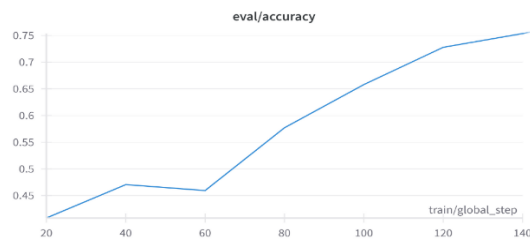
จากตาราง 2 การจำแนกอารมณ์ของเนื้อเพลงด้วยผลทดสอบประเมินประสิทธิภาพโมเดล Convolutional Neural Network (CNN) ได้ 0.7004 หรือประมาณ 70.04%

4.1.3 ผลการทดสอบโมเดล WangchanBERTa

กำหนดค่า Epochs เท่ากับ 6 Batch size เท่ากับ 64 Learning rate เท่ากับ 0.0001 Warmup step เท่ากับ 100 Weight decay เท่ากับ 0.2 และ Adam epsilon เท่ากับ 0.00000001 ในการฝึกโมเดล สามารถสังเกตค่า Training Loss และ Validation Loss ในกราฟของภาพที่ 15 และ Accuracy ในกราฟของภาพที่ 16



ภาพที่ 15 Training Loss vs Validation Loss Plot ของโมเดล WangchanBERTa



ภาพที่ 16 Accuracy Plot ของโมเดล WangchanBERTa

You are given a cropped image of a single layout block and an initial HTML representation of its content.										
Step	Training Loss	Validation Loss	Accuracy	F1 Micro	Precision Micro	Recall Micro	F1 Macro	Precision Macro	Recall Macro	Mo Samples
20	1.396800	1.316286	0.409088	0.408088	0.408088	0.408088	0.381562	0.441400	0.399161	272
40	1.258100	1.208405	0.470588	0.470588	0.470588	0.470588	0.451877	0.497525	0.487240	272
60	1.199900	1.203758	0.499559	0.499559	0.499559	0.499559	0.488710	0.484117	0.454448	272
80	1.085300	1.008187	0.577206	0.577206	0.577206	0.577206	0.574806	0.586072	0.575450	272
100	0.915600	0.870116	0.658088	0.658088	0.658088	0.658088	0.636741	0.677446	0.648956	272
120	0.635400	0.730797	0.727941	0.727941	0.727941	0.727941	0.724746	0.733991	0.726249	272
140	0.548900	0.673155	0.753676	0.753676	0.753676	0.753676	0.748559	0.753341	0.749488	272
160	0.295900	0.697142	0.763088	0.763088	0.763088	0.763088	0.776252	0.784555	0.780810	272
180	0.195200	0.690289	0.763088	0.763088	0.763088	0.763088	0.779134	0.781301	0.780810	272
200	0.113100	0.698444	0.779735	0.779735	0.779735	0.779735	0.779735	0.779216	0.779304	272

ตารางที่ 3 ผลลัพธ์การ Train

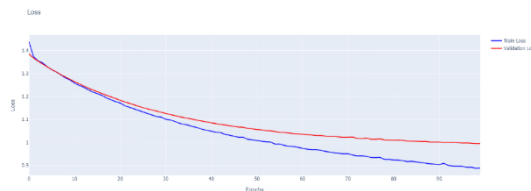
จากตารางที่ 3 การจำแนกอารมณ์ของเนื้อเพลงด้วยผลทดสอบประเมินประสิทธิภาพโมเดล WangchanBERTa ที่ Step ที่ 140 มีความเหมาะสมต่อการจำแนกเนื้อเพลงมากที่สุด โดยมีความแม่นยำ 0.753676 หรือประมาณ 75.37%

4.2 ผลการทดสอบโมเดลโดยแยกอารมณ์ของทำนองเพลง

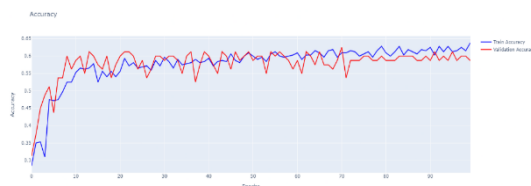
จากการทดสอบโมเดลการจำแนกอารมณ์ของเนื้อเพลงโดยใช้ทดสอบกับชุดฝึก 400 ชุด แบ่งเป็น class ละ 100 เพลง โดยผู้วิจัยมีการทดสอบประสิทธิภาพของโมเดลดังต่อไปนี้

4.2.1 ผลการทดสอบโมเดล *Convolutional Neural Network (CNN)* ด้วย *Mel Spectrogram*

กำหนดค่า Epochs เท่ากับ 100, Batch size เท่ากับ 16 และ Learning rate เท่ากับ 0.0001 ในการทดสอบโมเดล สามารถสังเกตค่า Loss ในกราฟของภาพที่ 17 และ Accuracy ในกราฟของภาพที่ 18



ภาพที่ 17 Loss Plot ของโมเดล CNN ด้วย Mel Spectrogram



ภาพที่ 18 Accuracy Plot ของโมเดล CNN Mel Spectrogram

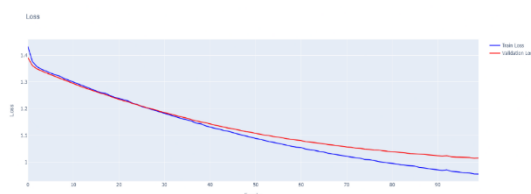
	precision	recall	f1-score	support
Happy	0.53	0.50	0.51	20
Sad	0.57	0.60	0.59	20
Angry	0.77	0.85	0.81	20
Calm	0.50	0.45	0.47	20
accuracy			0.60	80
macro avg	0.59	0.60	0.60	80
weighted avg	0.59	0.60	0.60	80

ตารางที่ 4

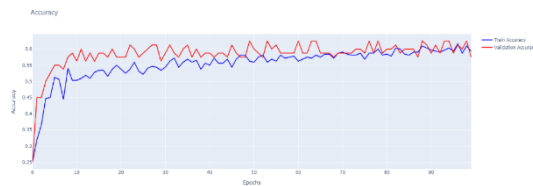
จากตารางที่ 4 การจำแนกอารมณ์ของทำนองด้วยผลทดสอบประเมินประสิทธิภาพโมเดล Convolutional Neural Network (CNN) ด้วย ด้วย Mel Spectrogram ได้ 0.60 หรือประมาณ 60.00%

4.2.2 ผลการทดสอบโมเดล *Convolutional Neural Network (CNN)* ด้วย *MFCC*

กำหนดค่า Epochs เท่ากับ 100 Batch size เท่ากับ 16 และ Learning rate เท่ากับ 0.0001 ในการทดสอบโมเดล สามารถสังเกตค่า Loss ในกราฟของภาพที่ 19 และ Accuracy ในกราฟของภาพที่ 20



ภาพที่ 19 Loss Plot ของโมเดล CNN ด้วย MFCC



ภาพที่ 20 Accuracy Plot ของโมเดล CNN ด้วย MFCC

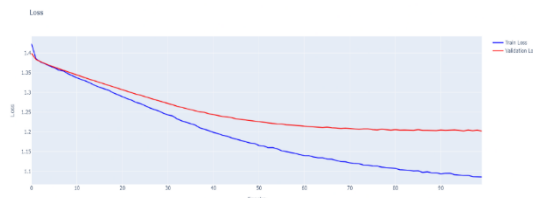
	precision	recall	f1-score	support
Happy	0.53	0.50	0.51	20
Sad	0.55	0.55	0.55	20
Angry	0.77	0.85	0.81	20
Calm	0.47	0.45	0.46	20
accuracy			0.59	80
macro avg	0.58	0.59	0.58	80
weighted avg	0.58	0.59	0.58	80

ตารางที่ 5

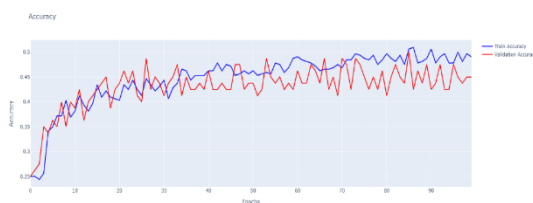
จากตารางที่ 5 การจำแนกอารมณ์ของทำนองด้วยผลทดสอบประเมินประสิทธิภาพโมเดล Convolutional Neural Network (CNN) ด้วย MFCC ได้ 0.59 หรือประมาณ 59.00%

4.2.3 ผลการทดสอบโมเดล Convolutional Neural Network (CNN) ด้วย Chromagram

กำหนดค่า Epochs เท่ากับ 100, Batch size เท่ากับ 16 และ Learning rate เท่ากับ 0.0001 ในการทดสอบโมเดล สามารถสังเกตค่า Loss ในกราฟของภาพที่ 21 และ Accuracy ในกราฟของภาพที่ 22



ภาพที่ 21 Loss Plot ของโมเดล CNN ด้วย Chromagram



ภาพที่ 22 Accuracy Plot ของโมเดล CNN ด้วย Chromagram

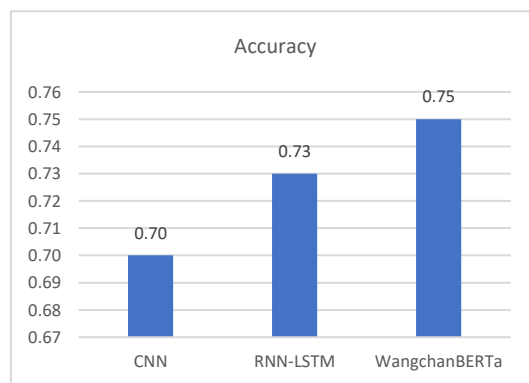
	precision	recall	f1-score	support
Happy	0.38	0.30	0.33	20
Sad	0.60	0.45	0.51	20
Angry	0.57	0.65	0.60	20
Calm	0.46	0.60	0.52	20
accuracy			0.50	80
macro avg	0.50	0.50	0.49	80
weighted avg	0.50	0.50	0.49	80

ตารางที่ 6

จากตารางที่ 6 การจำแนกอารมณ์ของทำนองด้วยผลทดสอบประเมินประสิทธิภาพโมเดล Convolutional Neural Network (CNN) ด้วย Chromagram ได้ 0.50 หรือประมาณ 50.00%

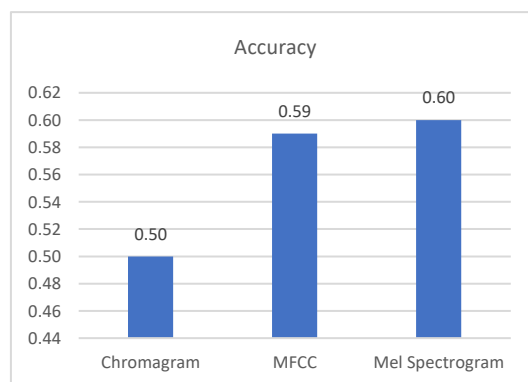
5. การอภิปรายผล

ผลการทดสอบโมเดลจำแนกอารมณ์ของเนื้อเพลงพบว่าโมเดล CNN มีค่า Accuracy เท่ากับ 0.70 โมเดล RNN แบบ LSTM มีค่า Accuracy เท่ากับ 0.73 และโมเดล WangchanBERTa มีค่า Accuracy สูงที่สุด อยู่ที่ 0.75 ดังภาพที่ 23



ภาพที่ 23 การเปรียบเทียบค่า Accuracy ในการจำแนกเนื้อเพลงของโมเดลแต่ละประเภท

ผลการทดสอบโมเดลจำแนกอารมณ์ของเสียงเพลง โดยใช้โมเดล CNN เป็นการทดสอบด้วยการ สกัดพีเจอร์ที่แตกต่างกัน 3 เทคนิค ได้แก่ Mel Spectrogram, MFCC และ Chromagram พบว่า Chromagram มีค่า Accuracy เท่ากับ 0.50 MFCC มีค่า Accuracy เท่ากับ 0.59 และ Mel Spectrogram มีค่า Accuracy สูงที่สุด อยู่ที่ 0.60 ดังภาพที่ 24



ภาพที่ 24 การเปรียบเทียบค่า Accuracy ในการจำแนกเสียงเพลงของเทคนิคแต่ละประเภท

ดังนั้นผู้ดำเนินการวิจัยจึงตัดสินใจเลือกโมเดล WangchanBERTa และเทคนิค Mel Spectrogram ในการนำไปจำแนกเนื้อเพลง และเสียงเพลง เนื่องจากมีค่า Accuracy สูงที่สุด

จากนั้นผู้ทำการทดลองได้คัดเลือกเพลงไทยจำนวน 50 เพลงพร้อมติดป้ายกำกับ เพื่อใช้โมเดลที่ได้คัดเลือกว่าเหมาะสมที่สุด จำแนกอารมณ์ของทำนองเพลง และเนื้อเพลง ในบทความนี้กำหนดหลักเกณฑ์ความสอดคล้องโดยอ้างอิงจากโมเดลวงล้ออารมณ์ 2 มิติของรัสเซล ที่มีแกน Arousal(Angry และ Calm) และ Valence(Sad และ Happy) โดย Aligned หมายถึง ทำนองเพลง และเนื้อเพลงมีอารมณ์ที่เหมือนกัน ตั้งอยู่บนแกนเดียวกัน ทิศทางเดียวกัน และถือว่ามีความสอดคล้องกัน Neutral หมายถึง ทำนองเพลง และเนื้อเพลงไม่ได้ตั้งอยู่บนแกนอารมณ์เดียวกัน และ Contrast หมายถึง ทำนองเพลง และเนื้อเพลงตั้งอยู่บนแกนอารมณ์เดียวกัน แต่อยู่ในทิศทางตรงกันข้าม ซึ่งถือว่าไม่สอดคล้องกัน

Index Label	Text Class	Lyrics	Music Class	Music	Align	Contrast
1. เพลงพญานาค (วงคาราบาว)	Angry	Angry	Angry	Angry	Aligned	Aligned
2. เพลงคนจน (วงคาราบาว)	Sad	Sad	Angry	Angry	Contrast	Contrast
3. เพลงรัก (วงคาราบาว)	Sad	Sad	Angry	Angry	Contrast	Contrast
4. เพลงคนจน (วงคาราบาว)	Angry	Sad	Angry	Angry	Contrast	Contrast
5. เพลงคนจน (วงคาราบาว)	Angry	Sad	Angry	Angry	Contrast	Contrast
6. เพลงคนจน (วงคาราบาว)	Sad	Angry	Sad	Sad	Aligned	Aligned
7. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
8. เพลงคนจน (วงคาราบาว)	Sad	Sad	Angry	Angry	Contrast	Contrast
9. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
10. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
11. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
12. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
13. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
14. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
15. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
16. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
17. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
18. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
19. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
20. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
21. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
22. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
23. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
24. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
25. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
26. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
27. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
28. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
29. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
30. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
31. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
32. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
33. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
34. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
35. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
36. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
37. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
38. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
39. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned
40. เพลงคนจน (วงคาราบาว)	Sad	Sad	Sad	Sad	Aligned	Aligned

ภาพที่ 25 ภาพตารางการทดสอบ

ผลการทดสอบพบว่าสามารถจำแนกเพลงที่มีอารมณ์ตรงข้ามกันของทำนองเพลง และเนื้อเพลงแม่นยำ 60%

6. สรุปผล

บทความนี้ได้กำหนดอารมณ์ทั้ง 4 Angry Calm Happy และ Sad โดยอ้างอิงโมเดลวงล้ออารมณ์ของรัสเซล นำมาใช้จำแนกองค์ประกอบของเพลง คือ ทำนองเพลง และเนื้อเพลง โดยมีการเก็บข้อมูลเพลงไทยด้วยวิธี Web scraping ทำการทดลองเปรียบเทียบโมเดล RNN แบบ LSTM CNN และโมเดลที่ผ่านการ Fine-tune ด้วย WangchanBERTa และเทคนิคการสกัดฟีเจอร์เสียงในโมเดล CNN 3 เทคนิค Mel Spectrogram Chromagram และ MFCC พบว่าโมเดล และเทคนิคที่มีความเหมาะสมที่สุด ได้แก่ โมเดลที่ผ่านการ Fine-tune ด้วย WangchanBERTa มีความแม่นยำอยู่ที่ร้อยละ 75 และเทคนิค Mel Spectrogram มีความแม่นยำอยู่ที่ร้อยละ 60 เมื่อนำโมเดลมาใช้ในการเปรียบเทียบระหว่างอารมณ์ของทำนองเพลง และเนื้อเพลงจำนวน 50 เพลงพบว่าสามารถจำแนกเพลงที่มีอารมณ์ของทำนองเพลง และเนื้อเพลงที่ไม่สอดคล้องกันมีความแม่นยำอยู่ที่ร้อยละ 60 ซึ่งโมเดลสำหรับการจำแนกนี้สามารถนำไปต่อยอดในการสร้างเพลย์ลิสต์อัตโนมัติ หรือการแนะนำเพลงที่ตรงต่ออารมณ์ที่ผู้ฟังรู้สึกอยู่ในขณะนั้น เพื่อตอบสนองความต้องการของผู้ฟังได้ดียิ่งขึ้น

เอกสารอ้างอิง

- [1] Russell's Circumplex Model, “แบบจำลองสำหรับการจำแนกอารมณ์และเนื้อเพลง ซึ่งจัดกลุ่มอารมณ์ในรูปแบบวงกลมสองมิติ,” [ออนไลน์], 2567, [สืบค้นวันที่ 20 พฤศจิกายน 2567], จาก https://archive.lib.cmu.ac.th/full/T/2557/enco50757wkv_ch2.pdf
- [2] อ.ดร.ณัฐโชติ พรหมฤทธิ์, “การวิเคราะห์ความรู้สึก (Sentiment Analysis),” [ออนไลน์], 2567, [สืบค้นวันที่ 20 พฤศจิกายน 2567], จาก <https://blog.pjjop.org/sentiment-analysis-101>
- [3] Towards Data Science, “Web Scraping Using Selenium,” [ออนไลน์], 2567, [สืบค้นวันที่ 20 พฤศจิกายน 2567], จาก <https://towardsdatascience.com/how-to-use-selenium-to-web-scrape-with-example-80f9b23a843a>
- [4] Medium, “ทำโมเดล RNN แบบ LSTM วิเคราะห์เนื้อเพลงไทยและจัดประเภท,” [ออนไลน์], 2567, [สืบค้นวันที่ 20 พฤศจิกายน 2567], จาก <https://medium.com/@mkp8541/emotion-recognition-using-lstm-7fb52dc36388>
- [5] Sorapong Somsorn, “ทำโมเดล CNN จำแนกเสียงเพลง,” [ออนไลน์], 2567, [สืบค้นวันที่ 20 พฤศจิกายน 2567], จาก <https://sorapong.medium.com/audio-classification>
- [6] ResearchGate, “ทำโมเดล WangchanBERTa วิเคราะห์เนื้อเพลงไทยและจัดประเภท,” [ออนไลน์], 2567, [สืบค้นวันที่ 20 พฤศจิกายน 2567], จาก https://www.researchgate.net/publication/348757168_WangchanBERTa_Pretraining_transformer-based_Thai_Language_Models
- [7] Leland Robert, “Understanding the Mel-Spectrogram,” [ออนไลน์], 2567, [สืบค้นวันที่ 20 พฤศจิกายน 2567], จาก <https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53>
- [8] Next Software Thailand, “MFCC: การสกัดลักษณะเด่นของหน่วยเสียง,” [ออนไลน์], 2567, [สืบค้นวันที่ 20 พฤศจิกายน 2567], จาก <https://nextsoftwarethailand.com/mfcc/>
- [9] Sirasit Boonklang, “Chromagram: การจับลักษณะฮาร์โมนิกและไพอเราะของคนตรี,” BorntoDev Co., Ltd., [ออนไลน์], 2567, [สืบค้นวันที่ 20 พฤศจิกายน 2567], จาก <https://www.borntodev.com/2021/09/17/มาทำความรู้จักกับ-librosa/>