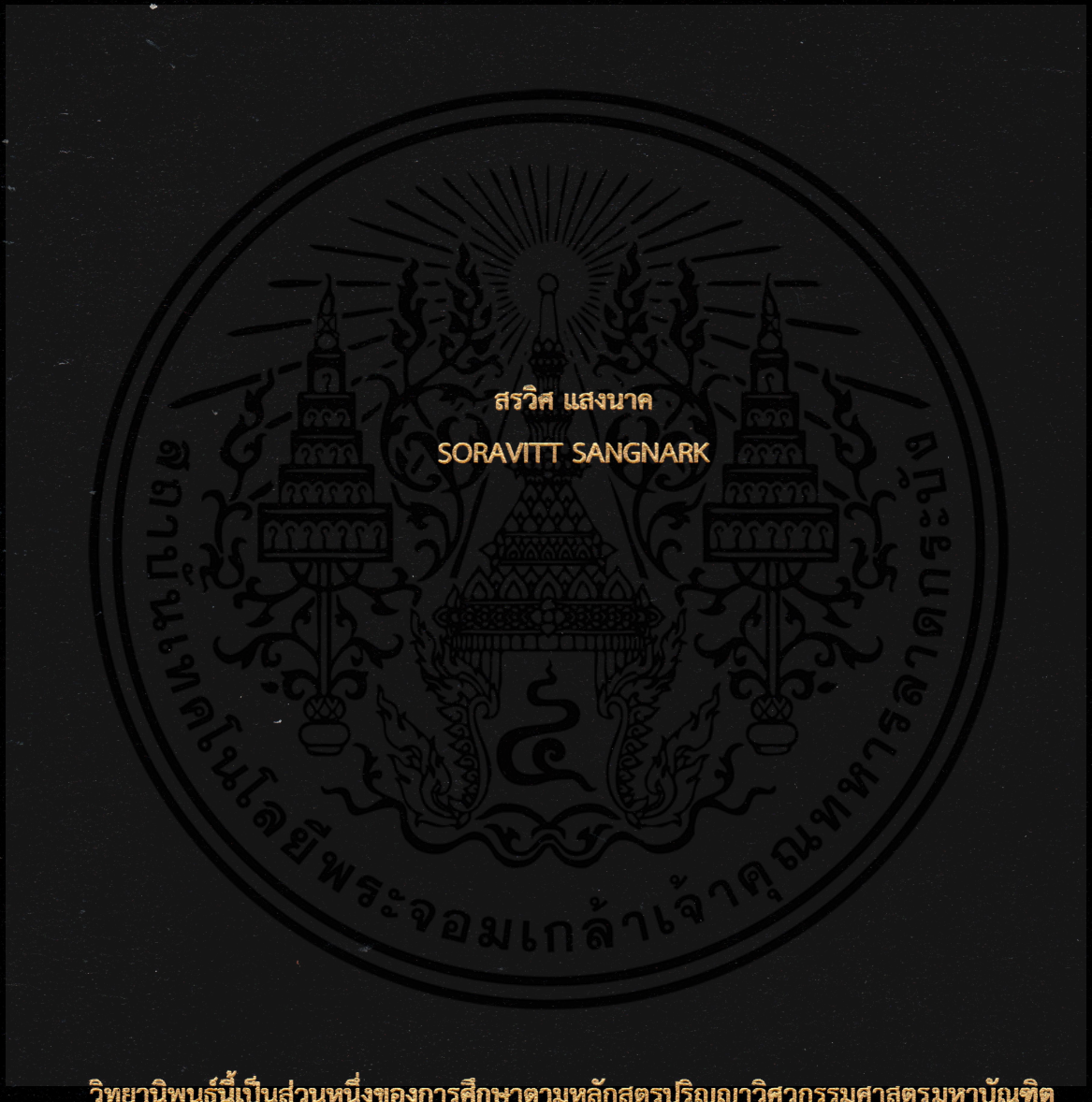
การรู้จำอารมณ์เพลงไทยโดยอ้างอิงโมเดลเพลงตะวันตก

THAI MUSIC EMOTION RECOGNITION BASED ON WESTERN MUSIC MODEL



สรวิศ แสงนาค

SORAVITT SANGNARK

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมสารสนเทศ

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2562

KMITL-2019-EN-M-230-003

การรู้จำอารมณ์เพลงไทยโดยอ้างอิงโมเดลเพลงตะวันตก

THAI MUSIC EMOTION RECOGNITION BASED ON WESTERN MUSIC MODEL

สรวิศ แสงนาค

SORAVITT SANGNARK

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมสารสนเทศ

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2562

KMITL-2019-EN-M-230-003

THAI MUSIC EMOTION RECOGNITION BASED ON WESTERN MUSIC MODEL

SORAVITT SANGNARK

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF ENGINEERING IN INFORMATION ENGINEERING

FACULTY OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2019

KMITL-2019-EN-M-230-003

หัวข้อวิทยานิพนธ์	การรู้จำอารมณ์เพลงไทยโดยอ้างอิงโมเดลเพลงตะวันตก
นักศึกษา	นายสรวิศ แสงนาค
รหัสประจำตัว	59601331
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมสารสนเทศ
พ.ศ.	2562
อาจารย์ที่ปรึกษาวิทยานิพนธ์	รศ.ดร.ชวลิต เบญจางคประเสริฐ

บทคัดย่อ

การรู้จำอารมณ์เพลง คือ การตรวจสอบอารมณ์เพลงโดยอ้างอิงจากคำนิยามอารมณ์ของผู้ฟัง ซึ่งงานวิจัยด้านนี้มีอัตราการเติบโตอย่างมากในปัจจุบัน เพราะพฤติกรรมการฟังเพลงของผู้ฟังที่เปลี่ยนแปลงไปอย่างต่อเนื่อง งานวิจัยด้านนี้นิยมศึกษาจากเพลงตะวันตก ในขณะที่เพลงไทยมีงานวิจัยที่น้อยมาก ดังนั้น วิทยานิพนธ์ฉบับนี้ได้เลือกศึกษาเพลงไทยซึ่งเป็นภาษาประจำชาติ และเป็นเพลงที่ได้รับความนิยมสูงสุดในพื้นที่การวิจัย โดยแบ่งออกเป็น 2 ส่วน ในส่วนแรก คือ ตรวจสอบอารมณ์เพลงไทยโดยโมเดลที่มีประสิทธิภาพสูงของเพลงตะวันตก และนำมาตรวจสอบว่าโมเดลที่มีประสิทธิภาพสูงสุดของเพลงไทยและเพลงตะวันตกเหมือนกันหรือแตกต่างกัน โดยหาค่าอารมณ์ Valence-Arousal ใช้วิธีการเรียนรู้ด้วยเครื่องแบบ Multiple linear regression และ K-Nearest neighbors ค่าอารมณ์เพลงไทยใช้ข้อมูลจาก Spotify API ผลลัพธ์ที่ได้ คือ อารมณ์ของเพลงไทยที่อ้างอิงจากคุณลักษณะเพลงตะวันตก โดยเพลงไทยแสดงค่า F-measure สูงสุด 41% จากโมเดล Multiple linear regression All model และเพลงตะวันตกแสดงค่า F-measure สูงสุด 51% จากโมเดล Multiple linear regression No Tempo model ซึ่งผลลัพธ์ทั้งสองมีความแตกต่างกันมาก เพราะโมเดลที่มีประสิทธิภาพสูงสุดและต่ำสุดของเพลงไทยและเพลงตะวันตก ให้ผลลัพธ์ที่แตกต่างกันโดยสิ้นเชิง และในส่วนที่สอง คือ วิเคราะห์อารมณ์เพลงไทย 155 เพลงโดยเฉพาะ ด้วยวิธี Multiple linear regression และ Support vector regression แบบ Linear ผลลัพธ์แสดงให้เห็นว่า Multiple linear regression มีค่าความถูกต้อง (Accuracy) 61.29%, ค่าความแม่นยำ (Precision) 65%, ค่าความระลึก (Recall) 61% และ F-measure 60% ซึ่งมีประสิทธิภาพสูงกว่า Support vector regression ทุกด้าน

Thesis	Thai Music Emotion Recognition Based on Western Music Model
Student	Mr.Soravitt Sangnark
Student ID.	59601331
Degree	Master of Engineering
Program	Information Engineering
Year	2019
Thesis Advisor	Assoc.Prof.Dr.Chawalit Benjangkaprasert

ABSTRACT

Music emotion recognition is the music emotion detected from people's annotations, which has rapidly grown in the present because the listener's behavior has been changing. Many research works have studied western music, but a few studied Thai music. Therefore, in this thesis, selected Thai songs for this work because the Thai language is native language and Thai song is most popular in the region of research. In this work, we divided into 2 parts. First, the Thai music was the evaluated set of a system based on western music training settings and investigated best model efficiency of Thai music and Western music have different or same. By using valence-arousal values, multiple linear regression, k-nearest neighbours to represent the emotional annotations from the music. We used valence and energy(arousal) from Spotify API to the investigated emotion of Thai music. As a result, the Thai music emotion according to the western music criteria could be understood. The highest f-measure of Thai music from multiple linear regression All model was 41% and the f-measure of western music from multiple linear regression No Tempo model was 51 %, which are very different because All model in western music was low efficiency than other models. Second, we used 155 popular Thai music for explore emotion and use valence-arousal(energy) values from Spotify API to investigate the results. By using multiple linear regression (MLR) and support vector regression (SVR), kinds of "Linear". Experimental results demonstrate that the multiple linear regression provides highest accuracy (61.29%), precision (65%), recall (61%) and f-measure (60%) which is more than support vector regression.

กิตติกรรมประกาศ

ความสำเร็จของชีวิตไม่ได้เกิดขึ้นจากแค่ใบปริญญา แต่ปริญญาใบนี้ได้เปิดโอกาสให้ข้าพเจ้าได้เรียนรู้ประสบการณ์ใหม่ ๆ มากมาย เปิดโลกทัศน์และมุมมอง ที่ข้าพเจ้าไม่เคยคิดว่าจะได้เจอ ไม่ใช่แค่ด้านวิชาการ แต่รวมไปถึงด้านอื่น ๆ ที่สามารถนำไปประยุกต์ใช้ในชีวิตประจำวัน

ตลอดเวลาที่ศึกษา ข้าพเจ้ารับรู้ได้ถึงการเติบโตทางด้านความคิดที่ดีของตัวเองผ่านงานวิจัย ซึ่งต้องใช้ความพยายาม และหลีกหนีไม่ได้ที่จะพบเจอกับความล้มเหลว แต่ด้วยความมุ่งมั่นที่ข้าพเจ้าต้องการไปให้ถึงเป้าหมาย ทำให้ข้าพเจ้าเอาชนะความล้มเหลว และก้าวข้ามขีดจำกัดของตัวเอง แม้จะเป็นแค่ก้าวเล็ก ๆ แต่ข้าพเจ้าภูมิใจในตัวเองเป็นอย่างยิ่ง ในทุกครั้งที่ได้ก้าวข้าม

ขอขอบคุณ รศ.ดร.ชวลิต เบญจางคประเสริฐ อาจารย์ที่ปรึกษา ที่เปิดโอกาสให้ข้าพเจ้าได้ศึกษาวิจัยในหัวข้อที่ข้าพเจ้าสนใจ คอยสนับสนุนและเป็นแรงผลักดันที่ดี ส่งผลให้ข้าพเจ้ามีความมุ่งมั่นที่จะศึกษาในเส้นทางวิชาการระดับสูงต่อไป

และสุดท้ายนี้ ขอขอบคุณ ครอบครัว ที่สนับสนุนให้ข้าพเจ้าศึกษาต่อ แม้ในช่วงแรกข้าพเจ้าจะรู้สึกไม่มั่นใจในศักยภาพตนเองที่ตัดสินใจศึกษาต่อ แต่แรงผลักดันจากครอบครัว คือ สิ่งสำคัญที่ทำให้ข้าพเจ้ามีความมั่นใจ ส่งผลให้ประสบความสำเร็จในก้าวนี้ และจะเป็นสิ่งสำคัญเสมอของก้าวต่อ ๆ ไปในอนาคต

สรวิศ แสงนาค

สารบัญ

หน้า

สารบัญ (ต่อ)

หน้า

สารบัญตาราง

สารบัญตาราง (ต่อ)

สารบัญรูป

สารบัญรูป (ต่อ)

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ตั้งแต่อดีตกาลจนถึงปัจจุบัน บทเพลงเป็นงานศิลปะที่มนุษย์สามารถเข้าถึงได้ง่ายที่สุด มีให้เลือกหลากหลาย และในปัจจุบันที่เทคโนโลยีได้มีส่วนเกี่ยวข้องกับชีวิตมนุษย์เป็นอย่างมาก วงการดนตรีก็ได้นำเทคโนโลยีเข้ามาช่วยอำนวยความสะดวกสบายให้มากยิ่งขึ้นเช่นกัน แม้ดนตรีจะมีพื้นฐานอยู่บนความเป็นศิลป์ แต่ความเป็นเหตุและผลของวิทยาศาสตร์ได้ก่อเกิดทฤษฎีต่าง ๆ เข้ามาช่วยยกระดับวงการดนตรีอย่างปฏิเสธไม่ได้ ไม่ว่าจะเป็นการใช้ในเครื่องดนตรี ระบบเสียง อุปกรณ์ทำเพลง หรือแม้แต่โน้ตดนตรีและเสียงประสาน ก็สามารถนำความรู้ด้านวิทยาศาสตร์มาอธิบายได้ แต่สิ่งหนึ่งในบทเพลงที่ยากจะอธิบายได้ก็คือ อารมณ์เพลง เพราะเพลงเดียวกันเมื่อคนฟังไม่ใช่คนเดียวกัน ก็อาจจะนิยามอารมณ์ไม่เหมือนกันขึ้นอยู่กับปัจจัยต่าง ๆ ในชีวิตส่วนตัวของผู้ฟัง หรือในกรณีคนฟังคนเดิม ได้ฟังเพลงเดิมในช่วงเวลาที่แตกต่างกัน ก็มีโอกาสที่จะนิยามอารมณ์เพลงนั้นแตกต่างออกไป ปัจจัยด้านอารมณ์เป็นปัจจัยหลักที่มนุษย์ใช้เลือกเพลงฟัง รวมไปถึงการเลือกเพลงประกอบสถานการณ์ต่าง ๆ เช่น ดนตรีประกอบฉากภาพยนตร์ เพลงประกอบงานสังสรรค์รื่นเริง เพลงฟังสบาย ๆ ในร้านกาแฟ และอื่น ๆ ก็ล้วนมีพื้นฐานมาจากอารมณ์

ในงานวิจัยนี้มีความสนใจศึกษาเกี่ยวกับอารมณ์เพลง โดยนำความรู้ด้านวิศวกรรมศาสตร์ ได้แก่ การประมวลผลสัญญาณ (Signal processing) การเรียนรู้ด้วยเครื่อง (Machine learning) ประยุกต์เข้ากับความรู้ด้านจิตวิทยา (Psychology) และด้านดนตรี (Musicology) เพื่อทำนายอารมณ์เพลง งานวิจัยด้านนี้กำลังได้รับความนิยมอย่างมากในปัจจุบัน อันมีผลมาจากแอปพลิเคชันฟังเพลงที่เรียกว่า Music streaming ที่ได้กลายเป็นช่องทางหลักในการฟังเพลง เพราะสะดวกสบาย มีเพลงให้เลือกฟังกว่าสิบล้านเพลงจากทั่วโลก และมีค่าใช้จ่ายที่ไม่สูง อีกทั้งยังมีระบบแนะนำเพลงที่ทำให้ผู้ฟังได้เลือกฟังเพลงตรงตามอารมณ์ หรือสถานการณ์ ดังนั้นการวิเคราะห์อารมณ์เพลงจะช่วยให้ผู้ให้บริการเข้าใจอารมณ์เพลงจากผู้ฟังมากขึ้น และนำข้อมูลอารมณ์เพลงที่วิเคราะห์ได้มาสร้างระบบแนะนำเพลงที่เหมาะสมให้ผู้ฟัง ส่งผลให้มูลค่าทางธุรกิจเพิ่มขึ้น แต่ทุกวันนี้งานวิจัยอารมณ์เพลงก็ยังไม่มีระบบที่ดีที่สุด เพราะอารมณ์นั้นยากจะเข้าใจ ดังนั้นการพัฒนาระบบวิเคราะห์อารมณ์เพลงจึงเป็นเรื่องที่ท้าทายต่อนักวิจัยในปัจจุบันเป็นอย่างมาก

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

เพลงยอดนิยมทั่วโลกในปัจจุบันมีพื้นฐานดนตรีอยู่บนแนวเพลงเดียวกัน ไม่ว่าจะเป็น ป๊อบ ร็อก แจ๊ส โฟล์ค และอื่น ๆ ซึ่งต่างพื้นที่ก็มีความนิยมในเพลงที่แตกต่างกันออกไป โดยปัจจัยหลักของความแตกต่างก็ คือ ภาษา ซึ่งเป็นส่วนสำคัญในการสื่อสารอารมณ์และควาหมายของเพลง เพลงที่ใช้ภาษาประจำประเทศจึงมีความนิยมในประเทศนั้น ๆ มากกว่า เพราะผู้ฟังส่วนใหญ่สามารถเข้าใจอารมณ์และความหมายของเพลงได้ แต่เพลงจากฝั่งตะวันตกซึ่งใช้ภาษาอังกฤษเป็นหลัก ก็ได้รับความนิยมมากที่สุด อาจเป็นเพราะว่าภาษาอังกฤษ คือ ภาษาหลักอันดับหนึ่งของโลก จึงมีผู้คนเข้าถึงเป็นจำนวนมาก แต่ก็ยังมีอีกมากที่ชื่นชอบเพลงฝั่งตะวันตกโดยที่ไม่เข้าใจความหมายของภาษาอังกฤษ โดยมีเหตุผลที่ชื่นชอบไปในด้านของเมโลดี้ ดนตรี เสียงประสาน อันเป็นโครงสร้างหลักของเพลง ในงานวิจัยนี้จึงสนใจที่จะสร้างโมเดลอารมณ์เพลงตะวันตก โดยใช้คุณสมบัติเพลงจากการประมวลสัญญาณเสียงที่ไม่เกี่ยวข้องกับภาษา และนำเพลงไทยเข้าไปทดสอบและทำนายอารมณ์จากโมเดลนั้น ผลลัพธ์จะแสดงให้เห็นว่าโมเดลที่มีประสิทธิภาพที่สุดของเพลงตะวันตกและเพลงไทย เป็นโมเดลเดียวกันหรือไม่

1.3 สมมติฐานของงานวิจัย

โมเดลที่มีประสิทธิภาพที่สุดของเพลงตะวันตกและเพลงไทย ควรเป็นโมเดลเดียวกัน เพราะเมื่อถอดภาษาออกจากบทเพลง โครงสร้างดนตรีต่าง ๆ มีความคล้ายคลึงกันเป็นอย่างมาก คุณสมบัติเพลงที่ใช้ในงานนี้ประกอบไปด้วย Pitch, Dynamic, Estimate Tempo, Tonality และ Timber ซึ่งเป็นปัจจัยพื้นฐานของดนตรี คุณสมบัติเพลงที่เป็นตัวแปรสำคัญต่ออารมณ์ คือ จังหวะ ในขณะที่ค่าพลังงานของเพลงน่าจะมีอิทธิพลต่ออารมณ์เพลงน้อยสุด

1.4 ขอบเขตการวิจัย

ผู้วิจัยเลือกเพลงตะวันตก 744 เพลง จากชุดข้อมูล MediaEval2013 [3] แบ่งเป็นชุดทดลอง 619 เพลง และชุดทดสอบ 125 เพลง มาสร้างโมเดลเพลงตะวันตก ในส่วนเพลงไทย เลือกใช้เพลงที่ผู้วิจัยเก็บสะสมจำนวน 125 เพลง โดยทำการซื้ออย่างถูกกฎหมายจาก ITunes และ Audio CD โดยเพลงไทยทั้งหมดเป็นเพลงยอดนิยม ค่าอารมณ์ของเพลงไทยอ้างอิงจาก Spotify API

การประมวลผลสัญญาณเสียงเลือกใช้ Pitch, Dynamic, Estimate Tempo, Tonality และ Timbre และการเรียนรู้ของเครื่องใช้วิธี Multiple linear regression, K-Nearest neighbors และ Support vector regression แบบ Linear ดำเนินการบนโปรแกรม Python ทั้งหมด โดยใช้ Librosa toolbox มาช่วยในการประมวลผลสัญญาณเสียง และ Scikit-learn ในการเรียนรู้ของเครื่อง

1.5 ขั้นตอนการวิจัย

ขั้นตอนการศึกษาในงานวิทยานิพนธ์ฉบับนี้ แบ่งขั้นตอนการศึกษาเป็น 6 บท โดยบทแรกนั้น จะเป็นการกล่าวถึงความเป็นมาและความสำคัญของปัญหา วัตถุประสงค์ สมมติฐานและขอบเขตการวิจัย ในส่วนของบทต่าง ๆ อีก 5 บท สามารถแบ่งเนื้อหาโดยสรุปได้ดังนี้

บทที่ 2 อธิบายชุดข้อมูล MediaEval2013 ซึ่งประกอบไปด้วยเพลงตะวันตกที่นำมาใช้ในงานวิจัยนี้ และแผนภาพ Valence-Arousal ที่ใช้อ้างอิงอารมณ์ในงานวิจัยเกี่ยวกับอารมณ์เพลง รวมไปถึงการใช้ Spotify API

บทที่ 3 อธิบายถึงการประมวลผลสัญญาณเสียงวิธีต่าง ๆ ที่ใช้ในงานวิจัยนี้ เลือกใช้วิธีอะไร กี่แบบ เหตุผลที่เลือกใช้ ซึ่งขั้นตอนนี้เป็นส่วนสำคัญอย่างมากต่อประสิทธิภาพของโมเดล

บทที่ 4 กล่าวถึงการเรียนรู้ด้วยเครื่อง (Machine learning) วิธี Multiple linear regression และ K-Nearest neighbors ซึ่งนำไปใช้ในการทำนายอารมณ์เพลง การเพิ่มประสิทธิภาพด้วยวิธี Backward Elimination และตรวจสอบประสิทธิภาพของโมเดลด้วยวิธีค่าความถูกต้อง (Accuracy), ค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และ F-measure

บทที่ 5 นำเสนอการออกแบบระบบและผลการทดลอง

บทที่ 6 สรุปผลการทดลองทั้งหมด รวมไปถึงข้อเสนอแนะ การนำไปใช้ประโยชน์และแนวทางในการทำวิจัยหัวข้อนี้ในอนาคต

บทที่ 2
Valence-Arousal และชุดข้อมูล

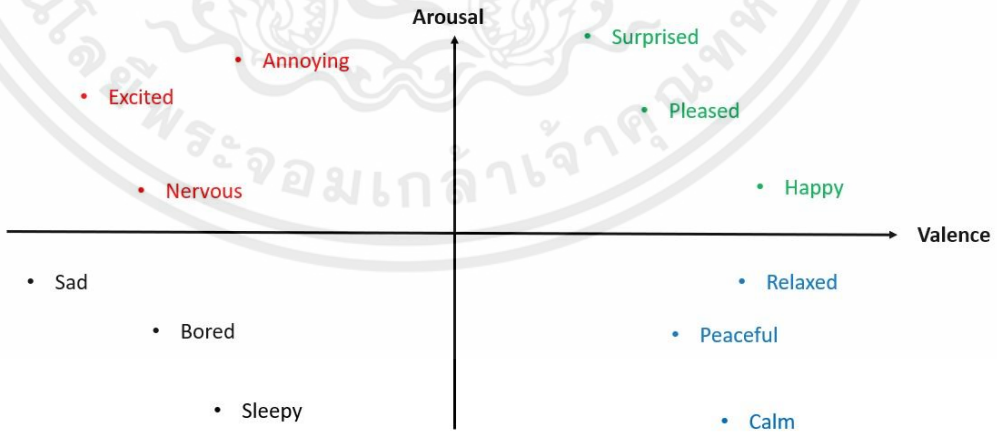
2.1 Valence-Arousal

แผนภาพ Valence-Arousal ในรูปที่ 2.1 เป็นแผนภาพ 2 มิติ ถูกนำมาใช้อ้างอิงในงานวิจัยด้านอารมณ์เพลงเป็นอย่างมาก เพื่อลดความไม่ชัดเจนในคำนิยามของอารมณ์ แผนภาพนี้ถูกคิดค้นโดย Russell [1] และพัฒนาเพิ่มเติมโดย Thayer [2]

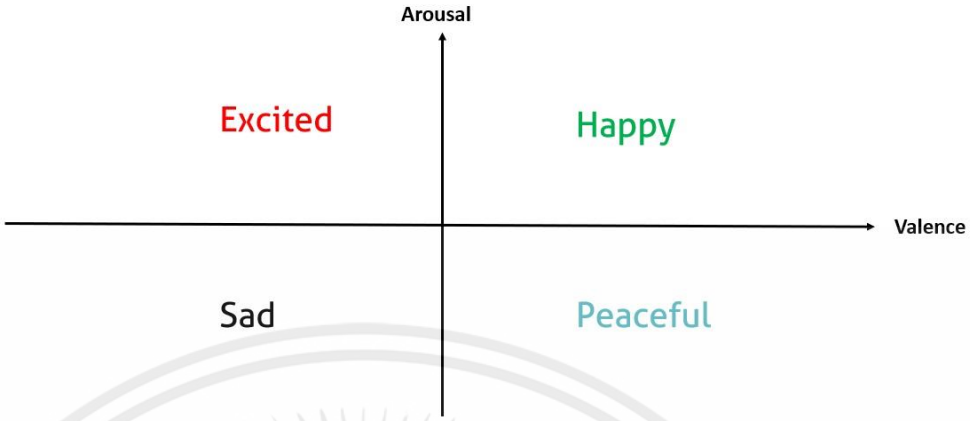
แกน X คือ Valence อธิบายความรู้สึก Positive และ Negative ตัวอย่าง เช่น เพลงที่ให้ค่า Valence มาก แสดงว่าเพลงนั้นแสดงความรู้สึก Positive ในขณะที่ค่า Valence น้อย แสดงว่าเพลงนั้นให้ความรู้สึกที่ Negative

แกน Y คือ Arousal อธิบายความรู้สึก Active และ Passive ตัวอย่าง เช่น เพลงที่ให้ค่า Arousal มาก แสดงว่าเพลงนั้นแสดงความรู้สึก Active คนฟังจะมีความตื่นตัว คึกคักเมื่อได้ฟัง ในขณะที่ค่า Arousal น้อยจะแสดงถึงความรู้สึก Passive ทำให้ผู้ฟังไม่รู้สึกตื่นตัว ซึ่งอาจเป็นเพลงเศร้า เพลงช้า หรือเพลงฟังสบาย ๆ

แผนภาพนี้จะแบ่งเป็น 4 ควอแดรนท์ (quadrant) ในแต่ละควอแดรนท์จะมีหลากหลายอารมณ์ตามพื้นที่ต่าง ๆ บนแผนภาพ เพื่อที่จะทราบอารมณ์จึงต้องทำการหาค่า Valence-Arousal และเพื่อลดความไม่ชัดเจนของอารมณ์ในแต่ละควอแดรนท์ วิทยานิพนธ์ฉบับนี้เลือกสรุปอารมณ์ในแต่ละควอแดรนท์เป็นเพียงอารมณ์เดียว ได้แก่ Happy, Excited, Sad และ Peaceful ตามลำดับ ดังในรูปที่ 2.2



รูปที่ 2.1 แผนภาพ Valence-Arousal



รูปที่ 2.2 คำนิยามอารมณ์ของแผนภาพ Valence-Arousal ที่ใช้ในงานนี้

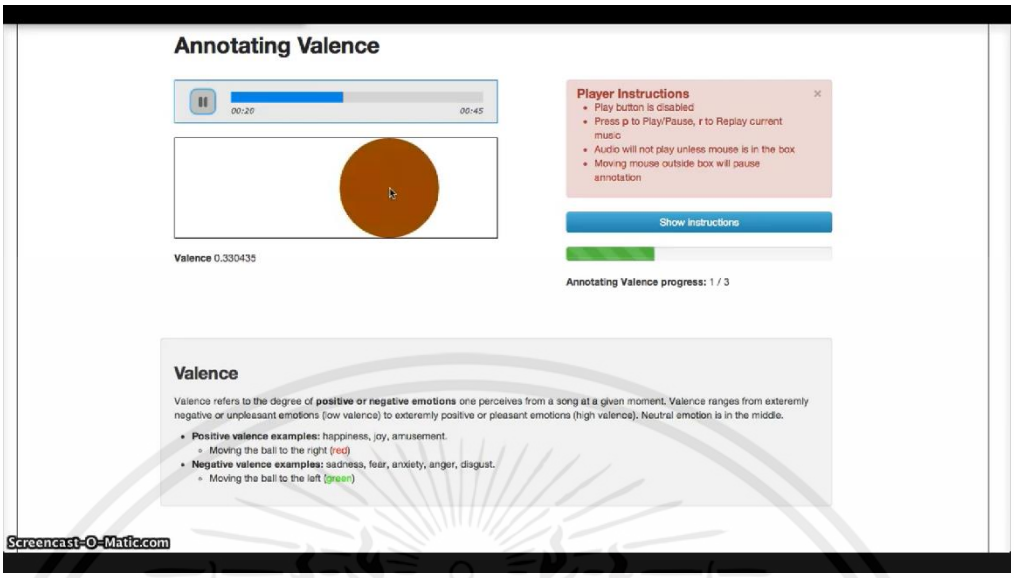
การที่จะทราบอารมณ์แต่ละเพลง โดยอ้างอิงจากแผนภาพ Valence-Arousal จะต้องหาค่าอารมณ์(Value) ไม่ใช่คำนิยาม(Label) ดังนั้น ชุดข้อมูลเพลง(Dataset) จะต้องมีค่าอารมณ์กำกับมาด้วย วิทยานิพนธ์ฉบับนี้จึงเลือกใช้ชุดข้อมูลเพลงตะวันตก MediaEval2013 [3]

2.2 ชุดข้อมูลเพลง (Dataset)

2.2.1 เพลงตะวันตก

MediaEval2013 คือ ชุดข้อมูลเพลงตะวันตก ถูกพัฒนาโดย MIREX ประกอบไปด้วย 744 เพลง จาก freemusicarchive.org แบ่งเป็นชุดทดลอง (Development set) 619 เพลง และชุดทดสอบ (Evaluate set) 125 เพลง มีทั้งหมด 8 แนวเพลง ได้แก่ ป็อบ ร็อก คลาสสิค แจ๊ส โฟล์ค บลูส์ คันทรี่ และอิเล็กทรอนิกส์ แต่ละเพลงมีความยาว 45 วินาที 16 บิต อัตราการสุ่มตัวอย่าง 44.1 กิโลเฮิรตซ์ (KHz) ระบบเสียงโมโน (Mono) และไฟล์ MP3

กระบวนการทดสอบอารมณ์เพลงของชุดข้อมูลนี้ คือ ให้ผู้เข้าร่วมการทดลอง 300 กว่าคน ทำแบบทดสอบออนไลน์ ฟังเพลงความยาว 45 วินาที และเลื่อนตัววัดอารมณ์ Valence-Arousal ไปตามอารมณ์ขณะฟังเพลงนั้น ๆ โดยด้านขวาสุดแสดงถึงค่าที่มาก และด้านซ้ายสุดแสดงถึงค่าที่ต่ำ ในสเกล -1 ถึง 1 การทดสอบจะแยก Valence-Arousal ไม่ได้ทำการทดสอบพร้อมกัน รูปที่ 2.3 แสดงตัวอย่างการใช้งานแบบทดสอบอารมณ์ Valence ขณะฟังเพลงในวินาทีที่ 20 ได้รับค่า Valence เท่ากับ 0.330435

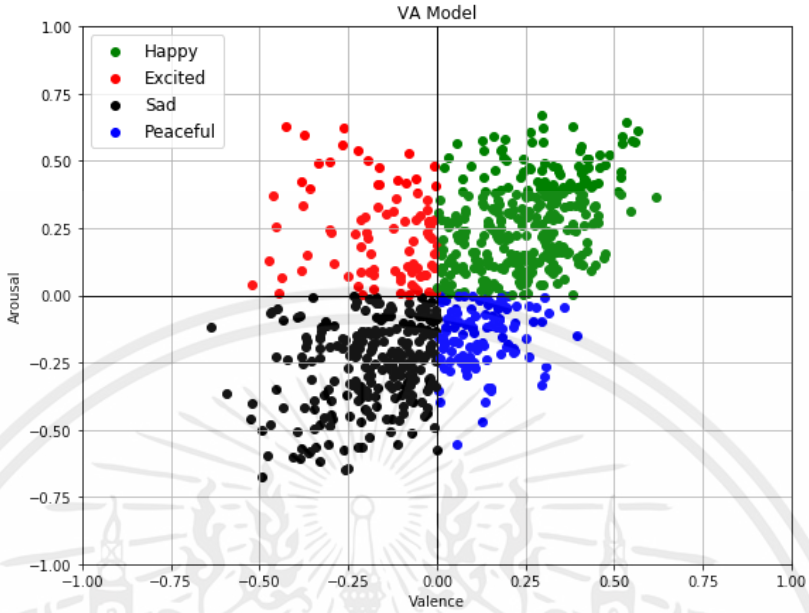


รูปที่ 2.3 ตัวอย่างการใช้งานแบบทดสอบอารมณ์ Valence ของ MediaEval2013 (https://bit.ly/2EkMrbf)

ผู้สร้างชุดข้อมูลเลือกที่จะตัดค่าอารมณ์ที่ผู้เข้าร่วมการทดลองเลือกในช่วง 15 วินาทีแรกออก เพราะอารมณ์ขณะเริ่มต้นไม่มีความเสถียรมากพอ ผลลัพธ์ที่ได้ คือ ค่าเฉลี่ยอารมณ์ของแต่ละเพลงในทุก 0.5 วินาที วิทยานิพนธ์ฉบับนี้ได้นำค่าอารมณ์ทุกช่วงเวลาของแต่ละเพลงมาหาค่าเฉลี่ยอารมณ์ และนำไปพล็อตบนแผนภาพ Valence-Arousal เพื่อหาคำนิยามอารมณ์ของแต่ละเพลง แสดงในตารางที่ 1.1 และรูปที่ 2.4

ตารางที่ 2.1 อารมณ์เพลงของชุดข้อมูล MediaEval2013

	Development	Evaluate	Total
Blues	82	18	100
Classic	98	18	116
Country	90	15	105
Electronic	76	16	92
Folk	62	9	71
Jazz	84	21	105
Pop	65	13	78
ROCk	62	15	77



รูปที่ 2.4 อารมณ์เพลงของ MediaEval2013 บนแผนภาพ Valence-Arousal

งานวิจัยด้านรู้จำอารมณ์เพลงนิยมศึกษาจากเพลงตะวันตก เพลงภาษาอื่น ๆ นิยมนำมาวิจัยไม่มาก เช่น วิเคราะห์อารมณ์เพลงศรีลังกา [4] และ วิเคราะห์เพลงอินเดีย [5] ในขณะที่เพลงไทยมีงานวิจัยที่น้อยมาก ดังนั้น วิทยานิพนธ์ฉบับนี้ได้เลือกศึกษาเพลงไทยซึ่งเป็นภาษาประจำชาติ และเป็นเพลงที่ได้รับความนิยมสูงสุดในพื้นที่การวิจัย

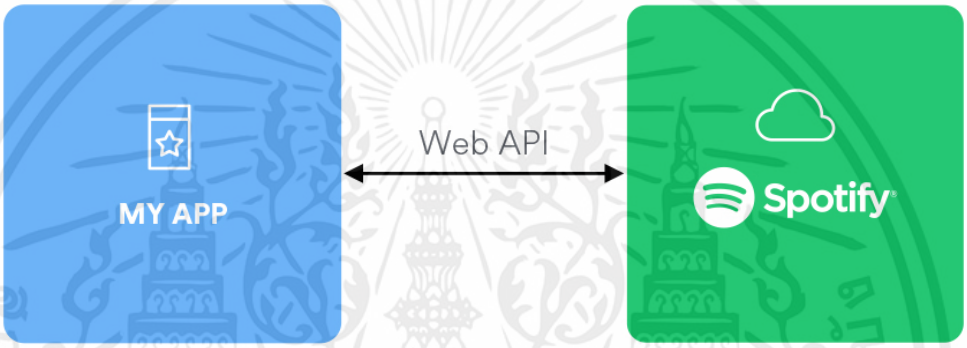
2.2.2 เพลงไทย

งานวิจัยนี้เลือกใช้เพลงจากชุดสะสมส่วนตัว ประกอบไปด้วยเพลงไทยยอดนิยม โดยทุกเพลงมาจากการสนับสนุนอย่างถูกลิขสิทธิ์จาก ITunes และ CD ไฟล์เพลงไทยถูกนำมาเปลี่ยนคุณภาพไฟล์เป็นมาตรฐานเดียวกับเพลงตะวันตก คือ ไฟล์ MP3 ความยาว 45 วินาที 16 บิต อัตราการสุ่มตัวอย่าง 44.1 กิโลเฮิรตซ์ (KHz) และระบบเสียงโมโน (Mono) การตัดต่อเพลงให้เป็น 45 วินาที เป็นการสุ่มท่อน ผู้วิจัยไม่ได้เปิดฟังระหว่างตัดต่อเพื่อความเป็นกลาง เพราะในเพลงเดียวกัน แต่ท่อนไม่เหมือนกันอาจให้อารมณ์ที่ต่างกัน

เพลงไทยไม่มีข้อมูลด้านอารมณ์อ้างอิง แต่จำเป็นต้องหาค่าอารมณ์เพื่อตรวจสอบความถูกต้อง ดังนั้น ผู้วิจัยเลือกที่จะดึงข้อมูลจาก Spotify API [6] มาระบุอารมณ์ โดยข้อมูลจาก Spotify API แสดงเป็นค่า Valence-Arousal เช่นกัน ไม่ได้เป็นคำนิยามอารมณ์

2.3 Spotify API

Spotify คือ ผู้ให้บริการ Music streaming อันดับหนึ่งของโลก ได้รับการยอมรับและยกย่องเป็นอย่างมากในระบบเพลย์ลิสท์(Playlist) แนะนำเพลงให้ผู้ฟัง ซึ่งเป็นจุดขายหลักของ Spotify เนื่องจากมีการวิเคราะห์ด้านเพลงด้วยระบบคอมพิวเตอร์อย่างจริงจัง และเปิดโอกาสให้นักวิจัยดึงข้อมูลต่าง ๆ ไปใช้ในงานวิจัยแบบไม่เสียค่าใช้จ่าย โดยต้องทำการลงทะเบียนเพื่อรับรหัสเข้าใช้งาน และ Spotify จะเก็บข้อมูลการใช้งานเมื่อเข้าไปใช้ ดังนั้นในงานวิจัยนี้จึงเลือกข้อมูลจาก Spotify API มาใช้ในอารมณ์เพลงไทย เพราะมีความน่าเชื่อถือสูง ระบบการทำงาน API เบื้องต้น แสดงในรูปที่ 2.5



รูปที่ 2.5 ระบบการทำงานเบื้องต้นของ Spotify API (https://spoti.fi/2RKRHru)

ผู้วิจัยดึงข้อมูลคุณสมบัติเพลง (Audio Feature) จาก Spotify API ข้อมูลที่ได้จะประกอบไปด้วยคุณสมบัติของเพลงมากมาย ดังรูปที่ 2.6 แต่ในงานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาด้านอารมณ์ จึงเลือกค่า Valence และค่า Energy ถูกนำมาใช้แทน Arousal

ค่า Valence-Arousal(Energy) จาก Spotify API มีสเกล 0 ถึง 1 ซึ่งแตกต่างจาก MediaEval2013 ดังนั้นผู้วิจัยจึงทดลองนำข้อมูลมานอร์มัลไลซ์ (Normalization) เพื่อให้เป็นมาตรฐานเดียวกัน ผลที่ออกมาพบว่าอารมณ์ของเพลงมีการเปลี่ยนแปลงไปอย่างมากจากเดิม ดังนั้นผู้วิจัยจึงไม่ทำการนอร์มัลไลซ์ แต่นำค่าอารมณ์จาก Spotify API ไปพล็อตบนแผนภาพ Valence-Arousal เพื่อให้ได้รับคำนิยามอารมณ์ของแต่ละเพลง และนำคำนิยามที่ได้นั้นมาใช้ตรวจสอบความถูกต้องเพียงอย่างเดียว ผลลัพธ์อารมณ์เพลงไทย แสดงในรูปที่ 2.7

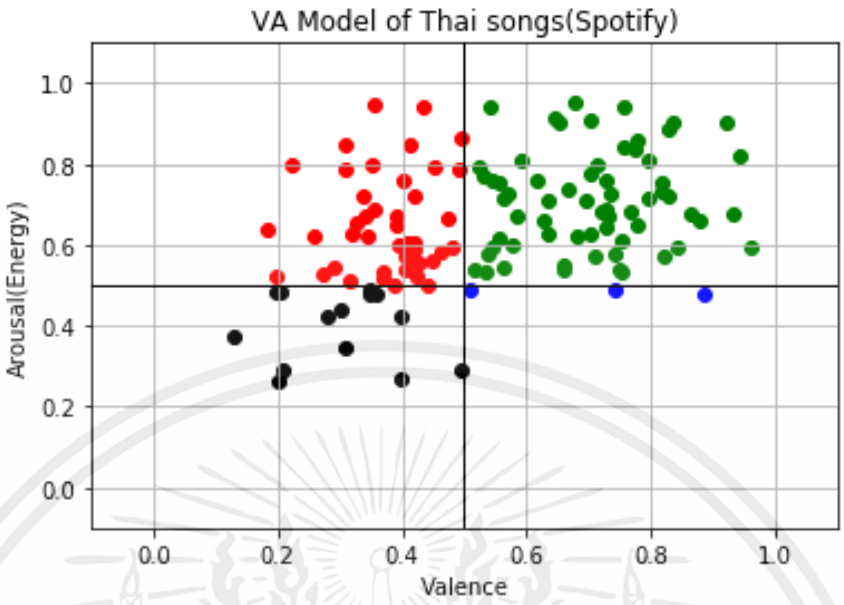
```
{u'acousticness': 0.134,
 u'analysis_url': u'https://api.spotify.com/v1/audio-analysis/7EvGAdKxTEysZMzjuqq8mW',
 u'danceability': 0.724,
 u'duration_ms': 284000,
 u'energy': 0.901,
 u'id': u'7EvGAdKxTEysZMzjuqq8mW',
 u'instrumentalness': 0,
 u'key': 2,
 u'liveness': 0.328,
 u'loudness': -5.572,
 u'mode': 1,
 u'speechiness': 0.0361,
 u'tempo': 122.036,
 u'time_signature': 4,
 u'track_href': u'https://api.spotify.com/v1/tracks/7EvGAdKxTEysZMzjuqq8mW',
 u'type': u'audio_features',
 u'uri': u'spotify:track:7EvGAdKxTEysZMzjuqq8mW',
 u'valence': 0.923}
```


รูปที่ 2.6 Audio feature ที่ได้รับจาก Spotify API

ตารางที่ 2.2 และรูปที่ 2.7 แสดงให้เห็นว่าอารมณ์เพลงไทยยอดนิยม มีความตื่นตัวสูงเป็นอย่างมาก คิดเป็น 85% ของเพลงไทยในงานวิจัยนี้ ในขณะที่เพลงที่มีความตื่นตัวต่ำมีน้อยมาก โดยเฉพาะ Peaceful มีเพียง 3 เพลงเท่านั้น ซึ่งเมื่อลองวิเคราะห์กับความเป็นจริง เพลงที่มีความตื่นตัวสูง ค่อนข้างเป็นเพลงยอดนิยมในไทยจริง ๆ เพราะสามารถนำเพลงเหล่านี้ไปเปิดตามงานเทศกาลต่าง ๆ ทำให้มีโอกาสได้ยินบ่อยมากขึ้น ส่งผลต่อทางจิตวิทยา เพิ่มโอกาสให้คนมาชอบเพลงมากขึ้น

ตารางที่ 2.2 จำนวนอารมณ์เพลงไทย 125 เพลง

	Happy	Excited	Sad	Peaceful
Emotions of Spotify API	64	43	15	3



รูปที่ 2.7 อารมณ์เพลงไทยบนแผนภาพ Valence-Arousal

2.4 Python and Librosa toolbox

Librosa [7] คือ Toolbox ที่ปฏิบัติการบน Python ใช้ในด้านประมวลผลสัญญาณเสียง (Audio signal processing) มีหลากหลายฟังก์ชันให้เลือกใช้ตามความเหมาะสม และภาษา Python เป็นภาษาที่นิยมใช้ในการเรียนรู้ด้วยเครื่อง (Machine learning) จึงทำให้มีความสะดวกในการจัดการข้อมูล

การประมวลผลสัญญาณเสียง จะกล่าวถึงในบทที่ 3 เกี่ยวกับวิธีประมวลผลสัญญาณเสียง แบบต่าง ๆ ที่ถูกนำมาใช้ในวิทยานิพนธ์ฉบับนี้ ในส่วนของการเรียนรู้ด้วยเครื่อง จะกล่าวถึงในบทที่ 4 เกี่ยวกับวิธีการเรียนรู้ด้วยเครื่อง ที่ถูกนำมาใช้ในวิทยานิพนธ์ฉบับนี้เช่นกัน

บทที่ 3
การประมวลผลสัญญาณเสียง

ในบทนี้จะกล่าวถึงเสียง และการประมวลผลสัญญาณเสียงที่ใช้ในวิทยานิพนธ์นี้ โมเดลจะมีประสิทธิภาพก็ต่อเมื่อข้อมูลที่นำมาสร้างโมเดลมีความถูกต้องและมีประสิทธิภาพ ดังนั้นขั้นตอนนี้จึงเป็นขั้นตอนที่สำคัญมาก หากนำข้อมูลที่ไม่มีประสิทธิภาพมาใช้งาน จะส่งผลให้ประสิทธิภาพโมเดลต่ำและยากที่จะพัฒนาให้ดีขึ้นได้ ดังนั้นหากจะสร้างโมเดล ผู้วิจัยจะต้องพิจารณาความเหมาะสมในการใช้ข้อมูล

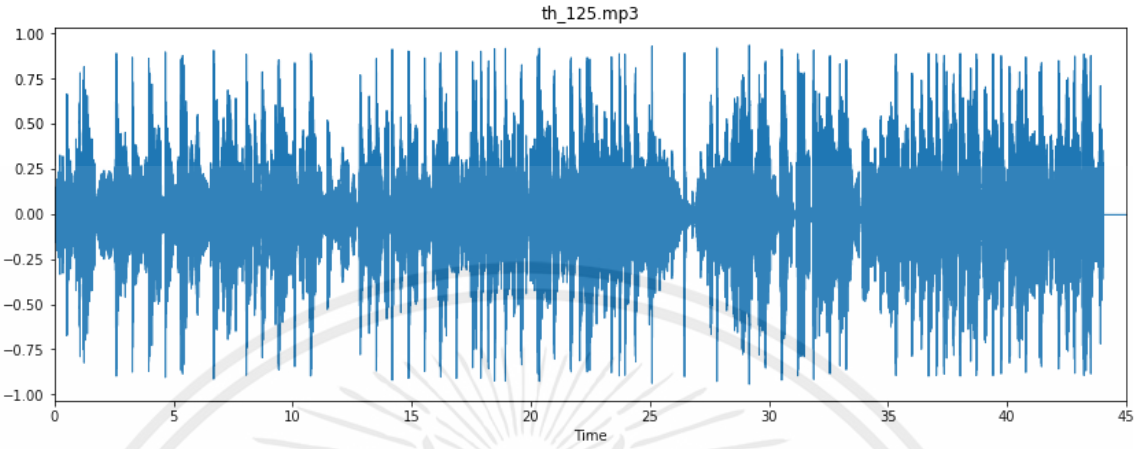
3.1 สัญญาณเสียง

เสียงเกิดจากการสั่นสะเทือนของวัตถุ โดยเดินทางผ่านอากาศ ในชีวิตประจำวันมนุษย์จะได้ยินเสียงต่าง ๆ อยู่ตลอดเวลา เช่น เสียงพูด เสียงเดิน เสียงเครื่องยนต์ และอีกมากมาย โดยเสียงที่มนุษย์สามารถได้ยินจะอยู่ในช่วงความถี่ 20 Hz ถึง 20,000 Hz เรียกว่า Audio Frequency หากนอกเหนือจากความถี่ช่วงนี้มนุษย์จะไม่สามารถได้ยิน แต่ใช่ว่าจะไม่มีเสียง ในปัจจุบันมีงานวิจัยที่นำเสียงมาประยุกต์มากมาย เช่น การสั่งการอุปกรณ์ด้วยเสียง การประมวลผลเสียงพูด การพิสูจน์หลักฐานทางเสียง การลดเสียงรบกวน และอื่น ๆ

สัญญาณเสียงประกอบไปด้วยสัญญาณไซนูซอยด์ (Sinusoid) ที่มีความถี่ต่างกันจำนวนมาก ทำให้เสียงมีความแตกต่างกัน ในรูปที่ 3.1 แสดงสัญญาณเสียงในโดเมนเวลา หากต้องการที่จะนำเสียงไปวิเคราะห์ข้อมูลควรแปลงสัญญาณจากโดเมนเวลาเป็นโดเมนความถี่ ด้วยการประมวลผลสัญญาณดิจิทัล (Digital Signal processing) โดยปกติแล้วสัญญาณเสียงจะอยู่ในรูปแบบสัญญาณแอนะล็อก (Analog) ดังนั้นต้องทำการแปลงสัญญาณแอนะล็อกเป็นสัญญาณดิจิทัลด้วยวิธีการสุ่มตัวอย่าง (Sampling) อัตราการสุ่มตัวอย่างที่นิยมใช้อยู่ที่ 22.05 และ 44.1 กิโลเฮิรตซ์

ในด้านเสียงดนตรีนั้น เกิดจากการสั่นสะเทือนของเครื่องดนตรี เสียงร้องหรือเสียงพูดซึ่งเกิดจากการสั่นสะเทือนของกล่องเสียงในลำคอมนุษย์ เมื่อนำเสียงดนตรี เสียงร้อง และเสียงอื่น ๆ มารวมกัน จึงทำให้เกิดเสียงเพลงที่ได้ยินในชีวิตประจำวัน

การวิเคราะห์สัญญาณเพลง จะต้องนำมาคำนวณด้วยสมการแบบต่าง ๆ เพื่อประมวลผลสัญญาณให้อยู่ในรูปแบบตามที่ต้องการ การประมวลผลสัญญาณเสียงมีให้เลือกใช้มากมาย ดังนั้น ผู้วิจัยต้องพิจารณาเรื่องความเหมาะสมว่างานวิจัยควรใช้คุณสมบัติสัญญาณแบบไหน



รูปที่ 3.1 สัญญาณเสียงเพลง th_125.mp3 รูปแบบโดเมนเวลา

3.2 การประมวลผลสัญญาณเสียง (Audio Signal processing)

การประมวลผลสัญญาณเสียง หมายถึง การประมวลผลสัญญาณเสียงทุกชนิดที่เกิดขึ้นบนโลกใบนี้ ไม่จำกัดว่าจะต้องเป็นเสียงดนตรี หรือเสียงที่เกิดจากมนุษย์เท่านั้น แต่รวมไปถึงเสียงที่เกิดจากสัตว์หลากหลายชนิดซึ่งมนุษย์ไม่สามารถที่จะได้ยินคลื่นเสียงเหล่านี้ได้ รวมไปถึงเสียงอีกมากมายที่ต้องใช้ความรู้เฉพาะทางในการอธิบายคุณลักษณะสัญญาณ ในปัจจุบันการประมวลผลสัญญาณเสียงมีให้เลือกใช้มากมายหลายแบบ ตามโจทย์และความเหมาะสมของงานวิจัย

งานวิจัยด้านรู้จำอารมณ์เพลง มีการเลือกใช้การประมวลผลสัญญาณเสียง และคุณสมบัติเสียงที่หลากหลายในปัจจุบัน เช่น [8] นำเสนอคุณสมบัติเสียงเพลง 5 แบบ ได้แก่ Pitch, Tempo, Tonality, Dynamic และ Timbre หรืองานวิจัย [9] ได้ใช้ Root mean square (RMS) energy, Mel-frequency cepstral coefficients (MFCCs), Zero-crossing rate (ZCR), F0 และ Voice probability (Voiceprob) งานวิจัยที่ [10]-[12] ได้ประยุกต์ใช้เนื้อเพลงในการวิจัย รวมไปถึงการนำเซนเซอร์กายภาพมาตรวจจับการตอบสนองของร่างกายขณะฟังเพลง [13]

วิทยานิพนธ์ฉบับนี้ หัวข้อวิจัยเกี่ยวกับเสียงเพลง ซึ่งมีต้นกำเนิดมาจากทฤษฎีดนตรีต่าง ๆ รวมกันสร้างสรรค์เป็นบทเพลงมากมายให้ผู้คนรับฟัง ดังนั้น การทำวิจัยเกี่ยวกับเสียงเพลง ผู้วิจัยจะต้องมีความรู้พื้นฐานด้านทฤษฎีดนตรี เพื่อที่จะนำความรู้มาอธิบายคุณลักษณะสัญญาณของเสียงเพลง และนำผลลัพธ์ที่ได้ไปประยุกต์ใช้ให้เกิดประโยชน์สูงสุด ผู้วิจัยเลือกสัญญาณเสียง 5 แบบ ได้แก่ Pitch, Dynamic, Tempo, Tonality และ Timbre ซึ่งใช้การประมวลผลสัญญาณเสียงแบบ Discrete Fourier Transform,

Root Mean Square Energy, Estimate Tempo, Chromagram และ Mel-frequency cepstral coefficients ตามลำดับ แสดงในตารางที่ 3.1

ตารางที่ 3.1 คุณสมบัติเพลง 5 แบบ

Features	No. of features	Methods
Pitch	1	Discrete Fourier transform
Dynamic	1	Root mean square energy
Tempo	1	Estimate tempo
Tonality	12	Chromagram
Timbre	13	Mel-frequency cepstral coefficients (MFCCs)

3.2.1 Discrete Fourier Transform

การแปลงฟูเรียร์ (Fourier Transform) ได้ถูกนำมาใช้เพื่อวิเคราะห์สัญญาณ เพื่อทราบว่าสัญญาณนั้น ๆ มีความถี่เท่าไหร่ประกอบอยู่ในสัญญาณบ้าง ซึ่งเป็นประโยชน์อย่างมากต่อหลายแขนงวิชา จึงได้มีการพัฒนาขั้นตอนวิธีเพื่อที่จะวิเคราะห์สัญญาณด้วยการแปลงฟูเรียร์ เพื่อนำมาประยุกต์ใช้งานได้ง่าย และเพิ่มความเร็วในการประมวลผลให้มากยิ่งขึ้น โดยมีการพัฒนาการแปลงฟูเรียร์ด้วยวิธี Discrete Fourier Transform (DFT) ดังสมการที่ (3.1) ซึ่งเป็นการคำนวณที่มีสัญญาณอินพุตและสัญญาณเอาต์พุตเป็นจำนวนที่มีลำดับจำกัด จึงสะดวกต่อการนำไปประยุกต์ ให้คอมพิวเตอร์หรือตัวประมวลผลมาช่วยคำนวณการแปลงฟูเรียร์ ซึ่งก็จะสะดวกต่อการนำไปใช้งานมากยิ่งขึ้น

$$\mathrm{F}[x(t)] = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t} dt \qquad (3.1)$$

เมื่อ x(t) คือ Discrete signal

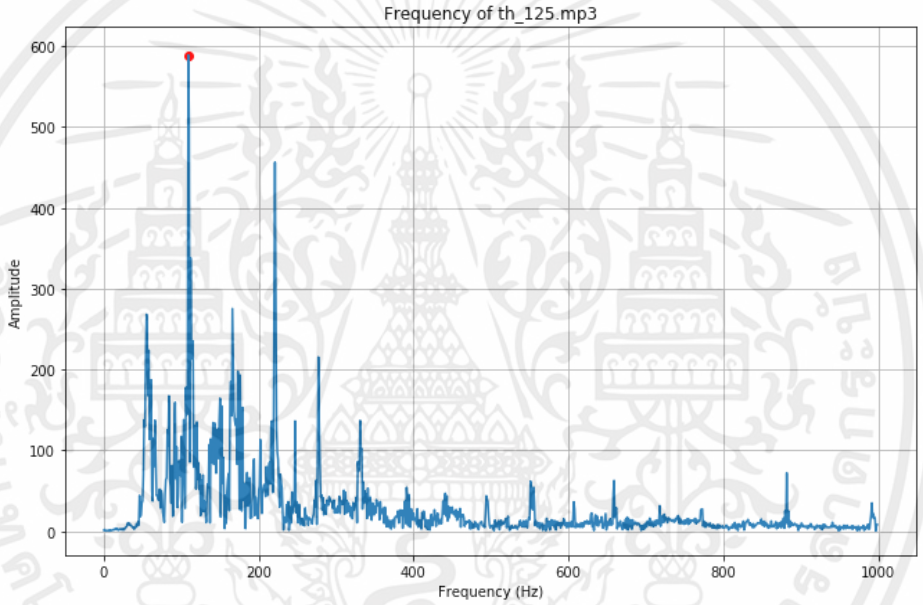
t คือ เวลา

แต่ DFT ก็ยังมีจำนวนครั้งของการคูณกันของจำนวนเชิงซ้อนอยู่มาก จึงต้องใช้เวลานานในการประมวลผล ดังนั้นจึงได้มีผู้พัฒนาต่อโดยลดจำนวนการคูณกันของจำนวนเชิงซ้อนนี้ลงอีก โดยการจัดรูปสมการของ DFT ใหม่ ซึ่งก็คือ การแปลงฟูเรียร์ด้วยวิธีฟาสฟูเรียร์ทรานสฟอร์ม (Fast Fourier Transform

: FFT) ซึ่งยังสามารถใช้คอมพิวเตอร์ หรือตัวประมวลผล ทำการคำนวณได้ อีกทั้งมีความเร็วเพิ่มขึ้น ดังนั้น จึงมีผู้สนใจนำไปพัฒนาขั้นตอนวิธีอีกหลากหลาย โดยลดจำนวนครั้งของการคูณกันของจำนวนเชิงซ้อนลง เพื่อเพิ่มความเร็วในการประมวลผลให้เพิ่มขึ้นอีก

ผู้วิจัยเคยทำการวิจัยเบื้องต้นเกี่ยวกับโน้ตดนตรี และได้ผลลัพธ์ว่าความถี่มูลฐาน(Fundamental frequency) คือ ตัวแปรสำคัญในการบ่งบอกโน้ตดนตรี ดังนั้นจึงได้นำความรู้นี้มาใช้ในงานวิจัยนี้

รูปที่ 3.2 แสดงตัวอย่างสัญญาณเพลง th_125.mp3 ในรูปแบบโดเมนความถี่ ซึ่งถูกเปลี่ยนรูปแบบจากสัญญาณโดเมนเวลาในรูปที่ 3.1 ด้วยวิธี FFT โดยตรวจจับความถี่ที่มีแอมพลิจูด (Amplitude) สูงสุด ซึ่งก็คือ ความถี่มูลฐานของเพลง และนำมาใช้ในงานนี้



รูปที่ 3.2 สัญญาณเสียงเพลง th_125.mp3 รูปแบบโดเมนความถี่

3.2.2 Root Mean Square Energy (RMS)

คำนวณหาค่าพลังงานโดยรวมของสัญญาณในโดเมนเวลา ดังสมการที่ (3.2)

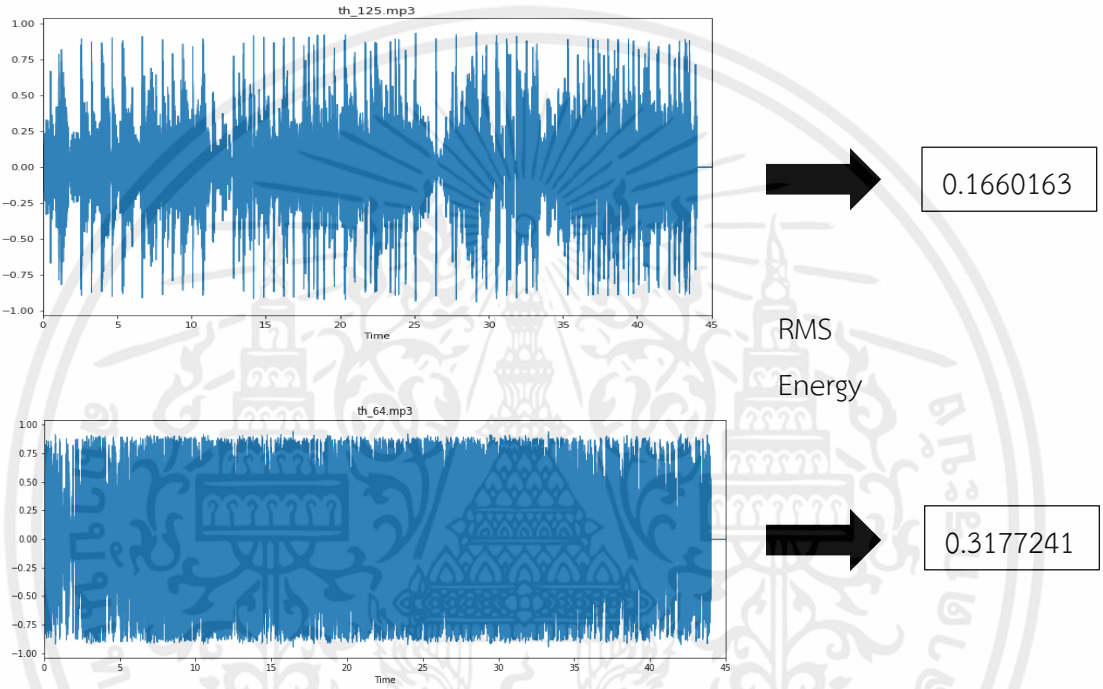
$$X_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|x_i|^2} \qquad (3.2)$$

เมื่อ x_i คือ สัญญาณเสียง

N คือ จำนวนข้อมูลตัวอย่าง (Number of Samples) สามารถหาได้จากสมการที่ (3.3)

$$\frac{\text{Number of samples}}{\text{Time (second)}} = \text{Sampling Rate(Hz)} \quad (3.3)$$

เพลงที่ใช้ในงานวิจัยนี้ ผู้วิจัยได้ลดทอนอัตราการการสุ่มตัวอย่าง (Sampling rate) ของทุกเพลงเป็น 22,050 Hz และมีความยาวของเพลง 45 วินาที เมื่อนำมาคำนวณหาจำนวนตัวอย่าง (Number of samples) ที่จะนำไปใช้ในสมการที่ 3.2 ได้ผลลัพธ์เท่ากับ 992,250 ข้อมูลตัวอย่าง



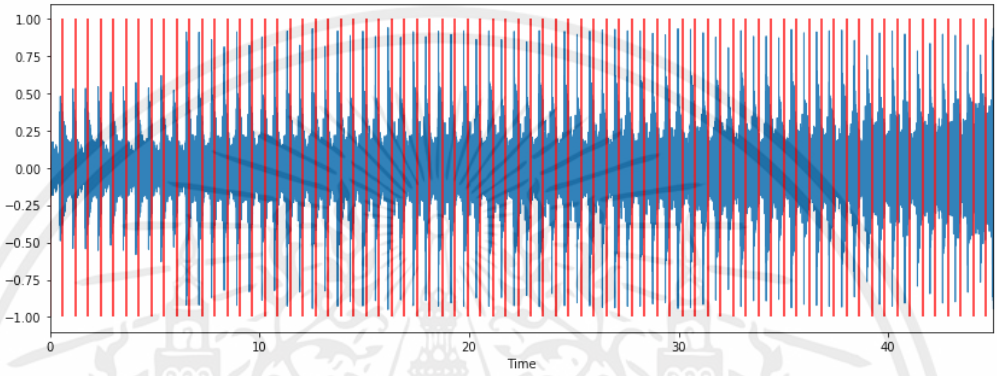
รูปที่ 3.3 ค่า RMS ของเพลง th_125.mp3 และ th_64.mp3

รูปที่ 3.3 แสดงสัญญาณเพลงในโดเมนเวลา 2 เพลงไทย ได้แก่ th_125.mp3 และ th_64.mp3 เมื่อดูแอมพลิจูดสูงสุด ต่ำสุด ของทั้งสองเพลง พบว่ามีค่าใกล้เคียงกัน แต่รูปแบบสัญญาณโดยรวม th_64.mp3 มีความหนาแน่นของสัญญาณมากกว่า th_125.mp3 อย่างเห็นได้ชัด เมื่อนำสัญญาณเพลงทั้งสองไปคำนวณ RMS ผลลัพธ์แสดงว่า th_64.mp3 มีค่าพลังงานมากกว่าเกือบหนึ่งเท่าตัว

3.2.3 Estimate Tempo

จังหวะ คือ คุณสมบัติเพลงสำคัญที่มีผลต่ออารมณ์ โดยเฉพาะในด้านความตื่นตัว เพลงที่ทำให้รู้สึกสนุกชวนเต้นมักจะมีจังหวะเร็ว ในขณะที่เพลงช้ามักจะมีจังหวะที่ช้า ทฤษฎีดนตรีด้านจังหวะโดยทั่วไปมีหลายรูปแบบ แต่ที่พบในเพลงเป็นส่วนมากในเพลงยอดนิยม คือ จังหวะ 4/4 ซึ่งสามารถใช้ได้

ทุกแนวเพลง รองลงมาคือจังหวะ 6/8 จะถูกใช้ในเพลงช้า นอกจากนี้ยังมีจังหวะอื่น ๆ อีกมากมาย ซึ่งต้องใช้ทักษะดนตรีขั้นสูงในการประพันธ์เพลง และให้ความรู้สึกที่ฟังยากจึงไม่เหมาะจะนำมาใช้ในเพลงยอดนิยมเท่าจังหวะ 4/4 และ 6/8 การหาค่าเฉลี่ยของจังหวะ จะทำการตรวจจับแอมพลิจูดของสัญญาณในคาบเวลา แสดงในรูปที่ 3.4 เมื่อสัญญาณเพลง (สีฟ้า) มีแอมพลิจูดที่เด่นชัด ระบบจะทำการตรวจจับ (สีแดง) แอมพลิจูดเหล่านั้น และนำไปคำนวณเพื่อแสดงผลลัพธ์ออกมาในหน่วย beats per minute (bpm)



รูปที่ 3.4 การตรวจจับแอมพลิจูดเพื่อหาค่าเฉลี่ยจังหวะเพลง

3.2.4 Chromagram

ผลลัพธ์ของ Chromagram แสดงให้เห็นโน้ตดนตรี (Musical note score) ที่เกิดขึ้นในแต่ละช่วงเวลานั้น ๆ ของเพลง โดยปกติแล้วในทุกช่วงเวลาของเพลง จะไม่ได้มีโน้ตดนตรีเกิดขึ้นเพียงแค่ตัวเดียว แต่จะมีเกิดขึ้นมากมาย เมื่อนำโน้ตทั้งหมดที่เกิดขึ้นในช่วงเวลาเดียวกัน มารวมกันตั้งแต่ 3 ตัวขึ้นไป จะทำให้เกิดสิ่งที่เรียกว่า คอร์ดเพลง (Chord) และเมื่อนำโน้ตดนตรีที่เกิดขึ้นทั้งหมดในเพลงมาวิเคราะห์ทฤษฎีดนตรีจะทำให้ทราบ คีย์เพลง (Musical key scale)

คีย์เพลง คือ ผลลัพธ์ที่งานวิจัยนี้ต้องการเพื่อนำมาวิเคราะห์อารมณ์เพลง โดยตามหลักทฤษฎีดนตรีมีทั้งหมด 12 คีย์ ได้แก่ C, C#, D, D#, E, F, F#, G, G#, A, A# และ B โดยแต่ละคีย์สามารถแบ่งเป็น 2 ประเภท ได้แก่ เมเจอร์ (Major) และ ไมเนอร์ (Minor) จากงานวิจัยทางสายดนตรีโดยเฉพาะ ผลลัพธ์แสดงให้เห็นว่าเมเจอร์ จะให้ความรู้สึกที่สนุกสนาน ในขณะที่ไมเนอร์จะให้ความรู้สึกที่เศร้าหม่นหมอง แต่ก็ไม่ได้ถูกต้องเสมอไป นักดนตรีสามารถนำเมเจอร์และไมเนอร์ มาแต่งเพลงได้ทุกแบบแล้วแต่ความเหมาะสม โดยโน้ตที่บ่งบอกว่าเพลงนี้ คือ คีย์/คอร์ด เมเจอร์ หรือไมเนอร์ จะดูได้จากตัวที่ 3 (Minor third) ของคีย์นั้น ๆ โดยปกติเมเจอร์และไมเนอร์ มีความสัมพันธ์ที่ตายตัวอยู่หนึ่งอย่าง คือ ตัวที่ 6 ของคีย์เมเจอร์จะเป็นคีย์ไมเนอร์ที่มีโน้ตดนตรีใน สเกล (Scale) เหมือนกันทั้งหมด (Relative)

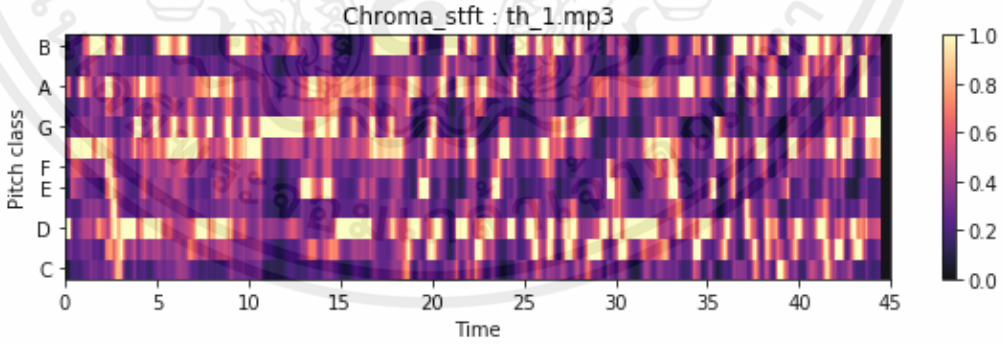
ตารางที่ 3.2 แสดงให้เห็นโน้ตทั้งหมดในคีย์ D major ได้แก่ D, Em, F#m, G, A, Bm และ C#m ซึ่งมีตัวที่ 6 เป็น Bm (B minor) ดังนั้นโน้ตทั้งหมดในคีย์ B minor คือ Bm, C#m, D, Em, F#m, G และ A ซึ่งเหมือนกับคีย์ D major ทุกโน้ต และเป็นอย่างนี้กับทุกคีย์ major

ตารางที่ 3.2 ความสัมพันธ์ของ D major และ B minor

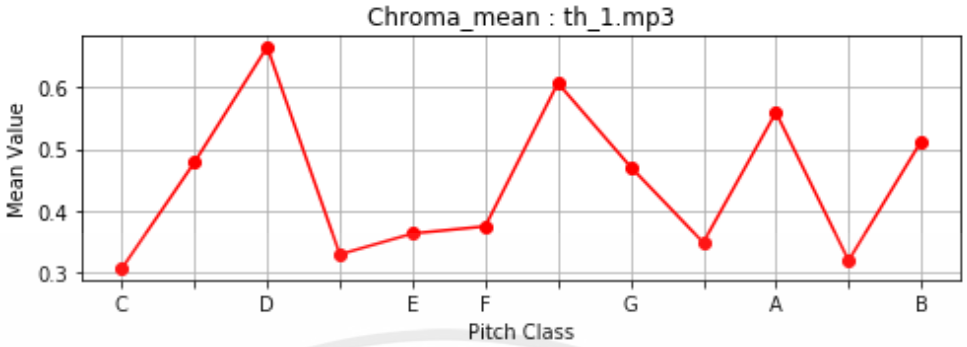
	ตัวที่ 1	ตัวที่ 2	ตัวที่ 3	ตัวที่ 4	ตัวที่ 5	ตัวที่ 6	ตัวที่ 7
D major	D	Em	F#m	G	A	Bm	C#m
B minor	Bm	C#m	D	Em	F#m	G	A

รูปที่ 3.5 แสดงตัวอย่างผลลัพธ์ของเพลงที่ใช้ Chromagram โดยแกน X คือ คีย์เพลง และ แกน Y คือ เวลา พบว่ามีค่าพลังงานเกิดขึ้นมากมายและแตกต่างกัน ยากที่จะสรุปคีย์เพลงได้ ดังนั้น ผู้วิจัยได้ใช้การหาค่าเฉลี่ย (Average) ของแต่ละโน้ต เพื่อดูผลลัพธ์ของโน้ตที่มีแอมพลิจูดที่เด่นชัด ดังรูปที่ 3.6 และนำมาคาดการณ์ว่าเพลงนี้คือคีย์อะไร และเป็น Major หรือ Minor

ผลลัพธ์ที่ได้แสดงออกมาว่าโน้ต D มีแอมพลิจูดที่สูงที่สุด จึงมีโอกาสเป็นคีย์ D และเมื่อดูโน้ตที่มีแอมพลิจูดสูงสุดสองอันดับรองลงมาพบว่า คือ โน้ต A และ F# ซึ่งการจะรู้ได้ว่าเป็นเมเจอร์หรือไมเนอร์ต้องดูตัวที่ 3 ของคีย์นั้น ๆ ซึ่งคีย์ D major มีตัวที่ 3 คือ F# ในขณะที่คีย์ D minor มีตัวที่ 3 คือ F ดังตารางที่ 3.3 ในกรณีโน้ตตัวที่ 5 ของสเกล D major อย่าง A จะอยู่ทั้งในเมเจอร์และไมเนอร์ ไม่สามารถใช้แยกประเภทเมเจอร์และไมเนอร์ได้



รูปที่ 3.5 Chromagram ของเพลง th_1.mp3



รูปที่ 3.6 ผลลัพธ์ค่าเฉลี่ย Chromagram ของ th_1.mp3 ในแต่ละคีย์เพลง

ตารางที่ 3.3 ความสัมพันธ์ของ D major และ D minor

	ตัวที่ 1	ตัวที่ 2	ตัวที่ 3	ตัวที่ 4	ตัวที่ 5	ตัวที่ 6	ตัวที่ 7
D major	D	Em	F#m	G	A	Bm	C#m
D minor	Dm	Em	F	Gm	Am	B	C

ผลลัพธ์ที่ได้ของเพลง th_1.mp3 เมื่ออ้างอิงจากทฤษฎีดนตรี มีแนวโน้มที่จะเป็นคีย์ D major และ F# minor ซึ่งจากการแกะเพลงตามทฤษฎีดนตรีเพลงนี้อยู่ในคีย์ D major ดังนั้นวิธีวิจัยนี้มีโอกาสที่จะได้ความถูกต้องตรงตามทฤษฎีดนตรี แต่เมื่อนำไปใช้กับหลาย ๆ เพลงที่อาจมีการใช้ทฤษฎีดนตรีที่ซับซ้อน พบว่าให้ผลลัพธ์ที่ไม่แม่นยำ แต่คีย์เพลงยังคงเป็นตัวแปรสำคัญต่ออารมณ์ ดังนั้นผู้วิจัยจึงนำค่าเฉลี่ยของทั้ง 12 คีย์มาใช้ในงานนี้

3.2.5 Timbre

Tone color หรือ Tone quality ทำให้รับรู้ว่าเสียงที่ได้ยินเป็นเสียงเครื่องดนตรีชนิดใด ในขณะที่เครื่องดนตรีต่าง ๆ เล่นโน้ตเดียวกัน โดยผู้วิจัยเลือกใช้วิธี Mel-frequency cepstral coefficients ซึ่งเป็นการปรับปรุงจากเทคนิคเซปตรัม หรือ Discrete cosine transform เนื่องจากมีการปรับสเกลของสเปกตรัมให้อยู่บนสเกลที่เหมาะสม สำหรับการรับฟังของมนุษย์และทำให้เก็บรายละเอียดสัญญาณเสียงช่วงความถี่ต่ำได้มากกว่า ซึ่งความถี่ต่ำจะบ่งบอกคุณลักษณะของเสียงได้ดีกว่าความถี่สูง โดยข้อมูลจะส่งผ่านชุดตัวกรอง (Filter bank) ใน Mel scale เพื่อเน้นความสำคัญของความถี่ที่อยู่ในช่วงกลางของชุดตัวกรองแต่ละตัว หาได้จากสมการ (3.4)

สมการ Discrete cosine transform เป็นดังนี้

$$c_m = \sum_{k=1}^{N} E_k \cos[m(k-\frac{1}{2})\frac{\pi}{N}] \quad (3.4)$$

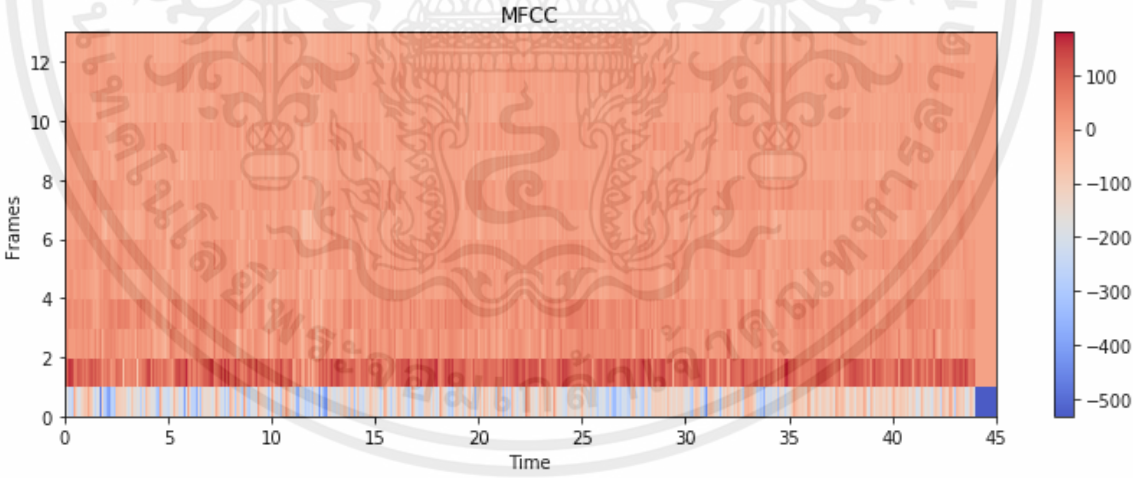
เมื่อ c คือ ค่าสัมประสิทธิ์เซปตรัม

m คือ จำนวนความถี่ Mel 1, 2, 3, ..., 13 หาจากสมการที่ (3.5)

แต่ละงานวิจัยมีการเลือกใช้จำนวน Feature ที่ได้จาก MFCCs แตกต่างกันไป ตั้งแต่ 1-40 โดยประมาณ โดยความถี่บน MFCCs สามารถหาได้จากการนำความถี่ปกติมาคำนวณในสมการที่ (3.5) แสดงผลลัพธ์ดังรูปที่ 3.7 โดยแกน X คือ mel 13 ลำดับ และ แกน Y คือ เวลา

$$mel(f) = 2595\log(1+\frac{f}{700}) \quad (3.5)$$

เมื่อ f คือ ความถี่ปกติ



รูปที่ 3.7 ตัวอย่างผลลัพธ์ของ Timbre

บทที่ 4
การเรียนรู้ด้วยเครื่อง

ในบทนี้กล่าวถึงการเรียนรู้ด้วยเครื่อง ซึ่งเป็นแขนงวิชาที่มีการนำไปใช้ในสายงานต่าง ๆ อย่างมาก ในปัจจุบัน การเรียนรู้ด้วยเครื่องมีความยืดหยุ่นในการประยุกต์กับโจทย์ต่าง ๆ ได้ตามที่ต้องการ สามารถแก้ปัญหาสมการต่าง ๆ ที่มนุษย์ไม่สามารถแก้ได้ ซึ่งผลลัพธ์จากการเรียนรู้ด้วยเครื่อง ก่อให้เกิดประโยชน์ในด้านต่าง ๆ มากมาย

4.1 การเรียนรู้ด้วยเครื่อง (Machine Learning)

การเรียนรู้ด้วยเครื่อง คือ กระบวนการที่ทำให้คอมพิวเตอร์สามารถคำนวณผลลัพธ์ได้อัตโนมัติ โดยอ้างอิงจากข้อมูลที่ผู้วิจัยได้วิเคราะห์ผลลัพธ์ และนำเข้าไปในระบบเพื่อให้ระบบเรียนรู้ กระบวนการนี้ทำให้คอมพิวเตอร์สามารถแก้สมการขั้นสูงที่มีความซับซ้อนและยากที่มนุษย์จะแก้สมการได้ การเรียนรู้ด้วยเครื่องในงานวิจัยนี้เลือกใช้แบบ Supervised learning

Supervised learning คือ กลุ่มของอัลกอริทึมที่สอนคอมพิวเตอร์ด้วยกลุ่มตัวอย่าง หรือ label แบบต่าง ๆ เมื่อระบบได้ทำการเรียนรู้จากข้อมูลตัวอย่าง ผลลัพธ์จะแสดงออกมาในรูปของ label นั้น ๆ ที่ได้เรียนรู้ไป เช่น ถ้าเรากำหนดคำนิยามอารมณ์เพลง เมื่อเอาคำนิยามของชุดข้อมูลทดลองไปให้ระบบเรียนรู้และสร้างโมเดลในการทำนายขึ้นมา เมื่อนำข้อมูลที่ต้องทดสอบเข้าไปในโมเดลเพื่อทราบคำนิยามอารมณ์เพลงนั้น โมเดลก็จะแสดงผลออกมาในรูปแบบของคำนิยาม

การเรียนรู้ด้วยเครื่องแบบ Supervised learning สามารถนำมาหาผลลัพธ์ได้ทั้งแบบ คำนิยาม (Label) และ ตัวเลข(Value) ซึ่งมาจากวิธีการที่แตกต่างกัน หากต้องการให้ผลลัพธ์ออกมาเป็นคำนิยาม เช่น หมา แมว ปลา ไก่ หรืออื่น ๆ ต้องใช้การจัดหมวดหมู่ (Classifier) เช่น Random forest [9], Naïve Bayes [10], Support vector machine [14] - [15] และหากต้องการให้ผลลัพธ์ออกมาเป็นตัวเลข เช่น 0.567, 0.2, 10 ต้องใช้การวิเคราะห์แบบถดถอย (Regression) [16]–[17] ซึ่งทั้ง 2 แบบ ถูกนำมาใช้ในงานวิจัยนี้ เริ่มต้นระบบจากวิธี Multiple linear regression

4.1.1 Multiple linear regression (MLR)

การประยุกต์ใช้ความรู้สถิติขั้นพื้นฐานโดยการหาความสัมพันธ์ของสองตัวแปร คือ X เป็นตัวแปรอิสระ(Independent variable) และ Y เป็นตัวแปรตาม (Dependent variable) ดังสมการที่ (4.1) แสดงสมการความสัมพันธ์ระหว่างตัวแปร X และ Y เรียกว่า Simple linear regression

$$Y = b_0 + b_1X_1 \quad (4.1)$$

เมื่อ Y คือ ตัวแปรตาม

X คือ ตัวแปรอิสระ

b_0 คือ ค่าสัมประสิทธิ์ บอกถึงจุดเริ่มต้นของค่า Y เมื่อตัวแปร X เท่ากับ 0

b_1 คือ ค่าสัมประสิทธิ์ของตัวแปร X_1

แต่ในสถานการณ์จริง ผลลัพธ์ต่าง ๆ ไม่ได้ขึ้นอยู่กับแค่สองตัวแปร แต่ขึ้นอยู่กับอีกหลายตัวแปร ทำให้มีความซับซ้อนและยากเกินที่มนุษย์จะแก้ได้ จึงต้องใช้คอมพิวเตอร์เข้ามาช่วย และเมื่อนำ MLR มาใช้ในงานวิจัยนี้จะแทนตัวแปรได้ ดังสมการที่ (4.2)

$$Y = b_0 + b_1X_1 + b_2X_2 + ... + b_iX_i \quad (4.2)$$

เมื่อ Y คือ ค่า Valence-Arousal

X คือ คุณสมบัติเพลง ได้แก่ Pitch, Dynamic, Tempo, Tonality

b_0 คือ ค่าสัมประสิทธิ์ บอกถึงจุดเริ่มต้นของค่า Y เมื่อตัวแปร X ทุกตัวเท่ากับ 0

i คือ จำนวนตัวแปร X

เมื่อนำสมการที่ (4.2) มาแทนค่า X ด้วยคุณสมบัติเพลง จะได้สมการที่ (4.3)

$$Y = b_0 + b_1(\text{Pitch}) + b_2(\text{Dynamic}) + b_3(\text{Tempo}) + b_4(\text{Tonality}[1]) + ... + b_{15}(\text{Tonality}[12]) \quad (4.3)$$

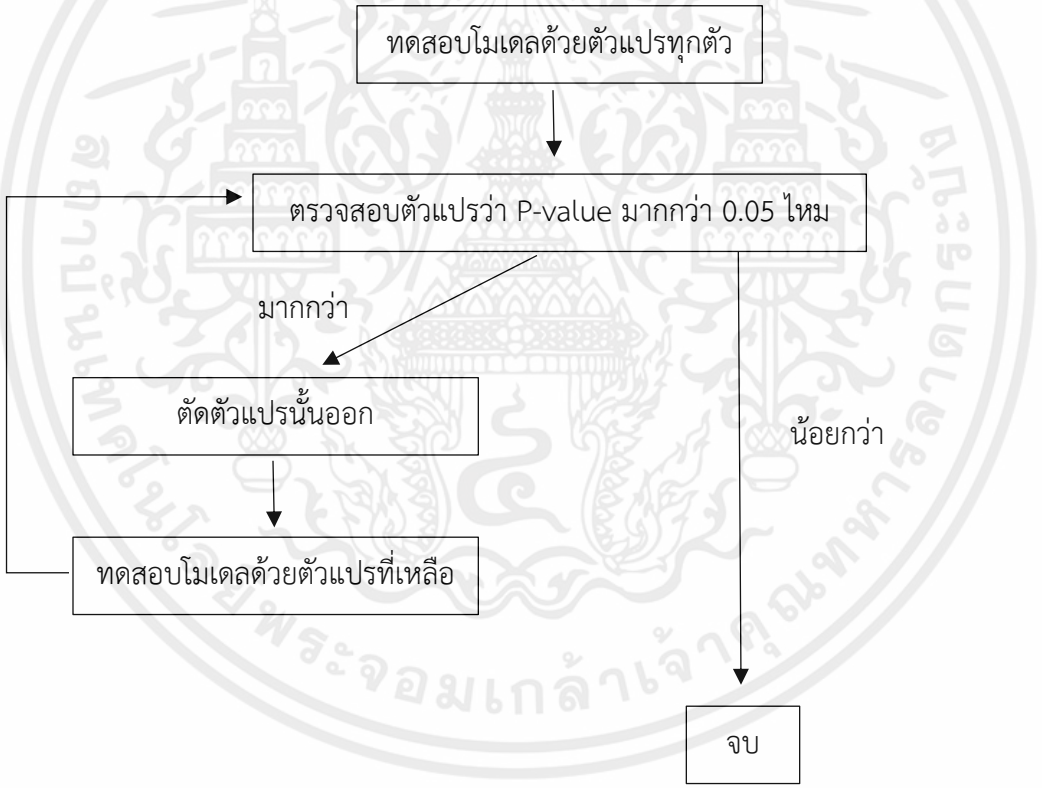
เมื่อได้ผลลัพธ์จากการคำนวณ ผู้วิจัยเลือกพัฒนาประสิทธิภาพโมเดลด้วยการตรวจสอบค่า P-value ของแต่ละตัวแปร หากตัวแปรไหนมี P-value มากกว่า ค่านัยสำคัญ(Significant level) ที่กำหนดไว้ มีโอกาสที่ตัวแปรนี้จะลดทอนประสิทธิภาพโมเดล หากตัดตัวแปรนี้ออกมีโอกาสที่ประสิทธิภาพโมเดลจะเพิ่มขึ้น

ค่านัยสำคัญมาจากการที่เรากำหนดค่าความเชื่อมั่น (Confident level) ในตัวแปรทั้งหมดของเรา งานวิจัยนี้เลือกที่จะกำหนดค่าความเชื่อมั่นของตัวแปรที่ 95% ซึ่งเป็นค่าที่นิยมใช้ ทั้งนี้สามารถปรับเปลี่ยนค่านี้ได้ตามความเหมาะสม เมื่อนำค่าความเชื่อมั่นมาคำนวณหาค่านัยสำคัญในสมการที่ (4.4)

จะได้ผลลัพธ์เท่ากับ 0.05 ซึ่งถ้าตัวแปรใดมีค่า P-value มากกว่า 0.05 ตัวแปรนั้นอาจลดทอนประสิทธิภาพของโมเดลได้ จึงต้องตัดตัวแปรนั้นออก

$$\text{Significant level} = 1 - \text{Confident level} \tag{4.4}$$

แต่ในความเป็นจริง ไม่ได้มีเพียงตัวแปรเดียวจากหลายตัวในโมเดลที่มีค่า P-value > 0.05 หากมีหลายตัวแปรที่ควรตัดออก ต้องตัดออกที่ละตัวแปรและดูประสิทธิภาพโมเดลที่เปลี่ยนไป ไม่ควรตัดทุกตัวออกพร้อมกัน เพราะในบางกรณี การตัดตัวแปรหนึ่งตัว อาจส่งผลให้ประสิทธิภาพของโมเดลเพิ่มขึ้นมาก และตัวแปรอื่น ๆ ที่เคยมีค่า P-value > 0.05 กลายเป็นว่าต่ำกว่าเกณฑ์ที่กำหนดไว้ ส่งผลให้ไม่จำเป็นต้องตัดตัวแปรนั้นออก วิธีนี้เรียกว่า Backward elimination แสดงในรูปที่ 4.1

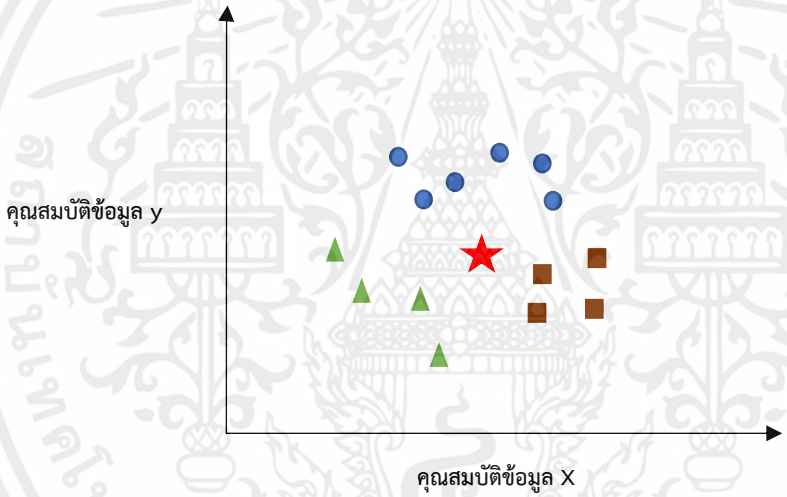


รูปที่ 4.1 Backward elimination

4.1.2 K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) เป็นหนึ่งวิธีการเรียนรู้ด้วยเครื่องแบบ Classification วิธีการ คือ หาความใกล้เคียงของข้อมูลที่ต้องการจะทราบ เมื่อเทียบกับข้อมูลอื่น ๆ ที่เป็นชุดทดลอง โดยใช้ Euclidean distance สมการที่ (4.5) ในการหาระยะทางของข้อมูล หากข้อมูลที่เราต้องการรู้คำตอบมีระยะทางใกล้เคียงกับข้อมูลไหนที่สุด ระบบจะตัดสินใจว่าเป็นข้อมูลประเภทเดียวกัน แต่วิธีนี้จะไม่จำกัดแค่เลือกแต่ระยะทางที่ใกล้ที่สุดอันดับแรกเท่านั้น เพราะถ้ากำหนดให้เปรียบเทียบแค่อันดับแรก อาจเกิดกรณีที่หลายข้อมูลมีระยะทางใกล้เคียงเท่ากัน ทำให้ระบบไม่สามารถตัดสินใจได้

ดังนั้น วิธีนี้จึงมีการกำหนดค่า K ซึ่ง คือจำนวนข้อมูลที่ต้องการเปรียบเทียบระยะทาง เช่น กำหนด K=3 ระบบจะทำการตรวจสอบข้อมูลสามอันดับแรกที่ใกล้ที่สุด และดูว่าข้อมูลทั้งสามนั้นมีคำตอบเป็นแบบใดมากสุด ระบบก็จะทำการตัดสินใจให้ข้อมูลนั้นตรงกับคำตอบที่มากที่สุด



รูปที่ 4.2 จำลองตัวอย่างข้อมูล เมื่อต้องการทราบประเภทของข้อมูลดาวสีแดง

รูปที่ 4.2 จำลองตัวอย่างของ KNN เมื่อกำหนดข้อมูลที่ต้องการทราบประเภท (รูปดาว) เป็นข้อมูลประเภทใด ซึ่งมีตัวเลือกอยู่ 3 ประเภท คือ สามเหลี่ยม สี่เหลี่ยม และวงกลม หากดูด้วยตาก็ยากที่จะระบุได้ว่าเป็นข้อมูลประเภทใด ดังนั้นจึงต้องทำการวัดระยะทางข้อมูลทุกตัวดังรูปที่ 4.3 ด้วยวิธี Euclidean distance ดังสมการที่ (4.5) และ สมการที่ (4.6)

สมการ Euclidean distance เมื่อข้อมูลมี 2 คุณสมบัติ

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (4.5)$$

เมื่อ

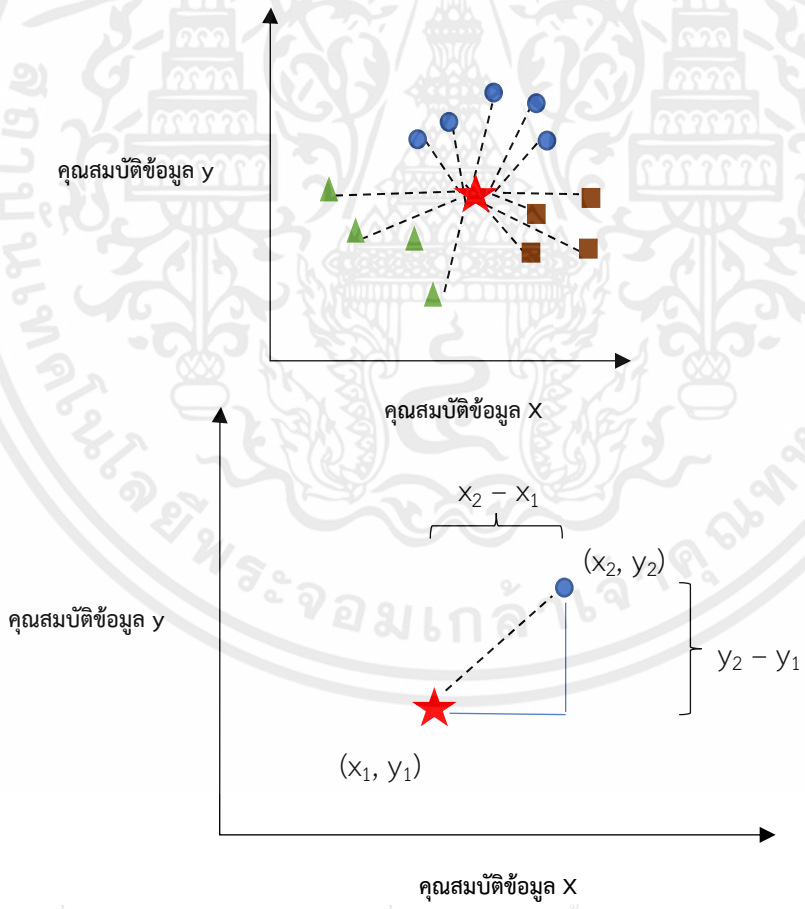
d คือ ระยะทาง

x คือ คุณสมบัติข้อมูล x

y คือ คุณสมบัติข้อมูล y

สมการ Euclidean distance เมื่อข้อมูลมีมากกว่า 2 คุณสมบัติ

$$d = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \quad (4.6)$$



รูปที่ 4.3 วัดระยะทางกับข้อมูลทุกตัว

ค่า K ควรกำหนดเป็นเลขคี่ เพื่อป้องกันปัญหาข้อมูลเป็นได้สองประเภท ตัวอย่าง ถ้ากำหนด K=4 เป็นเลขคู่ ข้อมูลที่ต้องการหาคำตอบอาจใกล้เคียงกับข้อมูลอีก 2 ประเภท ระบบจะไม่สามารถทำการตัดสินใจได้ การกำหนดค่า K มาก ไม่ได้หมายความว่าประสิทธิภาพของระบบจะเพิ่มขึ้น และอาจทำให้ประสิทธิภาพลดลงได้ เพราะไปครอบคลุมข้อมูลหลายประเภทที่ไม่ได้มีระยะทางใกล้เคียง

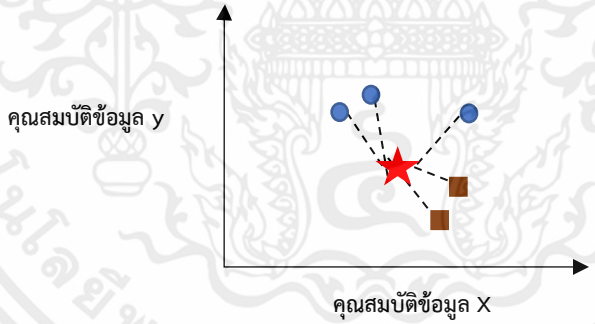
แต่ในกรณีที่ทำการวัดระยะทางข้อมูลทุกตัวแล้ว อาจมีความไม่ชัดเจนดังรูปที่ 4.4 เมื่อกำหนด K = 5 และแสดงผลออกมาว่านี่คือข้อมูล 5 ตัวที่มีระยะทางใกล้ที่สุด ข้อมูลวงกลมมีจำนวนข้อมูลใกล้เคียงมากกว่าสี่เหลี่ยม ทั้ง ๆ ที่ข้อมูลสี่เหลี่ยมมีระยะทางใกล้กว่า และดูเกาะกลุ่มกับข้อมูลดาวมากกว่า ดังนั้นจึงต้องนำระยะทางทั้งหมดมา Weighted nearest neighbors ด้วยสมการที่ (4.7)

$$w_i = \frac{d_k - d_i}{d_k - d_1} \quad ; d_k \neq d_1 \tag{4.7}$$

เมื่อ d_k คือ ระยะทางที่มากที่สุด

d_i คือ ระยะทางของข้อมูลที่นำมา Weight

d_1 คือ ระยะทางที่น้อยที่สุด

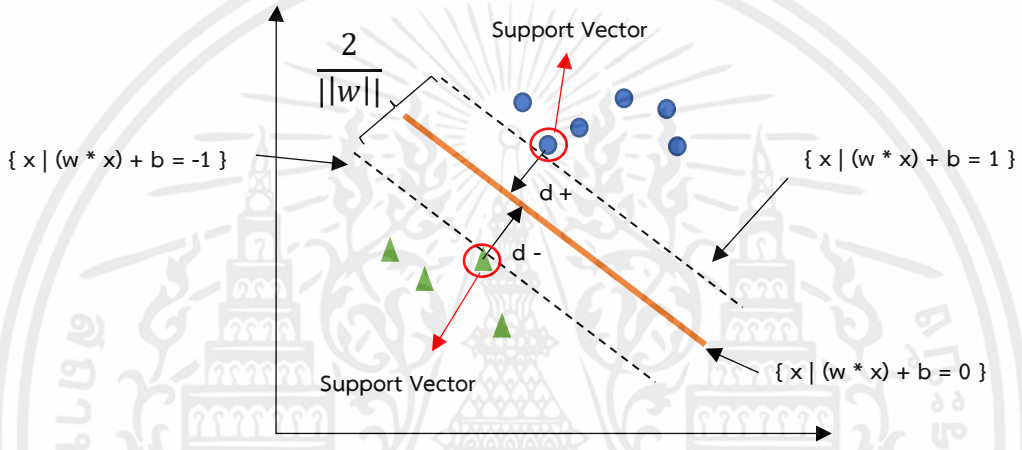


รูปที่ 4.4 เมื่อกำหนดค่า K = 5 แสดง 5 อันดับที่ใกล้เคียงที่สุด

เมื่อได้ผลลัพธ์จากการ Weight ให้นำข้อมูลประเภทเดียวกันมาบวกเข้าด้วยกัน ข้อมูลประเภทไหนมีค่ามากกว่า หมายความว่าข้อมูลที่เราต้องการหา เป็นประเภทเดียวกับข้อมูลนั้น

4.1.3 Support vector regression (SVR)

เทคนิคที่ใช้แก้ปัญหาทางด้านการรู้จำรูปแบบข้อมูล โดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการ เพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการการสอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งที่แยกแยะกลุ่มข้อมูลได้ดีที่สุด (Optimal Separating Hyperplane) เนื่องจากเมื่อนำข้อมูลมาวางลงใน space พบว่ามีเส้นตรงที่เป็นไปได้มากมายในการแยกแยะข้อมูล ดังนั้นจึงต้องหาเส้นแบ่งที่ดีที่สุดด้วยการเพิ่มเส้นขอบให้กับเส้นแบ่งทั้งสองข้าง ทำให้ได้เส้นใหม่ที่จะถือเป็นเส้นขอบของข้อมูลแต่ละฝั่ง



รูปที่ 4.5 จำลองรูปแบบข้อมูล Support vector regression

จากรูปที่ 4.5 hyperplane คือ เส้นสีส้ม ซึ่งเป็นเส้นหลักในการแบ่งแยกข้อมูลแบบ Linear ในรูปแบบสมการเส้นตรง y = mx + b โดยกำหนดค่า m เป็นค่า w เพื่อกำหนดเป็นสมการที่ (4.8) โดยจะสร้างเส้นตรงสองเส้นขนาบข้าง ดังสมการที่ (4.9) และ (4.10) เพื่อแบ่งว่า ข้อมูลที่เราต้องการคำตอบนั้นเป็นแบบ 1 หรือ -1 เส้นตรงทั้งสองเส้นนี้จะอยู่ติดกับข้อมูลแต่ละประเภท เรียกข้อมูลที่ติดกับเส้นขอบว่า Support Vector โดยมีระยะห่างจาก hyperplane เรียกว่า ระยะขอบ (Margin)

$$(w * x) + b = 0 \tag{4.8}$$

$$(w * x) + b = 1 \tag{4.9}$$

$$(w * x) + b = -1 \tag{4.10}$$

$$f(x) = \begin{cases} 1 & \text{if } (w * x) + b \geq 1 \\ -1 & \text{if } (w * x) + b \leq -1 \end{cases} \tag{4.11}$$

เมื่อ w คือ เวกเตอร์ตั้งฉากกับไฮเปอร์เพลน

b คือ ค่าคงที่แกน y ที่ได้จากข้อมูล X แต่ละค่า

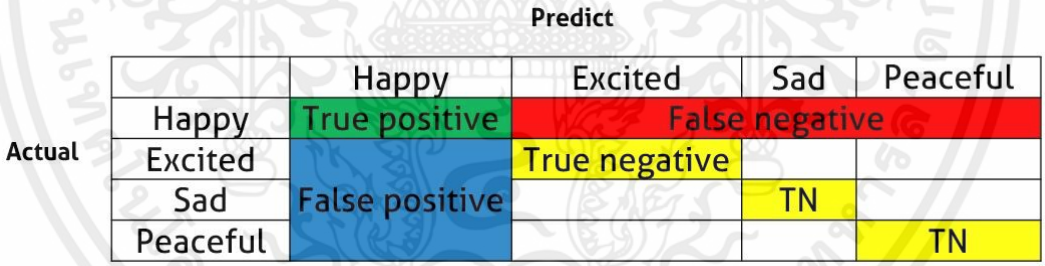
x คือ คุณสมบัติข้อมูลในแกน X

หากเส้นขอบของเส้นแบ่งใด แสดงค่า 2/||w|| มากที่สุด แสดงให้เห็นว่าข้อมูลสองชุดมีการแยกกันชัดเจนมากที่สุด ระบบจึงเลือกไฮเปอร์เพลนนั้น มาหาคำตอบให้กับข้อมูล

4.2 การวัดประสิทธิภาพ (Performance measurement)

ผลลัพธ์ที่ได้จากการเรียนรู้ด้วยเครื่องจะต้องนำมาวัดประสิทธิภาพในรูปแบบต่าง ๆ ในงานวิจัยนี้เลือกใช้ทั้งหมด 4 แบบ ด้วยกัน ได้แก่ ค่าความถูกต้อง (Accuracy), ค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และ F-measure ซึ่งแต่ละแบบจะให้ผลลัพธ์ที่บ่งบอกถึงประสิทธิภาพของโมเดลแต่ละด้าน สามารถคำนวณได้จากสมการที่ (4.12)-(4.16) ตามลำดับ การวัดประสิทธิภาพ ผู้วิจัยต้องเข้าใจนิยามของ True positive, True negative, False positive และ False negative แสดงใน Confusion matrix ตารางที่ 4.1

ตารางที่ 4.1 ตัวอย่าง Confusion matrix



ตารางที่ 4.2 ตัวอย่าง Confusion matrix อารมณ์ 4 แบบ

	Happy	Excited	Sad	Peaceful
Happy	41	0	8	5
Excited	12	0	2	1
Sad	5	1	21	2
Peaceful	9	0	10	8

แกน Y คือ อารมณ์ที่แท้จริง (Actual) แกน X คือ อารมณ์จากการทำนาย (Predict) ในตารางที่ 4.2 แสดงตัวอย่างการวัดประสิทธิภาพอารมณ์ Happy เมื่ออารมณ์จากการทำนายตรงกับอารมณ์ที่แท้จริง จะเรียกว่า True positive (สีเขียว) และอารมณ์จากการทำนายอื่น ๆ ที่ตรงกับอารมณ์ที่แท้จริง เรียกว่า True negative (สีเหลือง) สองค่านี้จะใช้ร่วมกัน เมื่อต้องการหาค่าความถูกต้อง ดังสมการที่ (4.12) ซึ่งบ่งบอกถึงความถูกต้องของระบบเมื่อเทียบกับจำนวนเพลงทั้งหมด แต่ค่าความถูกต้องมีความไม่น่าเชื่อถือในกรณีที่ข้อมูลกลุ่ม True ไม่เกิดการกระจายตัว

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{Total data}} \tag{4.12}$$

ค่าความแม่นยำ หรือ Precision บ่งบอกถึงประสิทธิภาพในการทำนายของโมเดล สามารถคำนวณได้โดยสมการที่ (4.13) โดย False positive (สีฟ้า) จากตารางที่ 4.1 และ 4.2 แสดงให้เห็นว่าโมเดลทำนายได้ผลลัพธ์เป็น Happy แต่เมื่อนำไปเปรียบเทียบกับอารมณ์ที่แท้จริง ผลลัพธ์กลับไปตรงกับอารมณ์อื่น ๆ ถึง 26 เพลง บ่งบอกได้ว่ามีประสิทธิภาพในด้านปริมาณ แต่ไม่มีประสิทธิภาพในด้านคุณภาพของการทำนาย เพราะทำนายอารมณ์ได้ Happy ตามต้องการ

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \tag{4.13}$$

ค่าความระลึก หรือ Recall บ่งบอกถึงประสิทธิภาพโมเดลเมื่อเทียบกับอารมณ์ที่แท้จริง แสดงให้เห็นว่าอารมณ์ที่แท้จริงแต่ละประเภท ตรงกับที่ทำนายเท่าไหร่ โดยค่า False negative จากภาพที่ และตารางที่ แสดงให้เห็นว่าอารมณ์ Happy ที่แท้จริงมีการส่งข้อมูลที่ผิดพลาดไปตรงกับอารมณ์อื่น ๆ ถึง 13 เพลง บ่งบอกว่าประสิทธิภาพเชิงปริมาณของโมเดลไม่ดี สามารถคำนวณได้โดยสมการที่ (4.14)

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \tag{4.14}$$

F-measure คือ การวัดประสิทธิภาพโดยรวมของระบบ มีความน่าเชื่อถือกว่าค่าความถูกต้อง เพราะนำค่าความแม่นยำ และค่าความระลึก ที่คำนวณความเสียหายในแต่ละด้านมาคำนวณในสมการที่ (4.15)

$$\text{F-measure} = 2 \text{ x } \frac{\text{Precision x Recall}}{\text{Precision + Recall}} \quad (4.15)$$

ค่าความแม่นยำ ค่าความระลึก และ F-measure ของอารมณ์แต่ละประเภท แต่เมื่อนำมาเปรียบเทียบประสิทธิภาพของโมเดล จะนำค่าของทุกอารมณ์มาหาค่าเฉลี่ยของค่านั้น เพื่อสรุปว่าโมเดลนั้นมีค่าความแม่นยำ ค่าความระลึก และ F-measure เท่าไหร่ โดยสมการที่ (4.16) มีการนำไป Weighting กับปริมาณข้อมูล จึงมีความน่าเชื่อถือกว่าค่าความถูกต้อง ที่ใช้แต่ข้อมูลในกลุ่ม True โดยไม่นำมา Weighting กับปริมาณข้อมูลทั้งหมด ในงานวิจัยด้านการเรียนรู้ด้วยเครื่องมีการอภิปรายผล F-measure อยู่เสมอ

$$\text{Average} = \frac{\sum(\text{Value x Amount of actual data})}{\text{Total data}} \quad (4.16)$$

เมื่อ Value คือ ค่าความแม่นยำ ค่าความระลึก และ F-measure ของอารมณ์ทุกประเภท

Amount of actual data คือ จำนวนข้อมูลของอารมณ์ที่แท้จริง ในอารมณ์นั้น ๆ

ตารางที่ 4.2 แสดงให้เห็นว่า เมื่อนำไปคำนวณค่าความถูกต้อง แล้ว ได้ผลลัพธ์ 56% ซึ่งในด้านอารมณ์ค่อนข้างเพียงพอต่อความน่าเชื่อถือ แต่เมื่อวิเคราะห์ผลลัพธ์จริง ๆ จะพบว่าข้อมูลที่ถูกต้องไม่เกิดการกระจายตัว เพราะ Excited มีค่าเท่ากับ 0 ซึ่งหมายความว่าไม่มีการทำนายอารมณ์นี้ได้ถูกต้องเลย ค่าความถูกต้อง 56% นั้นมาจาก Happy เป็นส่วนใหญ่ ในกรณีแบบนี้จะถือว่าค่าความถูกต้อง ไม่มีความน่าเชื่อถือ ดังนั้น F-measure จึงเป็นทางเลือกที่น่าเชื่อถือที่สุด ในการวัดประสิทธิภาพของระบบกรณีที่ข้อมูลที่ถูกต้องไม่เกิดการกระจายตัว แต่การจะหาค่า F-measure จำเป็นจะต้องหาค่าความแม่นยำ และค่าความระลึกก่อน

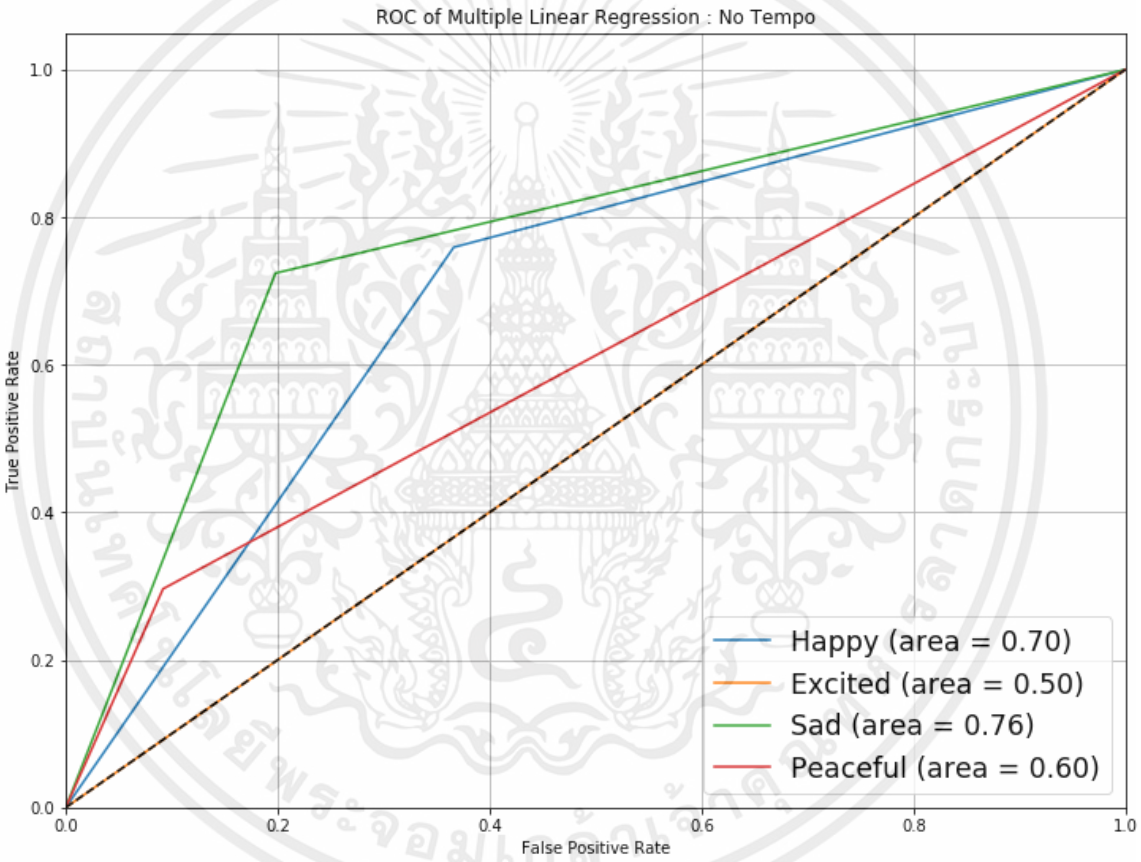
4.3 ROC Curve

Receiver operating characteristice curve หรือ ROC curve เป็นวิธีวัดประสิทธิภาพโมเดลวิธีหนึ่ง แสดงกราฟความสัมพันธ์ระหว่าง True positive และ False positive ซึ่งสองค่านี้บ่งบอกประสิทธิภาพในด้าน Precision ดังสมการที่ (4.13) หมายความว่าผลลัพธ์ของพื้นที่ใต้กราฟ หรือ Area under the curve (AUC) ของ ROC curve ดังรูปที่ 4.6 จะแสดงให้เห็นว่าโมเดลมีประสิทธิภาพในการทำนายแต่ละข้อมูลแม่นยำเพียงใด ซึ่งโมเดลนี้มีอัตราการทำนายอารมณ์ Sad มากที่สุด มีค่าเฉลี่ยโมเดลที่

0.71 ตารางที่ 4.3 แสดงระดับพื้นที่ใต้กราฟ หากข้อมูลมีค่าเข้าใกล้ 1 แสดงว่า โมเดลมีประสิทธิภาพสูง และไม่ควรต่ำกว่า 0.5

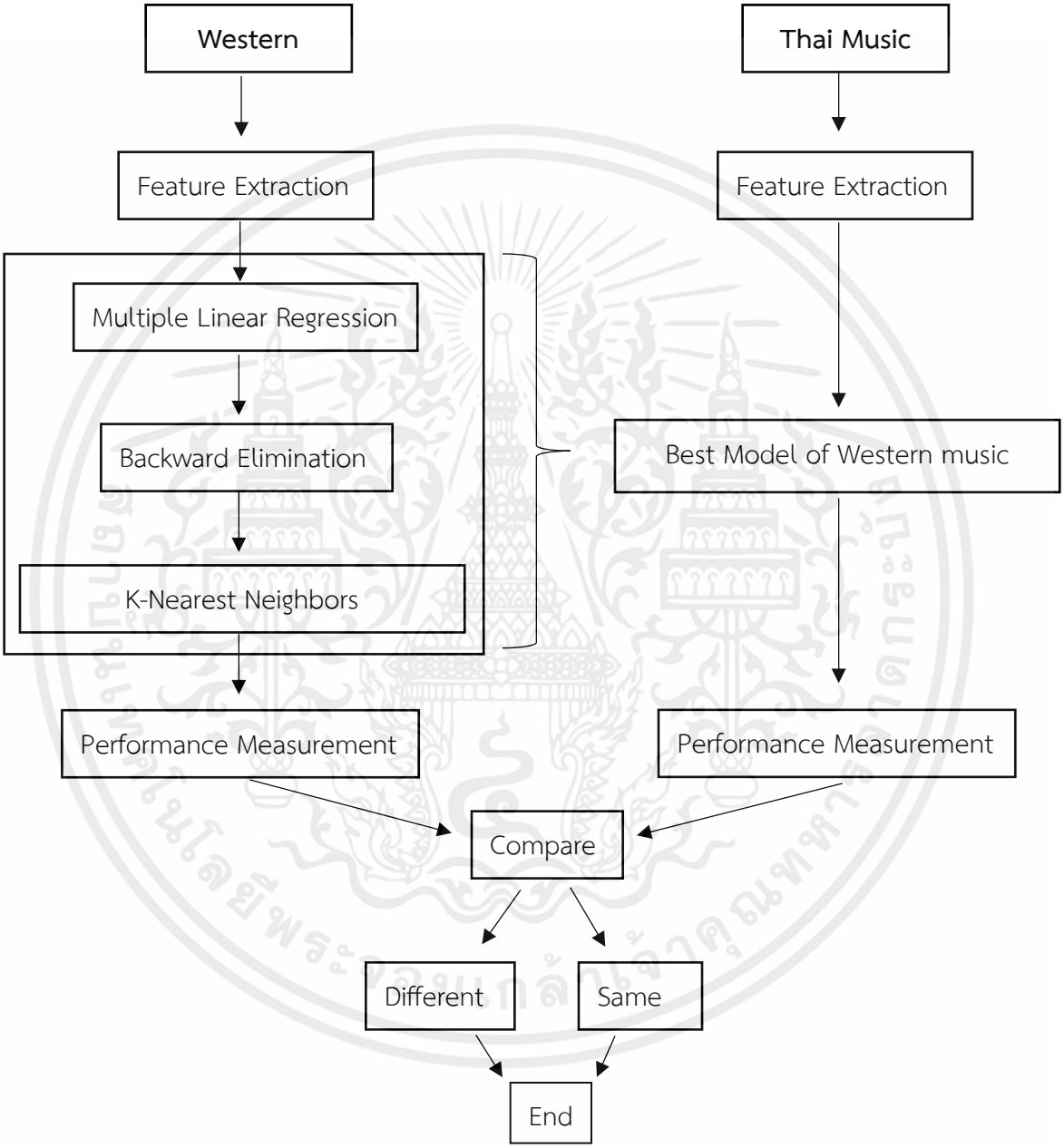
ตารางที่ 4.3 ระดับค่าประสิทธิภาพกราฟ ROC

	Excellent	Good	Fair	Poor	Fail	Worthless
AUC	1-0.9	0.89-0.8	0.79-0.7	0.69-0.6	0.59-0.5	< 0.5



รูปที่ 4.6 กราฟ ROC

บทที่ 5
การออกแบบระบบและผลการทดลอง

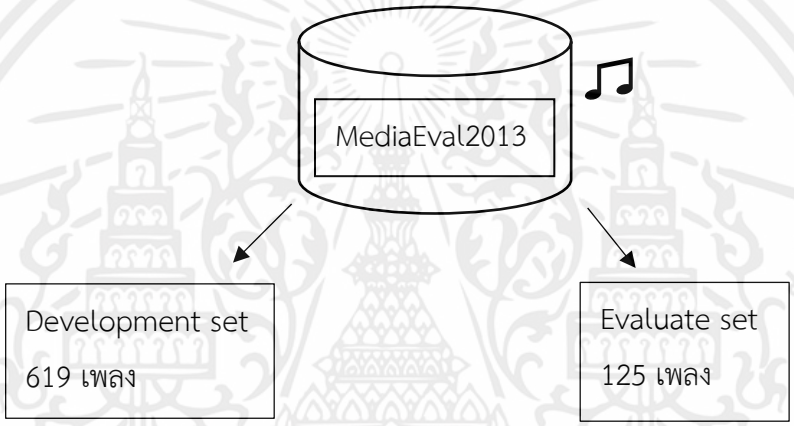


รูปที่ 5.1 การออกแบบระบบรู้จำอารมณ์เพลง

5.1 ระบบรู้จำอารมณ์เพลง

ระบบรู้จำอารมณ์เพลงในงานวิทยานิพนธ์ฉบับนี้ สร้างขึ้นจากพื้นฐานความรู้ด้านประมวลผลสัญญาณ การเรียนรู้ของเครื่อง จิตวิทยา และทฤษฎีดนตรี เริ่มต้นระบบจากการจัดเพลงในชุดข้อมูล MediaEval2013 ให้กลายเป็น 2 กลุ่ม ได้แก่ ชุดทดลอง และชุดทดสอบ ดังรูปที่ 5.1 ซึ่งผู้สร้าง MediaEval2013 ได้กำหนดมาแล้วว่าเพลงไหน คือ ชุดทดลอง และชุดทดสอบ

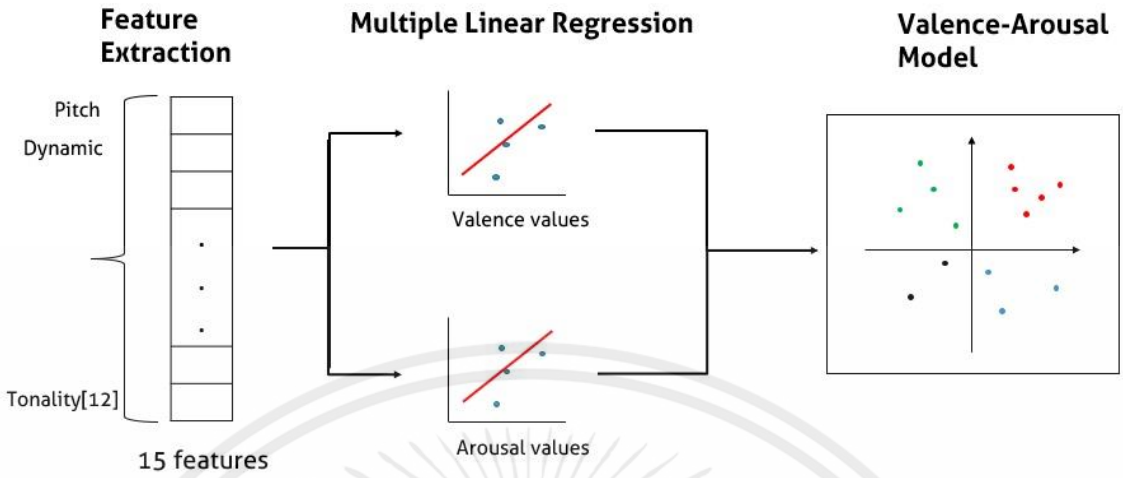
ผลลัพธ์จากการแบ่งชุดข้อมูล ได้ชุดทดลอง 619 เพลง และ ชุดทดสอบ 125 เพลง แต่ละเพลงจะมีค่าอารมณ์ในแต่ละช่วงเวลา ซึ่งผู้วิจัยได้นำหาค่าเฉลี่ยอารมณ์ Valence-Arousal แต่ละเพลง และนำมาใช้ในงานนี้



รูปที่ 5.2 ชุดทดลอง และ ชุดทดสอบ MediaEval2013

ค่าอารมณ์ Valence-Arousal คือ ค่าที่ใช้กำหนดคำนิยาม โดยนำค่าไปพล็อตบนแผนภาพ Valence-Arousal เพื่อทราบว่าแต่ละเพลงอยู่บริเวณไหนของแผนภาพ ซึ่งแต่ละบริเวณก็จะมีอารมณ์แตกต่างกันไป ในงานวิจัยนี้เพื่อลดความไม่ชัดเจนของอารมณ์ จึงกำหนดแต่ละควอแดรนท์ให้มีเพียงคำนิยามอารมณ์เดียว คือ Happy ควอแดรนท์ 1, Excited ควอแดรนท์ 2, Sad ควอแดรนท์ 3 และ Peaceful ควอแดรนท์ 4 หมายความว่า ไม่ว่าเพลงจะอยู่บริเวณไหนของแผนภาพ คำนิยามอารมณ์ของเพลง ๆ นั้น ก็จะเป็นคำนิยามอารมณ์ในควอแดรนท์ที่ตรงกับเพลงนั้น

การหาค่า Valence-Arousal ต้องใช้การเรียนรู้ด้วยเครื่องแบบ Regression เพราะต้องการคำตอบประเภทตัวเลข วิทยานิพนธ์ฉบับนี้เลือก Multiple linear regression (MLR) เพื่อหาค่าอารมณ์ เนื่องจากวิธีนี้เป็นวิธีพื้นฐานและให้ประสิทธิภาพที่ดี เหมาะแก่การเริ่มต้นศึกษางานด้านนี้



รูปที่ 5.3 Feature extraction และ Multiple linear regression

เมื่อทำการแบ่งเพลง MediaEval2013 แล้วต้องนำชุดทดลองมาทำกระบวนการประมวลผลสัญญาณและดึงคุณลักษณะเด่น (Feature extraction) ประกอบไปด้วยคุณสมบัติของเพลง 4 แบบ ได้แก่ Pitch, Dynamic, Tempo และ Tonality ดังตารางที่ 5.1 โดย Timbre จะยังไม่ถูกนำมาใช้

ตารางที่ 5.1 คุณสมบัติเพลง 4 แบบใน MLR

Features	No. of features	Methods
Pitch	1	Discrete Fourier transform
Dynamic	1	Root mean square energy
Tempo	1	Estimate tempo
Tonality	12	Chromagram

Tonality จะแตกต่างกับอีก 3 คุณสมบัติ เพราะจะใช้ถึง 12 ฟีเจอร์ที่ได้จาก Tonality ซึ่งแสดงถึงคีย์เพลงตามหลักทฤษฎีดนตรี ดังนั้นคุณสมบัติของเพลงทั้งหมดที่ใช้ในระบบนี้ตอนเริ่มต้น จะมีทั้งหมด 15 ฟีเจอร์ (Feature) เหตุผลที่เลือกคุณสมบัติเพลงเหล่านี้ เพราะเป็นคุณสมบัติพื้นฐานในด้านทฤษฎีดนตรี

นำข้อมูลที่ได้ทั้งหมดมาเก็บบันทึกเป็นไฟล์ .csv เพื่อง่ายต่อการเรียกใช้ใน Python โดยไฟล์ .csv จะประกอบไปด้วยคุณสมบัติเพลงทั้ง 15 ฟีเจอร์ รหัสเพลง ค่าอารมณ์ Valence-Arousal คำนิยามอารมณ์และแนวเพลง เมื่อทำการจัดระเบียบไฟล์แล้ว จึงนำข้อมูลไปหาค่า Valence-Arousal โดยวิธี Multiple

linear regression ในส่วนนี้จะแยกโมเดลเพื่อหาค่า Valence-Arousal ซึ่งไม่สามารถที่จะหาพร้อมกันได้ ดังรูปที่ 5.3 ดังนั้น ค่าความแม่นยำ ค่าความผิดพลาด และอื่น ๆ ของทั้งสองค่าอารมณ์ จะแตกต่างกัน

เมื่อได้ค่า Valence-Arousal ของแต่ละเพลงแล้ว จึงนำทั้งสองค่านี้มาพล็อตลงบนแผนภาพ เพื่อที่จะทำให้เพลงมีคำนิยามอารมณ์ที่แน่ชัด ได้แก่ Happy, Excited, Sad และ Peaceful ซึ่งงานวิจัยนี้ แม้จะเริ่มต้นจากการหาค่าเป็นตัวเลข แต่วิธีการทดสอบประสิทธิภาพโมเดลที่ใช้ในงานวิจัยนี้ จะตรวจสอบความแม่นยำจากการให้ผลลัพธ์คำนิยามที่ตรงกัน

เมื่อแต่ละเพลงได้รับคำนิยามอารมณ์แล้ว ผู้วิจัยจะทำการตรวจสอบประสิทธิภาพของระบบด้วยค่าความถูกต้อง เป็นวิธีแรก เพื่อที่จะทราบว่าค่าความถูกต้อง ให้ประสิทธิภาพที่ดีไหม จำเป็นต้องใช้ค่า F-measure ไหม

หลังจากตรวจสอบค่าความถูกต้อง และทำ Backward elimination ได้นำโมเดลที่มีประสิทธิภาพที่สุดทั้งหมด ไปทดสอบใน KNN ที่กำหนดค่า K เท่ากับ 1 ถึง 30 เพื่อหาโมเดลที่มีประสิทธิภาพที่สุดจาก K-NN และนำโมเดลจาก MLR และ KNN ที่ดีที่สุด ไปทดสอบกับเพลงไทย

เพลงไทย 125 เพลงจะมีค่าอารมณ์ Valence-Arousal จาก Spotify API โดยมีค่าระหว่าง 0 ถึง 1 ซึ่งแตกต่างจาก MediaEval2013 แต่ไม่สามารถทำการ Normalize ได้ เพราะค่าอารมณ์ที่ละเอียดอ่อน ทำให้อารมณ์เพลงเกิดการเปลี่ยนแปลงไปจากเดิมอย่างมาก ผู้วิจัยจึงนำค่า Valence-Arousal ไปพล็อตลงบนแผนภาพ Valence-Arousal โดยแบ่งเป็น 4 ควอแดรนท์ บนสเกล 0 ถึง 1 เพื่อระบุคำนิยามอารมณ์ Happy, Excited, Sad และ Peaceful ให้กับเพลงไทยเช่นกัน

เมื่อระบุอารมณ์ให้เพลงไทยทั้ง 125 เพลงแล้ว จึงนำเพลงไทยทั้งหมดไปตรวจสอบอารมณ์ในโมเดลที่มีประสิทธิภาพที่สุดจากพลงตะวันตก เพื่อที่จะทราบอารมณ์เพลงไทย โมเดลที่มีประสิทธิภาพที่สุดในการทำนายเพลงไทย และเปรียบเทียบว่าโมเดลที่มีประสิทธิภาพที่สุดระหว่างเพลงไทย และเพลงตะวันตกมีความคล้ายคลึงกันไหม หากว่าไม่ มีความแตกต่างที่สามารถอธิบายได้เช่นไร

5.2 ผลการทดลอง

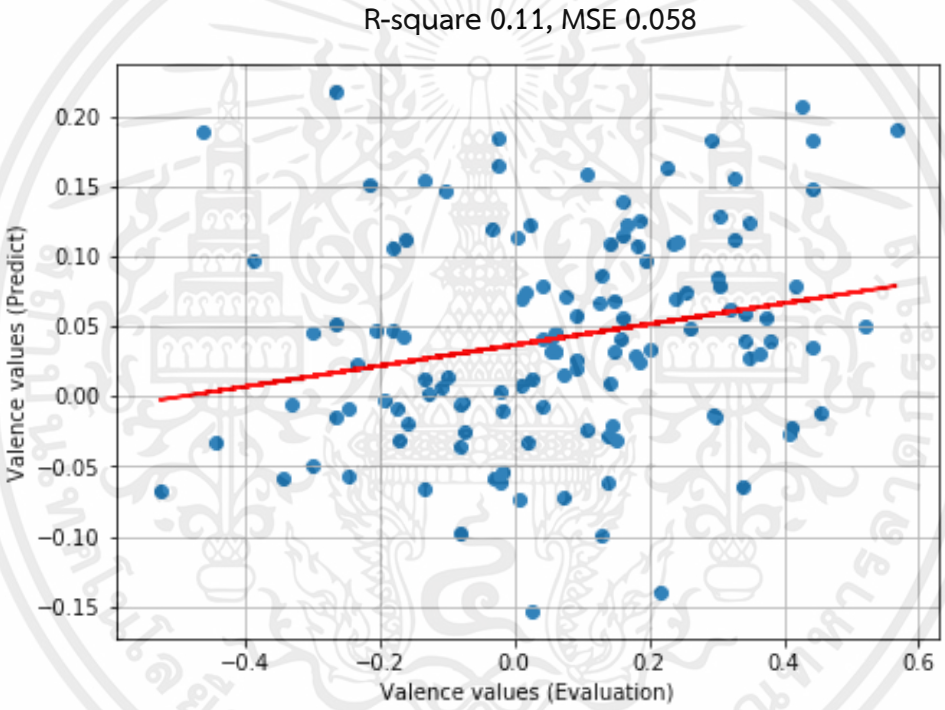
5.2.1 Multiple linear regression (MLR)

ผลลัพธ์จาก MLR จะแยกผลลัพธ์ของ Valence-Arousal ในตารางที่ 5.2 แสดงค่า R-squared และ MSE ในส่วนของ Valence ค่า R-squared เท่ากับ 0.11 ค่า MSE เท่ากับ 0.058 และ Arousal ค่า R-squared เท่ากับ 0.525 ค่า MSE เท่ากับ 0.042 ผลลัพธ์ของ Valence-Arousal ในส่วนนี้ค่อนข้างแตกต่างกัน โดยที่ Valence มีค่า R-square ที่ต่ำมาก แต่ในงานด้านอารมณ์ การที่ได้ค่าประมาณนี้ไม่ใช่เรื่องผิดปกติ เนื่องจากอารมณ์มีความไม่ชัดเจนในการนิยาม และในงานวิจัยอื่น ๆ ที่ใช้วิธีคล้าย ๆ กัน ก็ได้ผลลัพธ์ของ Valence ที่ต่ำเช่นกัน [17]

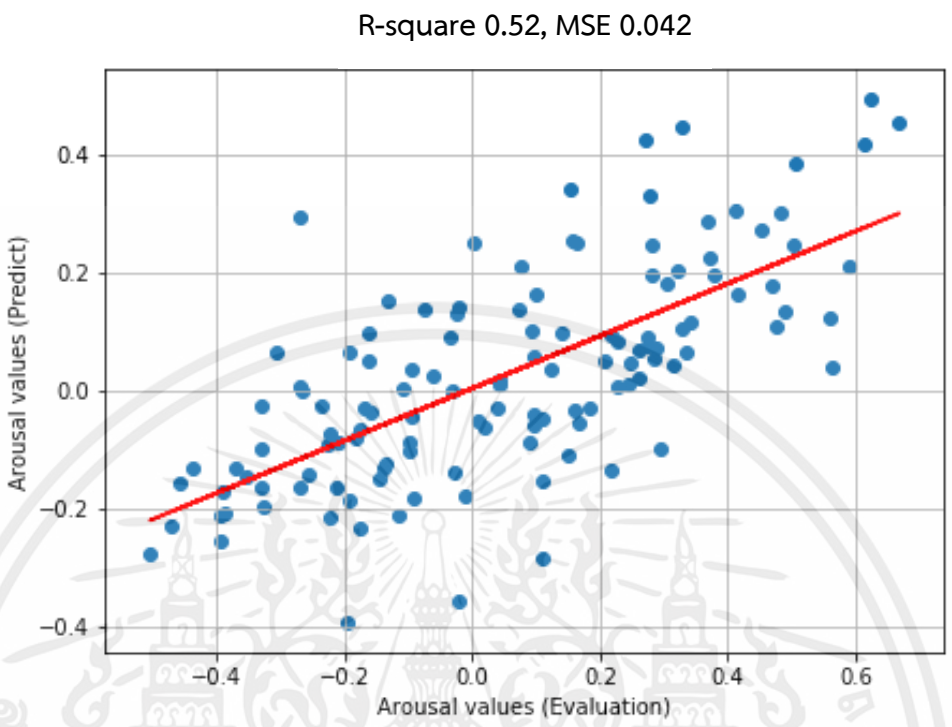
ตารางที่ 5.2 ค่า R-square และ MSE จาก MLR

	R-square	MSE
Valence	0.11	0.058
Arousal	0.52	0.042

รูปที่ 5.4 และ 5.5 แสดงความสัมพันธ์ของค่า Valence-Arousal โดยแกน X คือ ผลลัพธ์ที่ทำนายได้ และแกน Y คือ ค่าจริง ผลลัพธ์แสดงให้เห็นว่า Arousal ที่มีค่า R-square 0.52 มีรูปแบบข้อมูลที่เป็นเชิง Linear มากกว่า Valence อย่างเห็นได้ชัด



รูปที่ 5.4 ความสัมพันธ์ระหว่าง Valence ค่าจริง และ Valence จากการทำนาย



รูปที่ 5.5 ความสัมพันธ์ระหว่าง Arousal ค่าจริง และ Arousal จากการทำนาย

ตารางที่ 5.3 และ 5.4 แสดงค่า P-value และค่าสัมประสิทธิ์ของแต่ละคุณสมบัติเพลงของ Valenceโดยที่ค่า b_0 มีค่าสัมประสิทธ์เท่ากับ -0.2413 และจะมีค่า P-value เท่ากับ 0 เสมอ เพราะเป็นค่าที่บ่งบอกผลลัพธ์เมื่อทุกคุณสมบัติเพลงมีค่าเท่ากับ 0 หมายความว่าถ้าคุณสมบัติเพลงทุกแบบมีค่าอินพุตเท่ากับ 0 ค่า Arousal จะเท่ากับ -0.8131

เมื่อพิจารณา Valence-Arousal พบว่าค่าสัมประสิทธิ์ของ Dynamic มีค่ามากที่สุด ที่ Valence เท่ากับ 0.5925 และ Arousal เท่ากับ 1.2750 หมายความว่า Dynamic คือ คุณสมบัติเพลงที่ส่งผลต่ออารมณ์มากที่สุด โดยเฉพาะ Arousal ที่แสดงผลในด้านความตื่นตัว และเมื่อดูที่ค่า P-value จะพบว่ามีความสัมพันธ์กับค่าสัมประสิทธิ์แบบผกผัน เมื่อค่าสัมประสิทธิ์มากค่า P-value จะยิ่งต่ำ

ตารางที่ 5.3 ค่าสัมประสิทธิ์ และ P-value ของ Valence

ลำดับค่าสัมประสิทธิ์	คุณสมบัติเพลง	ค่าสัมประสิทธิ์	P-value
0	b_0	-0.2413	0
1	Pitch	0.00002374	0.437
2	Dynamic	0.5925	0
3	Tempo	0.0004	0.427
4	Tonality[1]	-0.0248	0.737
5	Tonality[2]	0.0931	0.322
6	Tonality[3]	0.0324	0.645
7	Tonality[4]	0.1314	0.150
8	Tonality[5]	0.0509	0.489
9	Tonality[6]	-0.0495	0.534
10	Tonality[7]	0.0368	0.671
11	Tonality[8]	0.0679	0.353
12	Tonality[9]	0.0004	0.996
13	Tonality[10]	0.1316	0.060
14	Tonality[11]	-0.0295	0.725
15	Tonality[12]	-0.0165	0.825

ตารางที่ 5.4 ค่าสัมประสิทธิ์ และ P-value ของ Arousal

ลำดับค่าสัมประสิทธิ์	คุณสมบัติเพลง	ค่าสัมประสิทธิ์	P-value
0	b_0	-0.8131	0
1	Pitch	0.00001424	0.599
2	Dynamic	1.2750	0
3	Tempo	0.0010	0.016
4	Tonality[1]	0.0386	0.556
5	Tonality[2]	0.1730	0.038
6	Tonality[3]	0.1758	0.005
7	Tonality[4]	0.1305	0.107
8	Tonality[5]	0.0743	0.256
9	Tonality[6]	0.2044	0.004
10	Tonality[7]	0.0824	0.284
11	Tonality[8]	0.1498	0.021
12	Tonality[9]	0.1116	0.166
13	Tonality[10]	0.1920	0.002
14	Tonality[11]	0.0013	0.986
15	Tonality[12]	0.1771	0.008

คุณสมบัติเพลงอื่น ๆ นอกจาก Dynamic มีค่า P-value > 0.05 เป็นส่วนใหญ่ โดยเฉพาะ Pitch ที่มีค่า P-value ที่สูงมากทั้ง Valence 0.437 และ Arousal 0.599 ซึ่งเมื่อพิจารณาความสัมพันธ์ของคุณสมบัติเพลงและ Valence-Arousal ของเพลงตะวันตกทั้งหมด ในรูปที่ 5.6 จะพบว่า Pitch มีการเกาะกลุ่มของข้อมูลเป็นอย่างมาก และไม่ได้มีความสัมพันธ์ที่เพิ่มขึ้นแบบ Linear ซึ่งหมายความว่าค่า Pitch ที่มาก ไม่ได้ส่งผลให้ค่า Valence-Arousal เพิ่มขึ้นอย่างเป็นรูปแบบ

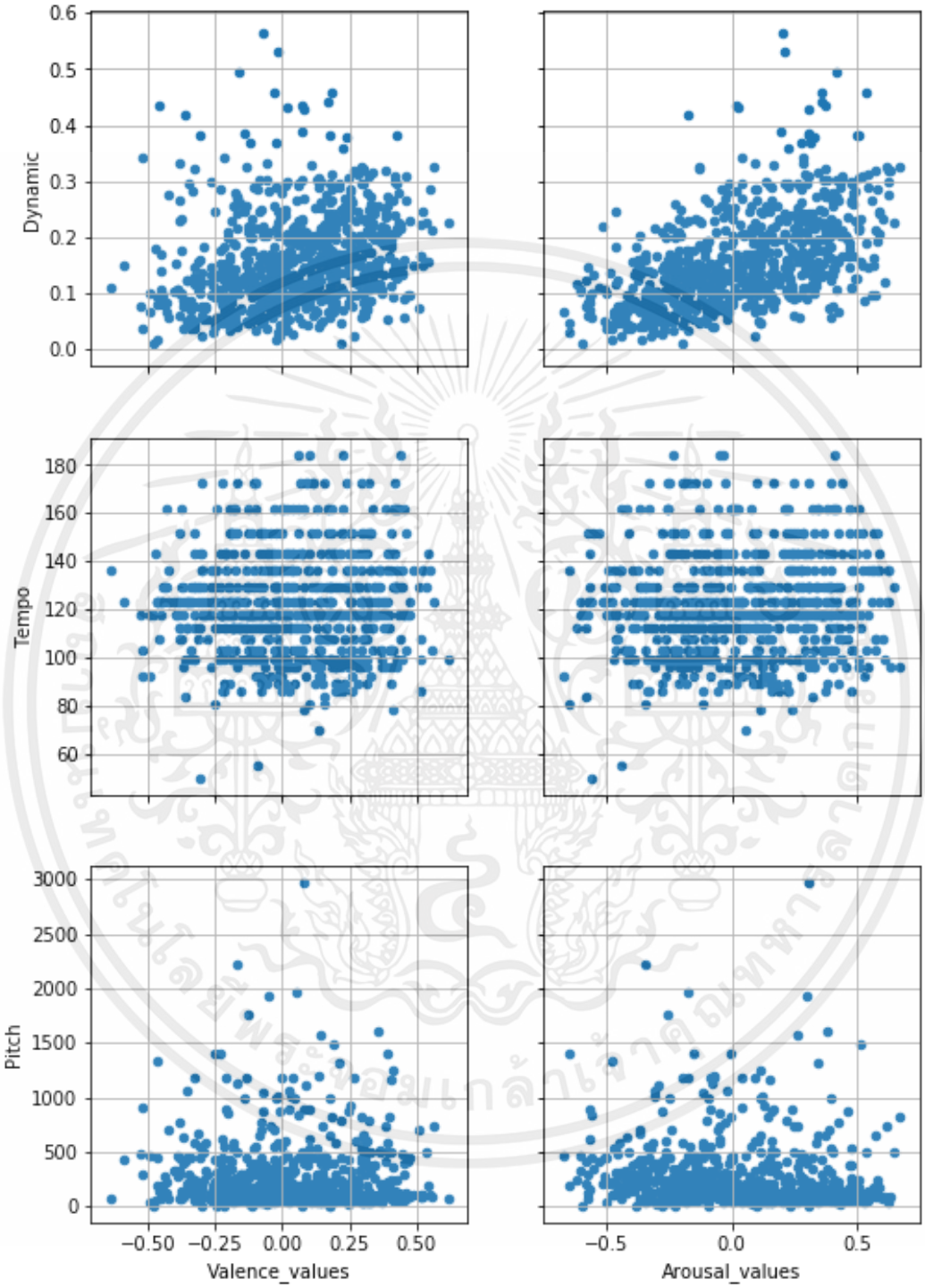
Tempo มีค่า P-value ของ Valence-Arousal ที่แตกต่างกัน ดังนี้ Valence 0.427 และ Arousal 0.016 ซึ่ง Valence มีค่ามากกว่า 0.05 จึงควรตัดคุณสมบัติเพลงนี้ออก แต่ Arousal ไม่ได้มีค่าที่เกิน 0.05 จึงไม่ควรตัดคุณสมบัติเพลงออก ผลลัพธ์ทั้งสองค่าแสดงให้เห็นความแตกต่างของผลลัพธ์เป็นอย่างมาก จึงนำไปพิจารณารูปแบบข้อมูล

ในรูปที่ 5.6 แสดงให้เห็นว่า Tempo ไม่ได้มีการกระจายตัวของผลลัพธ์ที่ดี เพราะผลลัพธ์ที่ออกมาแสดงว่าการวิเคราะห์ Tempo มีความซ้ำซ้อนของผลลัพธ์เป็นอย่างมาก ซึ่งในความเป็นจริงแล้ว ควรจะแตกต่างเฉพาะเพลง และมีการกระจายตัวของข้อมูลที่หลากหลายและเป็น Linear

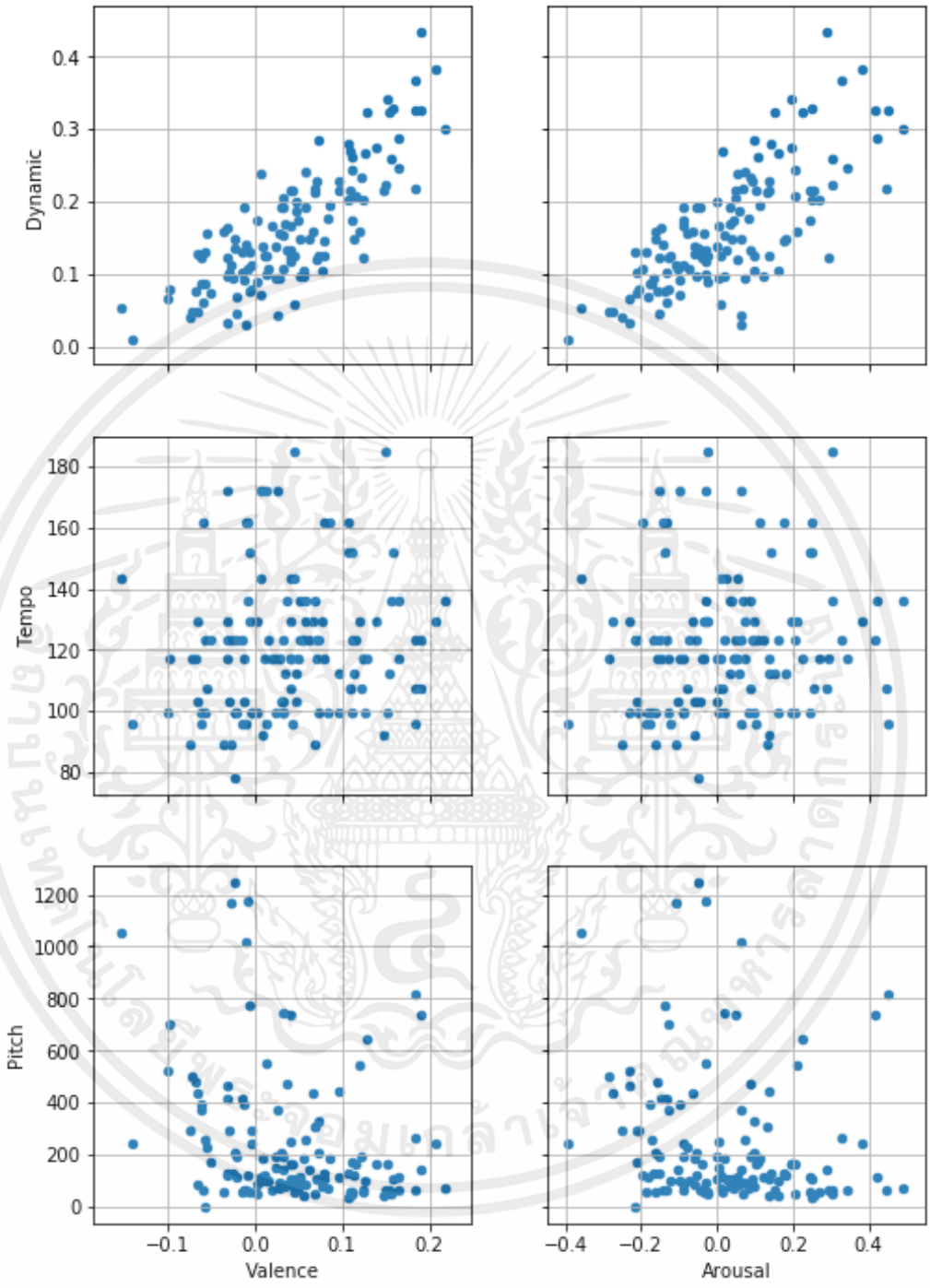
ผู้วิจัยจึงทดลองฟังเพลง และตรวจสอบโดยอ้างอิงผลลัพธ์ Tempo ในงานนี้ พบว่าผลลัพธ์ของหลาย ๆ เพลง ไม่ตรงกับจังหวะที่แท้จริงของเพลง ซึ่ง Tempo ควรจะส่งผลต่ออารมณ์มากที่สุด แต่เมื่อผลลัพธ์ออกมาไม่ตรงกับความเป็นจริง ทำให้ประสิทธิภาพของ Tempo ที่ส่งผลต่ออารมณ์ลดลงไปเป็นอย่างมาก ดังนั้น แม้จะมีค่า P-value ของ Arousal ต่ำกว่า 0.05 แต่ตัดสินใจตัด Tempo ออก เพื่อดูว่าประสิทธิภาพของระบบจะเพิ่มขึ้นหรือไม่

Dynamic ที่มีค่า P-value เท่ากับ 0 เมื่อตรวจสอบรูปแบบข้อมูลพบว่ามีความสัมพันธ์ที่เพิ่มขึ้นแบบ Linear แสดงว่าเมื่อ Dynamic เพิ่ม Valence-Arousal ก็จะเพิ่ม ตรงกับค่า P-value ที่เท่ากับ 0 ซึ่งหมายความว่า คุณสมบัติเพลงนี้มีผลต่ออารมณ์มากที่สุด ไม่ควรตัดออก

รูปที่ 5.7 แสดงความสัมพันธ์ของคุณสมบัติเพลงและ Valence-Arousal ในชุดทดสอบเพลงตะวันตก 125 เพลง ผลลัพธ์แสดงให้เห็นว่ารูปแบบข้อมูลที่ทำนายออกมามีรูปแบบความสัมพันธ์กับคุณสมบัติเพลงคล้ายคลึงกับรูปที่ 5.6



รูปที่ 5.6 ความสัมพันธ์ระหว่าง Valence-Arousal และ Pitch, Dynamic, Tempo ของ MediaEval2013 744 เพลง



รูปที่ 5.7 ความสัมพันธ์ระหว่าง Valence-Arousal และ Pitch, Dynamic, Tempo ของ MediaEval2013 ชุดทดสอบ

ตารางที่ 5.5 สรุปค่า P-value ของ 3 คุณสมบัติเพลง ได้แก่ Pitch, Dynamic และ Tempo เหตุผลที่ตัด Tonality ออกจากการพิจารณา P-value เพราะว่าในงานวิจัยนี้ต้องการใช้ทั้ง 12 คีย์เพลง การที่จะตัดคีย์ใดคีย์หนึ่งออกไป อาจทำให้ไม่ได้ผลสรุปที่แน่ชัด เพราะโมเดลนี้ถูกนำมาใช้ร่วมกันกับทุกเพลงที่คีย์เพลงแตกต่างกัน ดังนั้นจึงมุ่งศึกษาไปที่ค่า P-value ของ Pitch, Dynamic และ Tempo เนื่องจากเป็น 3 คุณสมบัติเพลงที่เลือกใช้เพียง 1 ฟีเจอร์ ดังนั้นการตัดคุณสมบัติเหล่านี้ออก จะทำให้เห็นได้ชัดเจนกว่า Tonality ว่าคุณสมบัติใดส่งผลต่อประสิทธิภาพของโมเดล

ตารางที่ 5.5 P-value

	Pitch	Dynamic	Tempo
Valence	0.437	0	0.427
Arousal	0.599	0	0.016

ผลลัพธ์จากตารางที่ 5.5 แสดงให้เห็นว่า ไม่ควรที่จะตัด Dynamic ในขณะที่ Pitch และ Tempo ควรตัดออก เพื่อดูว่าประสิทธิภาพของโมเดลจะเพิ่มขึ้นหรือไม่ แต่ไม่ควรที่จะตัดพร้อมกันทีเดียว เพราะในขณะที่ตัด 1 ตัว อาจส่งผลให้คุณสมบัติเพลงอื่น ๆ มีประสิทธิภาพขึ้นมาได้ จึงต้องทำการตัดทีละตัว วิธีนี้เรียกว่า Backward elimination

เมื่อทำการ Backward elimination มีโมเดลทั้งหมด 4 แบบ ได้แก่ All, No Pitch, No Tempo, No Pitch+Tempo ดังตารางที่ 5.6 เมื่อนำไปตรวจสอบค่าความถูกต้อง พบว่า No Tempo ให้ผลลัพธ์ที่ดีที่สุด คือ 56% ในขณะที่ All และ No Pitch+Tempo ได้ผลลัพธ์ที่ต่ำที่สุด คือ 54.4% ซึ่งไม่ได้แตกต่างกับผลลัพธ์ที่ดีที่สุดเท่าไหร่ ผลลัพธ์ในส่วนนี้ แสดงในตารางที่ 5.7 โดยในส่วนการคำนวณค่าความถูกต้องเรียงลำดับผลลัพธ์ ดังนี้ Happy, Excited, Sad และ Peaceful

ตารางที่ 5.6 โมเดลทั้งหมด หลังจาก Backward elimination

Models	Features
All model	Pitch, Dynamic, Tempo and Tonality [1-12]
No Pitch	Dynamic, Tempo and Tonality [1-12]
No Tempo	Pitch, Dynamic and Tonality [1-12]
No Pitch+Tempo	Dynamic and Tonality [1-12]

ตารางที่ 5.7 ค่าความถูกต้องของแต่ละโมเดล

	Accuracy calculate	Accuracy
All	(41 + 0 + 19 + 8) / 125	54.4%
No Pitch	(41 + 0 + 20 + 8) / 125	55.2%
No Tempo	(41 + 0 + 21 + 8) / 125	**56%**
No Pitch+Tempo	(40 + 0 + 20 + 8) / 125	54.4%

เมื่ออ้างอิงกับวิธี Backward elimination ที่ควรตัดคุณสมบัติเพลง P-value > 0.05 ออก สรุปได้ว่าค่าสัมประสิทธิ์ที่มาก ซึ่งส่งผลต่ออารมณ์มาก จะมีค่า P-value ที่ต่ำ แสดงว่าไม่ควรตัด Dynamic ออก จากนั้นตรวจสอบค่าความถูกต้อง ของโมเดล ได้ผลลัพธ์ดังตารางที่ 5.7 ซึ่งถือว่ามีประสิทธิภาพค่อนข้างดี แต่ต้องทดสอบหาวิธีที่จะเพิ่มประสิทธิภาพมากกว่านี้ จึงทำการตรวจสอบค่า P-value ทุกคุณสมบัติเพลง

เมื่อตรวจสอบ Confusion matrix ของทั้ง 4 โมเดล ตารางที่ 5.8, 5.10, 5.12 และ 5.14 พบว่าข้อมูลไม่มีการกระจายตัวเท่าที่ควร Happy มีความถูกต้อง 41 เพลง ยกเว้นโมเดล No Pitch+Tempo ที่ได้ผลลัพธ์ 40 เพลง แต่เมื่อนำไปคิดเฉพาะผลลัพธ์ที่ถูกต้องแล้ว(True positive) พบว่าเป็น Happy มากกว่า 50%

ในขณะที่ Sad มีความถูกต้องรองลงมาอยู่ที่ 19-21 เพลง ซึ่งโมเดล No Tempo มีประสิทธิภาพที่สุดก็เพราะว่าสามารถทำนาย Sad ได้ 21 เพลง มากกว่าโมเดลอื่น ๆ และ Peaceful ทุกโมเดลแสดงผลลัพธ์ที่ถูกต้องเท่ากับ 8 เพลงทั้งหมด

แต่ทุกโมเดลมีจุดด้อยประสิทธิภาพเหมือนกัน คือ ไม่สามารถทำนาย Excited ได้ถูกต้องแม้แต่เพลงเดียว ส่งผลให้โมเดลทั้งหมดไม่มีการกระจายตัวของข้อมูลที่ดี ดังนั้น ค่าความถูกต้อง จึงไม่ใช่ตัวเลือกที่มีความน่าเชื่อถือในการวัดประสิทธิภาพของระบบ ต้องทำการตรวจสอบค่าความแม่นยำ ค่าความระลึก และ F-measure

ตารางที่ 5.8 Confusion matrix MLR โมเดล All

	Happy	Excited	Sad	Peaceful
Happy	41	0	8	5
Excited	13	0	2	0
Sad	5	1	19	4
Peaceful	9	0	10	8

ตารางที่ 5.9 ค่าความแม่นยำ ค่าความระลึก และ F-measure MLR โมเดล All

	Precision	Recall	F-measure
Happy	60%	76%	67%
Excited	0%	0%	0%
Sad	49%	66%	56%
Peaceful	47%	30%	36%
Average	**48%**	**54%**	**50%**

ตารางที่ 5.10 Confusion matrix MLR โมเดล No Pitch

	Happy	Excited	Sad	Peaceful
Happy	41	0	7	6
Excited	13	0	2	0
Sad	6	0	20	3
Peaceful	9	0	10	8

ตารางที่ 5.11 ค่าความแม่นยำ ค่าความระลึก และ F-measure MLR โมเดล No Pitch

	Precision	Recall	F-measure
Happy	59%	76%	67%
Excited	0%	0%	0%
Sad	51%	69%	59%
Peaceful	47%	30%	36%
Average	**48%**	**55%**	**50%**

ตารางที่ 5.12 Confusion matrix MLR โมเดล No Tempo

	Happy	Excited	Sad	Peaceful
Happy	41	0	8	5
Excited	12	0	1	2
Sad	5	1	21	2
Peaceful	9	0	10	8

ตารางที่ 5.13 ค่าความแม่นยำ ค่าความระลึก และ F-measure MLR โมเดล No Tempo

	Precision	Recall	F-measure
Happy	61%	76%	68%
Excited	0%	0%	0%
Sad	53%	72%	61%
Peaceful	47%	30%	36%
Average	**49%**	**56%**	**51%**

ตารางที่ 5.14 Confusion matrix MLR โมเดล No pitch tempo

	Happy	Excited	Sad	Peaceful
Happy	40	1	5	8
Excited	12	0	1	2
Sad	6	0	20	3
Peaceful	9	0	10	8

ตารางที่ 5.15 ค่าความแม่นยำ ค่าความระลึก และ F-measure MLR โมเดล No Pitch Tempo

	Precision	Recall	F-measure
Happy	60%	74%	66%
Excited	0%	0%	0%
Sad	56%	69%	62%
Peaceful	38%	30%	33%
Average	**47%**	**54%**	**50%**

ตารางที่ 5.9, 5.11, 5.13 และ 5.15 คำนวณค่าความแม่นยำ ค่าความระลึก และ F-measure ของทั้ง 4 โมเดล โดยแยกเป็นอารมณ์ และนำมาสรุปเป็นค่าเฉลี่ยของแต่ละโมเดลในตารางที่ 5.16 ซึ่งโมเดล No Tempo ยังคงเป็นโมเดลที่มีประสิทธิภาพที่สุดที่ F-measure 51% ในขณะที่อีก 3 โมเดลมีค่า F-measure 50% ซึ่งไม่ได้แตกต่างกับ No Tempo มากนัก แต่ผลลัพธ์ได้ยืนยันว่า No Tempo เป็นโมเดลที่มีประสิทธิภาพมากที่สุด ทั้งในด้านค่าความถูกต้อง และ F-measure รวมถึงมีค่าเฉลี่ยของค่าความแม่นยำ ค่าความระลึก มากที่สุดเช่นกัน

ตารางที่ 5.16 สรุปค่าความถูกต้อง และ F-measure ของ MLR

	All	No Pitch	No Tempo	No Pitch+Tempo
Accuracy	54.4%	55.2%	**56%**	54.4%
F-measure	50%	50%	**51%**	50%

โมเดล All และ No Pitch+Tempo แสดงผลลัพธ์ค่าความถูกต้อง 54.4% และ F-measure เท่ากันทั้ง 2 โมเดล ซึ่งเป็นผลลัพธ์ที่มีประสิทธิภาพต่ำที่สุดใน 4 โมเดล แต่เมื่อเปรียบเทียบตารางที่ 5.9 และ ตารางที่ 5.15 พบว่าในแต่ละอารมณ์ไม่ได้มีค่าต่าง ๆ เท่ากันทุกอารมณ์ ซึ่งในส่วนนี้ยากจะสรุปผลลัพธ์ของแต่ละอารมณ์ได้ แต่เมื่อนำมาคิดค่าเฉลี่ยโมเดล พบว่าค่าความแม่นยำไม่ได้เท่ากัน โมเดล No Pitch+Tempo ได้ผลลัพธ์ค่าความแม่นยำ 47% ซึ่งต่ำที่สุด เพียงแต่เมื่อนำมาหาค่าเฉลี่ย F-measure ผลลัพธ์ของทั้ง 2 โมเดลจึงมีความน่าเชื่อถือไม่ต่างกัน แสดงในตารางที่ 5.17

โมเดล All แม้จะมีประสิทธิภาพด้านค่าความแม่นยำ ดีกว่า แต่ในภาพรวมก็ยังมีประสิทธิภาพที่ต่ำที่สุด ซึ่งแสดงให้เห็นว่าการใช้คุณสมบัติเพลงทุกแบบที่งานวิจัยนี้ได้เลือกมา อาจไม่เหมาะสมกับการนำมาทำนายอารมณ์เพลง และทั้ง 4 โมเดล ยังด้อยประสิทธิภาพในการทำนายอารมณ์ Excited เป็นอย่างมาก ทั้ง ๆ ที่เพลงตะวันตก และเพลงไทย มีเพลงอยู่ในอารมณ์นี้เป็นอย่างมาก โดยเฉพาะเพลงไทยที่มีอารมณ์ Excited เป็นอันดับสองรองจาก Happy

ตารางที่ 5.17 เปรียบเทียบประสิทธิภาพโมเดล All และ No Pitch+Tempo

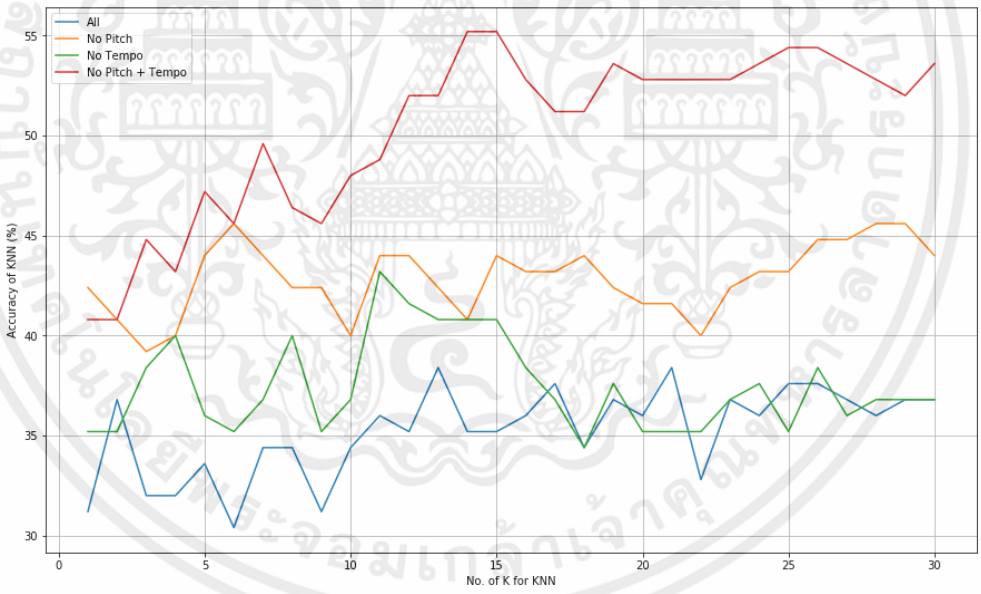
	Precision	Recall	F-measure	Accuracy
All	48%	54%	50%	54.4%
No Pitch+Tempo	47%	54%	50%	54.4%

ผลลัพธ์ของทั้ง 4 โมเดล แสดงให้เห็นว่าโมเดล No Tempo มีประสิทธิภาพสูงสุด แต่เมื่อต้องการโมเดลอื่น ๆ ไปทำนายเพลงไทยเพื่อเปรียบเทียบประสิทธิภาพ ผู้วิจัยจึงนำทั้ง 4 โมเดลไปขั้นตอนต่อไป เพราะประสิทธิภาพโดยรวมของทั้ง 4 โมเดล มีความใกล้เคียงกันมาก ดังนั้นเมื่อนำไปเข้าสู่กระบวนการ KNN และทำนายเพลงไทย โมเดลที่ประสิทธิภาพต่ำที่สุดในเพลงตะวันตก อาจกลายเป็นโมเดลที่มีประสิทธิภาพสูงสุดทั้งในส่วน KNN และ เพลงไทย

5.2.2 K-Nearest neighbors (KNN)

นำทั้ง 4 โมเดล เข้าสู่กระบวนการ KNN โดยกำหนดค่า K ที่ 1-30 เพื่อตรวจสอบประสิทธิภาพว่าโมเดลต่าง ๆ ให้ประสิทธิภาพในแต่ละค่า K เท่าไหร่ และโดยรวมของโมเดลมีประสิทธิภาพเป็นเช่นใด โดยนำโมเดลที่มีประสิทธิภาพที่สุดของ KNN ไปทำนายเพลงไทยร่วมกับ 4 โมเดลจาก MLR

รูปที่ 5.8 แสดงประสิทธิภาพของ 4 โมเดล เมื่อกำหนดค่า K ที่ 1-30 โดย All คือ เส้นสีฟ้า, No Pitch คือ เส้นสีส้ม, No Tempo คือ เส้นสีเขียว และ No Pitch+Tempo คือ เส้นสีแดง



รูปที่ 5.8 ค่าความถูกต้อง ของ KNN เมื่อกำหนดค่า K ที่ 1 ถึง 30

รูปที่ 5.8 แสดงให้เห็นว่าโมเดล No Pitch+Tempo ที่ค่า K 14 และ 15 มีประสิทธิภาพสูงสุดที่ 55.2% และมีประสิทธิภาพเฉลี่ยโมเดลอยู่ที่ 50.32% ในขณะที่โมเดล No tempo ที่มีประสิทธิภาพที่สุด

ใน MLR กลับกลายเป็นอันดับ 3 ใน KNN ซึ่งแสดงผลลัพธ์สูงสุดแค่ 43.2% และมีประสิทธิภาพเฉลี่ยเพียง 37.44% แสดงในตารางที่ 5.18

โมเดล All ในวิธี KNN แสดงประสิทธิภาพโมเดลต่ำที่สุดอย่างเห็นได้ชัด มีประสิทธิภาพสูงสุดที่ 38.4% และประสิทธิภาพเฉลี่ยแค่ 35.27% ซึ่งต่ำมาก ๆ เมื่ออ้างอิงถึงผลลัพธ์ในวิธี MLR ซึ่งโมเดล All เป็นโมเดลที่มีประสิทธิภาพต่ำสุดเช่นกัน ดังนั้น เป็นไปได้ว่าคุณสมบัติเพลงที่เลือกใช้ในงานวิจัยนี้ ไม่เหมาะสมที่จะนำมาทำนายอารมณ์เพลง ต้องมีการตัดคุณสมบัติเพลงบางอย่างออก ซึ่งจากผลลัพธ์ของ MLR และ KNN แสดงให้เห็นว่า ผลลัพธ์ที่ดีที่สุดมาจากโมเดลที่ไม่มีคุณสมบัติเพลง Tempo ร่วมอยู่ด้วย ได้แก่ No Tempo จาก MLR และ No Pitch+Tempo จาก KNN

ตารางที่ 5.18 ประสิทธิภาพวิธี KNN

	ค่า K ที่ประสิทธิภาพสูงสุด	ประสิทธิภาพเฉลี่ย	ประสิทธิภาพสูงสุด
All	13 และ 21	35.27%	38.4%
No Pitch	6, 28 และ 29	42.85%	45.6%
No Tempo	11	37.44%	43.2%
No Pitch+Tempo	14 และ 15	**50.32%**	**55.2%**

ค่าความถูกต้อง ที่ดีที่สุดของ KNN เท่ากับ 55.2% จากโมเดล No Pitch+Tempo ที่ K = 14 และ 15 ซึ่งเท่ากับโมเดล No Pitch ใน MLR ที่มีประสิทธิภาพเป็นอันดับ 2 ดังนั้นผู้วิจัยจึงนำโมเดลนี้ไปตรวจสอบ Confusion matrix ในตารางที่ 5.19 และ 5.21 เพื่อดูการกระจายตัวของข้อมูลที่ถูกต้องว่าค่าความถูกต้อง มีความน่าเชื่อถือหรือไม่

ตารางที่ 5.19 Confusion matrix โมเดล K = 14

	Happy	Excited	Sad	Peaceful
Happy	46	1	7	0
Excited	13	0	2	0
Sad	11	0	18	0
Peaceful	12	0	10	5

ตารางที่ 5.20 ค่าความแม่นยำ ค่าความระลึก และ F-measure โมเดล K = 14

	Precision	Recall	F-measure
Happy	57%	85%	68%
Excited	0%	0%	0%
Sad	47%	62%	54%
Peaceful	100%	19%	32%
Average	**57%**	**55%**	**49%**

ตารางที่ 5.21 Confusion matrix โมเดล K = 15

	Happy	Excited	Sad	Peaceful
Happy	46	1	7	0
Excited	13	0	2	0
Sad	11	0	18	0
Peaceful	11	0	11	5

ตารางที่ 5.22 ค่าความแม่นยำ ค่าความระลึก และ F-measure โมเดล K = 15

	Precision	Recall	F-measure
Happy	56%	85%	68%
Excited	0%	0%	0%
Sad	49%	62%	55%
Peaceful	100%	19%	32%
Average	**57%**	**55%**	**49%**

Confusion matrix ตารางที่ 5.19 และ 5.21 แสดงข้อมูลที่ถูกต้อง ซึ่งยังคงไม่มีการกระจายตัว ไม่มีโมเดลใดที่ทำนาย Excited ได้ถูกต้อง และข้อมูลที่ถูกต้องส่วนมาก คือ Happy เช่นเดิม แต่มีจำนวนที่มากกว่า MLR

Peaceful และ Sad มีข้อมูลที่ถูกต้องน้อยกว่าทุกโมเดลของ MLR ดังนั้นค่าความถูกต้อง ไม่ใช่ค่าที่มีความน่าเชื่อถือเพียงพอ ผู้วิจัยจึงต้องทำการคำนวณค่าความแม่นยำ ค่าความระลึก และ F-measure แสดงให้เห็นในตารางที่ 5.20 และ 5.22

ผลลัพธ์แสดงให้เห็นว่า KNN ให้ค่าเฉลี่ยของค่าความแม่นยำ 57% ซึ่งเยอะกว่าโมเดล MLR ที่มีประสิทธิภาพสูงสุดอย่าง No Tempo มาก เป็นเพราะว่าโมเดล KNN มีประสิทธิภาพในการทำนายอารมณ์ Peaceful โดยไม่เสียคุณภาพเชิงปริมาณ หมายความว่าโมเดลให้ผลลัพธ์เป็น Peaceful และผลลัพธ์ไม่ไปตรงกับอารมณ์อื่น ๆ แสดงให้เห็นในคอลัมน์ของ Peaceful ตารางที่ 5.19 และ 5.21

ในขณะที่ MLR มีประสิทธิภาพด้านค่าความแม่นยำ ต่ำกว่าเพราะโมเดลสามารถทำนายได้ Peaceful ถูกต้อง แต่ผลลัพธ์แสดงออกมาตรงกับอารมณ์อื่น ๆ แสดงให้เห็นในคอลัมน์ของอารมณ์ Peaceful ของ MLR ตัวอย่างในตารางที่ 5.2 โมเดล No Tempo ของ MLR สามารถทำนายอารมณ์ Peaceful ได้ถูกต้อง 8 เพลง แต่ไปตรงกับอารมณ์ Happy 5 เพลง, Excited 2 เพลง และ Sad 2 เพลง

ผลลัพธ์ค่าเฉลี่ยค่าความระลึก จาก KNN เท่ากับ 55% ทั้ง 2 โมเดล ซึ่งตรงกับโมเดล No Pitch ของ MLR ที่มีค่าความระลึก และค่าความถูกต้อง เท่ากัน ค่าความระลึก ของอารมณ์ Peaceful ทั้ง 2 โมเดลเท่ากับ 19% ซึ่งต่ำมาก แตกต่างกับค่าความแม่นยำในอารมณ์นี้ที่สูง และเมื่อเปรียบเทียบกับ MLR ค่าความระลึกของ Peaceful แสดงผลลัพธ์ที่ต่ำกว่ามาก เป็นเพราะว่าข้อมูลจริงให้ผลลัพธ์เชิงคุณภาพที่ไม่ดี ข้อมูลจริงของอารมณ์ Peaceful ในตารางที่ 5.18 และ 5.20 แสดงผลลัพธ์ตรงกับอารมณ์อื่น ๆ ถึง 22 เพลง

ค่าเฉลี่ย F-measure แสดงผลลัพธ์ของทั้ง 2 โมเดล เท่ากับ 49% ซึ่งต่ำกว่าโมเดลจาก MLR ทั้งหมด ทั้ง ๆ ที่ค่าความแม่นยำ ค่าความระลึก มีค่าเฉลี่ยค่อนข้างมากกว่า โดยเฉพาะอารมณ์ Peaceful เป็นเพราะว่าเมื่อนำมาคำนวณหา F-measure จากสมการที่ (4.15) ผลลัพธ์ F-measure ของอารมณ์ Peaceful เท่ากับ 32% ซึ่งต่ำกว่าทุกโมเดล MLR และ อารมณ์ Sad ก็ให้ค่า F-measure ของอารมณ์ที่ต่ำกว่าทุกโมเดลใน MLR เช่นกัน

ดังนั้นเมื่อนำมาคำนวณค่าเฉลี่ย F-measure ด้วยสมการที่ (4.16) แสดงให้เห็นว่าประสิทธิภาพโดยรวมของโมเดล KNN ต่ำกว่า MLR ทุกโมเดล แม้จะมีค่าความถูกต้องที่สูงก็ตาม ซึ่งสนับสนุนกับทฤษฎีความไม่น่าเชื่อถือของค่าความถูกต้อง เมื่อข้อมูลที่ถูกต้องไม่เกิดการกระจายตัว

แสดงวิธีการคำนวณ F-measure อารมณ์ Peaceful ของ KNN โดยต้องคำนวณค่าความแม่นยำ ค่าความระลึก ดังสมการที่ (5.1) และ (5.2) ตามลำดับ

$$\text{Precision}_{peaceful} = \frac{5}{5+0} \text{ x } 100\%$$

$$= 100\% \tag{5.1}$$

$$\text{Recall}_{peaceful} = \frac{5}{5+22} \text{ x } 100\%$$

$$= 19\% \tag{5.2}$$

$$\text{F-measure}_{peaceful} = 2 \text{ x } \frac{100\% \text{ x } 19\%}{100\% + 19\%}$$

$$= 32\% \tag{5.3}$$

หาค่าเฉลี่ยของ F-measure ด้วยสมการที่ (4.16) แสดงวิธีการคำนวณและผลลัพธ์ในสมการที่ (5.4)

$$\text{F-measure}_{average} = \left(\frac{0.68(54) \text{ x } 0(15) \text{ x } 0.32(27) \text{ x } 0.54(29)}{125} \right) \text{x } 100\%$$

$$= 49\% \tag{5.4}$$

ดังนั้น แม้ค่า F-measure ของ KNN จะต่ำกว่า MLR ทุกโมเดล แต่ก็ไม่ได้แตกต่างกันมาก ผู้วิจัยจึงนำ 2 โมเดลจาก KNN บวกกับ 4 โมเดลจาก MLR รวมเป็น 6 โมเดลที่นำไปทำนายอารมณ์เพลงไทย แสดงให้เห็นในตารางที่ 5.23 และสรุปประสิทธิภาพของ 6 โมเดลจากการทำนายเพลงตะวันตก ในตารางที่ 5.24

ตารางที่ 5.23 โมเดล 6 แบบ ในการทำนายอารมณ์เพลงไทย

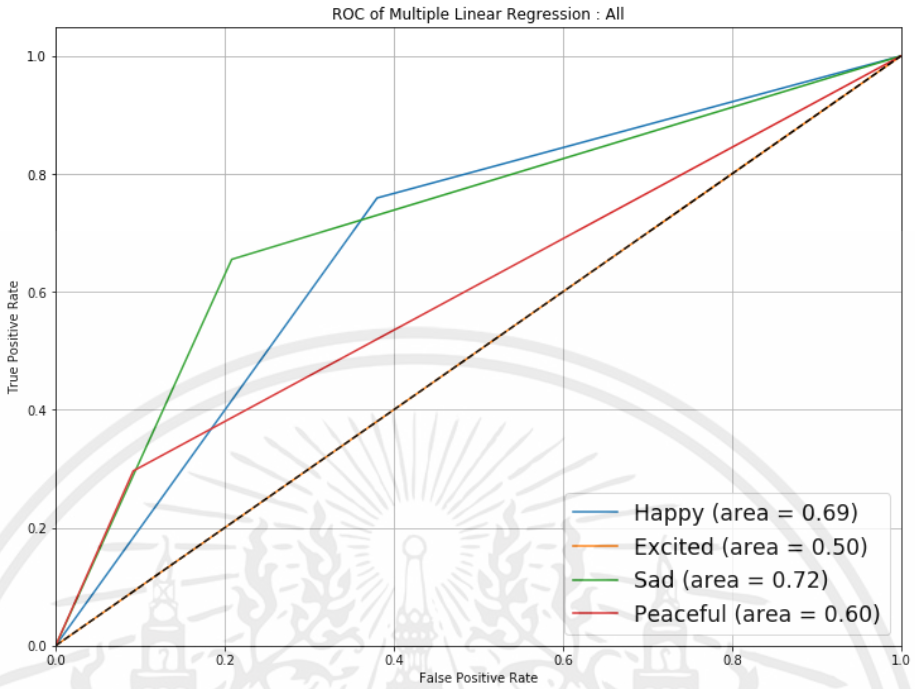
Models	Features
All model	MLR with Pitch, Dynamic, Tempo and Tonality [1-12]
No Pitch	MLR with Dynamic, Tempo and Tonality [1-12]
No Tempo	MLR with Pitch, Dynamic and Tonality [1-12]
No Pitch+Tempo	MLR with Dynamic and Tonality [1-12]
K_14	KNN at K=14 from MLR No Pitch+Tempo
K_15	KNN at K=15 from MLR No Pitch+Tempo

ตารางที่ 5.24 ประสิทธิภาพของ 6 โมเดลที่นำไปทำนายอารมณ์เพลงไทย

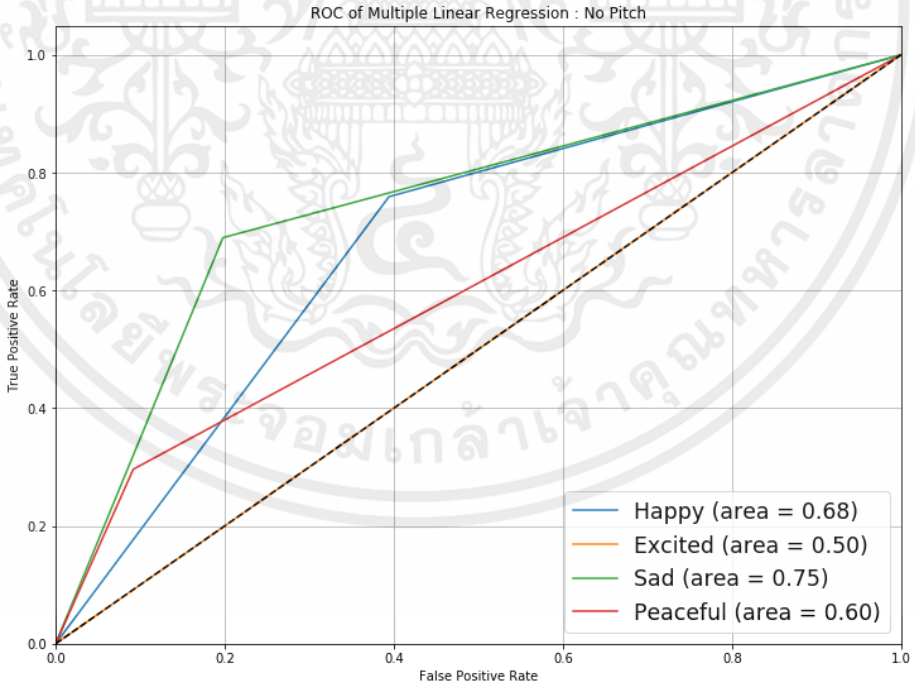
	Accuracy	Precision	Recall	F-measure
All	54.4%	48%	54%	50%
No Pitch	55.2%	48%	55%	50%
No Tempo	56.0%	49%	56%	**51%**
No Pitch+Tempo	54.4%	47%	54%	50%
K_14	55.2%	57%	55%	49%
K_15	55.2%	57%	55%	49%

รูปที่ 5.9-5.14 แสดงกราฟ ROC ของ 6 โมเดล และสรุปผลลัพธ์ทั้งหมดในตารางที่ 5.25 โดยผลลัพธ์แสดงไปในทิศทางเดียวกันว่า Sad คือ อารมณ์ที่มีค่า AUC มากที่สุด ซึ่งมากกว่า 70% ในทุกโมเดลทั้ง ๆ ที่ Happy น่าจะเป็นอารมณ์ที่มีค่า AUC มากสุด เพราะตารางที่ 5.8, 5.10, 5.12, 5.14, 5.19 และ 5.21 แสดงผลลัพธ์ว่า Happy มีค่า True positive และ ค่าความแม่นยำมากสุด แต่เมื่อนำข้อมูลมาพิจารณาพบว่าอารมณ์ Sad มีการสูญเสียประสิทธิภาพการทำนายข้อมูลในอัตราส่วนต่อ True positive ที่น้อยกว่า Happy

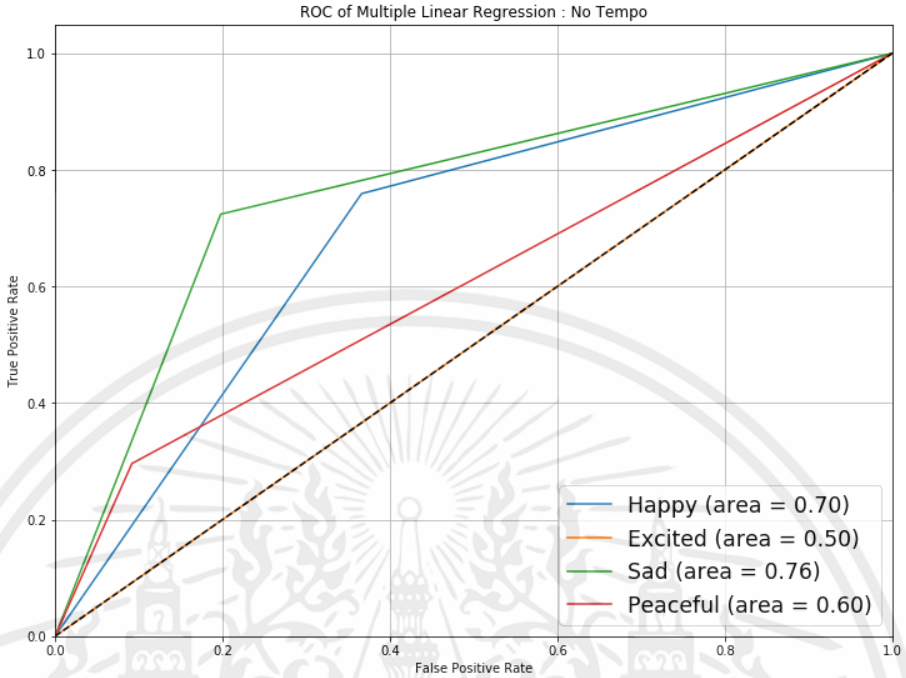
No Tempo คือ โมเดลเดียวที่มีค่า ROC อารมณ์ Happy 70% แต่ก็ยังต่ำกว่าค่า ROC ต่ำสุดของอารมณ์ Sad ที่ 71% จากโมเดล K_14 และ K_15 ในขณะที่ Excited ซึ่งแสดงผลใน Confusion Matrix ว่าไม่สามารถมีโมเดลใดทำนายได้ถูกต้อง มีค่า ROC 50% เท่ากันทุกโมเดล ซึ่งผ่านมาตรฐานค่า ROC ที่ไม่ควรต่ำกว่า 50%



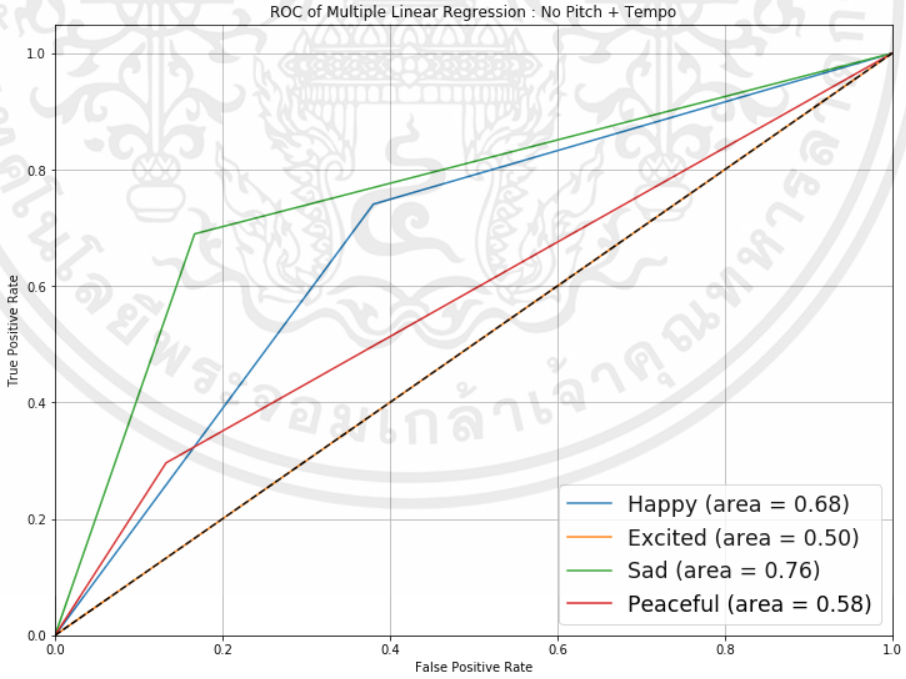
รูปที่ 5.9 กราฟ ROC โมเดล All



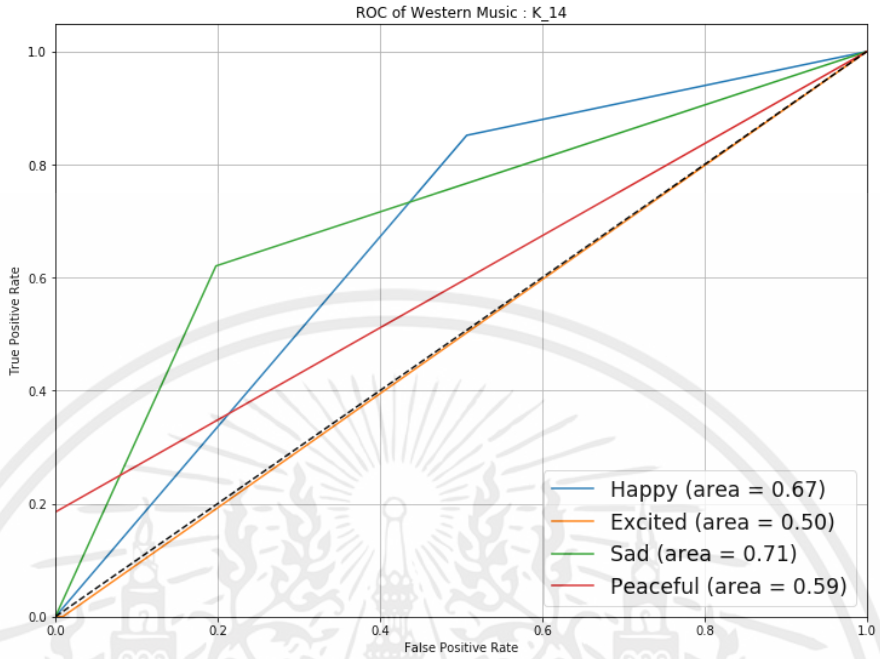
รูปที่ 5.10 กราฟ ROC โมเดล No Pitch



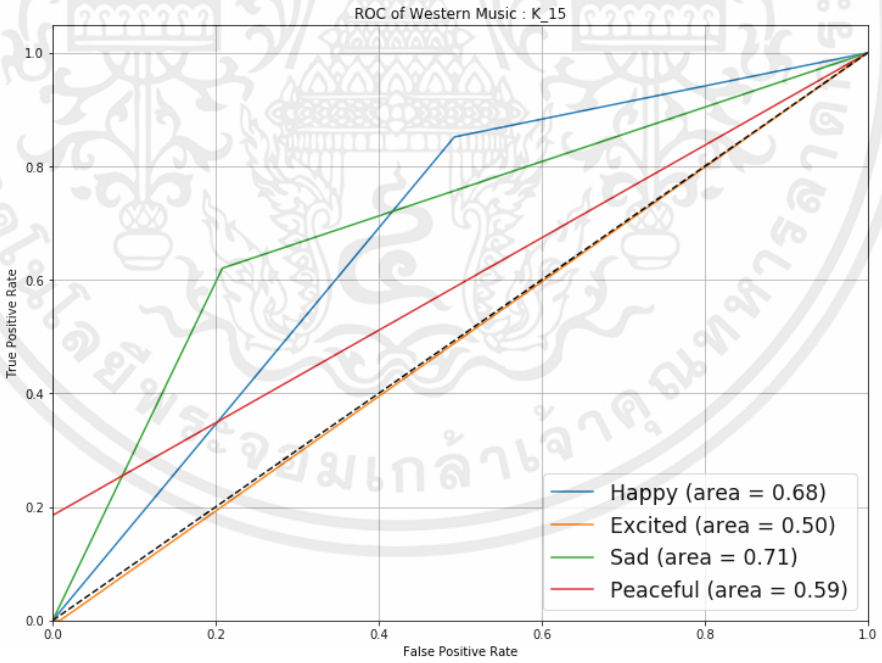
รูปที่ 5.11 กราฟ ROC โมเดล No Tempo



รูปที่ 5.12 กราฟ ROC โมเดล No Pitch+Tempo



รูปที่ 5.13 กราฟ ROC โมเดล K_14



รูปที่ 5.14 กราฟ ROC โมเดล K_15

ตารางที่ 5.25 ประสิทธิภาพของกราฟ ROC เพลงตะวันตกจาก 6 โมเดล

	Happy	Excited	Sad	Peaceful
All	69%	50%	72%	60%
No Pitch	68%	50%	75%	60%
No Tempo	**70%**	**50%**	**76%**	**60%**
No Pitch+Tempo	68%	50%	76%	58%
K_14	67%	50%	71%	59%
K_15	68%	50%	71%	59%

5.2.3 ทำนายอารมณ์เพลงไทย

นำทั้ง 6 โมเดลมาทำนายอารมณ์เพลงไทย 125 เพลง โดยอ้างอิงค่าอารมณ์จาก Spotify API เพื่อดูผลลัพธ์ว่าโมเดลที่มีประสิทธิภาพที่สุดของเพลงตะวันตกและเพลงไทย เหมือนกันหรือแตกต่างกัน ซึ่งเพลงตะวันตก แสดงผลลัพธ์ว่าโมเดล No Pitch+Tempo จาก MLR คือ โมเดลที่มีประสิทธิภาพที่สุด และโมเดล All แสดงประสิทธิภาพที่ต่ำทั้ง MLR และ KNN

เริ่มต้นจากการตรวจสอบค่าความถูกต้อง เหมือนเพลงตะวันตก ผลลัพธ์ในตารางที่ 5.26 แสดงผลลัพธ์ว่าโมเดล K_14, K_15 และ All คือ โมเดลที่มีค่าความถูกต้อง สูงสุดของเพลงไทย โดยเฉพาะโมเดล All ที่มีประสิทธิภาพต่ำที่สุดของเพลงตะวันตก แต่กลับมีค่าความถูกต้อง มากที่สุดในเพลงไทย ในขณะที่ No Tempo จาก MLR แสดงผลลัพธ์ที่ต่ำ ทั้ง ๆ ที่เป็นโมเดลที่มีประสิทธิภาพที่สุดในเพลงตะวันตก ผลลัพธ์เช่นนี้ ทำให้ผลลัพธ์ของเพลงตะวันตกและเพลงไทยแตกต่างกันมาก สิ่งที่น่าสนใจ คือ โมเดล All สามารถทำนายอารมณ์ Excited ได้ถูกต้อง จากที่ไม่เคยมีโมเดลใด ทำนายถูกในเพลงตะวันตกและไทย

ตารางที่ 5.26 ค่าความถูกต้องของเพลงไทย

	Accuracy calculate	Accuracy
All	(64 + 1 + 2 + 0) / 125	53.6%
No Pitch	(64 + 0 + 2 + 0) / 125	52.8%
No Tempo	(63 + 0 + 2 + 0) / 125	52%
No Pitch+Tempo	(63 + 0 + 3 + 0) / 125	52.8%
K_14	(62 + 0 + 5 + 0) / 125	53.6%
K_15	(62 + 0 + 5 + 0) / 125	53.6%

แม้ค่าความถูกต้องจะดูแตกต่างกันมาก แต่จากผลลัพธ์เพลงตะวันตก ทำให้ไม่สามารถเชื่อถือค่าความถูกต้องได้ หากข้อมูลที่ถูกต้องไม่มีการกระจายตัว ดังนั้นจึงต้องทำการตรวจสอบข้อมูล โดยตารางที่ 5.27, 5.29, 5.31, 5.33, 5.35 และ 5.37 แสดง Confusion matrix และ ตารางที่ 5.28, 5.30, 5.32, 5.34, 5.36 และ 5.38 แสดงค่าความแม่นยำ ค่าความระลึก และ F-measure ของเพลงไทย ทั้ง 6 โมเดล ตามลำดับตารางที่ 5.26

ตารางที่ 5.27 Confusion matrix โมเดล All เพลงไทย

	Happy	Excited	Sad	Peaceful
Happy	64	0	0	0
Excited	39	1	2	1
Sad	9	0	2	4
Peaceful	3	0	0	0

ตารางที่ 5.28 ค่าความแม่นยำ ค่าความระลึก และ F-measure โมเดล All เพลงไทย

	Precision	Recall	F-measure
Happy	56%	100%	72%
Excited	100%	2%	5%
Sad	50%	13%	21%
Peaceful	0%	0%	0%
Average	**69%**	**54%**	**41%**

ตารางที่ 5.29 Confusion matrix โมเดล No Pitch เพลงไทย

	Happy	Excited	Sad	Peaceful
Happy	64	0	0	0
Excited	40	0	3	0
Sad	9	0	2	4
Peaceful	3	0	0	0

ตารางที่ 5.30 ค่าความแม่นยำ ค่าความระลึก และ F-measure โมเดล No Pitch เพลงไทย

	Precision	Recall	F-measure
Happy	55%	100%	71%
Excited	0%	0%	0%
Sad	40%	13%	20%
Peaceful	0%	0%	0%
Average	**33%**	**53%**	**39%**

ตารางที่ 5.31 Confusion matrix โมเดล No Tempo เพลงไทย

	Happy	Excited	Sad	Peaceful
Happy	63	0	0	1
Excited	39	0	3	1
Sad	8	0	2	5
Peaceful	9	0	0	0

ตารางที่ 5.32 ค่าความแม่นยำ ค่าความระลึก และ F-measure โมเดล No Tempo เพลงไทย

	Precision	Recall	F-measure
Happy	56%	98%	71%
Excited	0%	0%	0%
Sad	40%	13%	20%
Peaceful	0%	0%	0%
Average	**33%**	**52%**	**39%**

ตารางที่ 5.33 Confusion matrix โมเดล No pitch tempo เพลงไทย

	Happy	Excited	Sad	Peaceful
Happy	63	0	0	1
Excited	39	0	3	1
Sad	8	0	3	4
Peaceful	3	0	0	0

ตารางที่ 5.34 ค่าความแม่นยำ ค่าความระลึก และ F-measure โมเดล No Pitch Tempo เพลงไทย

	Precision	Recall	F-measure
Happy	56%	98%	71%
Excited	0%	0%	0%
Sad	50%	20%	29%
Peaceful	0%	0%	0%
Average	**35%**	**53%**	**40%**

ตารางที่ 5.35 Confusion matrix โมเดล K_14 เพลงไทย

	Happy	Excited	Sad	Peaceful
Happy	62	0	1	1
Excited	40	0	3	0
Sad	10	0	5	0
Peaceful	3	0	0	0

ตารางที่ 5.36 ค่าความแม่นยำ ค่าความระลึก และ F-measure โมเดล K_14 เพลงไทย

	Precision	Recall	F-measure
Happy	54%	97%	69%
Excited	0%	0%	0%
Sad	56%	33%	42%
Peaceful	0%	0%	0%
Average	**34%**	**54%**	**40%**

ตารางที่ 5.37 Confusion Matrix โมเดล K_15 เพลงไทย

	Happy	Excited	Sad	Peaceful
Happy	62	0	2	0
Excited	40	0	3	0
Sad	10	0	5	0
Peaceful	3	0	0	0

ตารางที่ 5.38 ค่าความแม่นยำ ค่าความระลึก และ F-measure โมเดล K_15 เพลงไทย

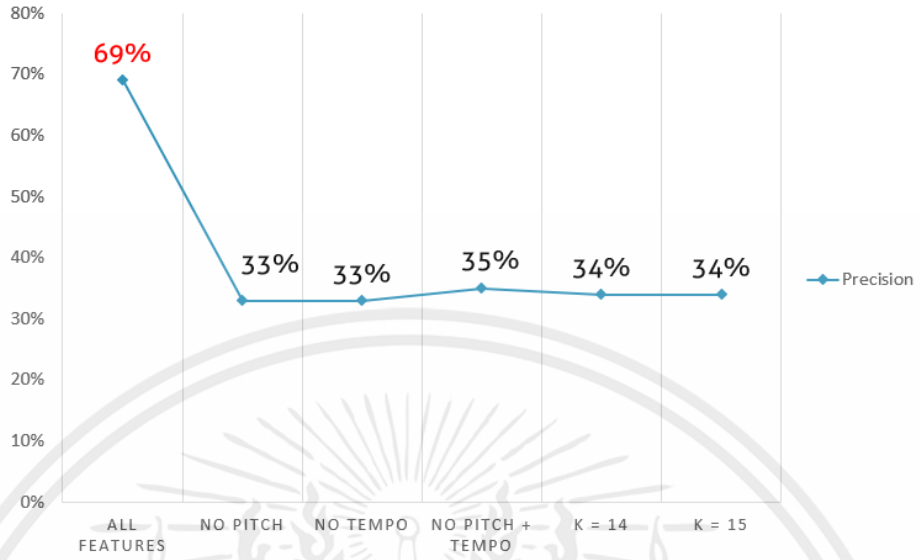
	Precision	Recall	F-measure
Happy	54%	97%	69%
Excited	0%	0%	0%
Sad	50%	33%	40%
Peaceful	0%	0%	0%
Average	**34%**	**54%**	**40%**

ผลลัพธ์ของค่าเฉลี่ย F-measure และ ค่าความถูกต้อง ทั้ง 6 โมเดล สรุปไว้ในตารางที่ 5.39 ซึ่งโมเดล All ยังคงเป็นโมเดลที่มีประสิทธิภาพที่สุดในเพลงไทย ที่ F-measure 41% ในขณะที่อีก 5 โมเดลมีค่า F-measure 39-40% ซึ่งอาจจะไม่ได้แตกต่างกับโมเดล All มากนัก แต่ผลลัพธ์ได้ยืนยันว่าโมเดล All เป็นโมเดลที่มีประสิทธิภาพมากที่สุด ทั้งในด้านค่าความถูกต้อง และ F-measure

ตารางที่ 5.39 ค่าเฉลี่ย F-measure และค่าความถูกต้องเพลงไทยและเพลงตะวันตก ทั้ง 6 โมเดล

	All	No Pitch	No Tempo	No Pitch+Tempo	K_14	K_15
Accuracy	**53.6%**	52.8%	52%	52.8%	53.6%	53.6%
F-measure	**41%**	39%	39%	40%	40%	40%

เหตุผลที่โมเดล All เป็นโมเดลที่มีประสิทธิภาพที่สุดในเพลงไทย เพราะว่าสามารถทำนาย Excited ได้ถูกต้อง แม้จะแค่ 1 เพลง ซึ่งไม่มีโมเดลใดในวิทยานิพนธ์ฉบับนี้สามารถทำนายได้ถูกต้อง ทั้งในเพลงตะวันตกและเพลงไทย ด้วยเหตุนี้จึงทำให้ค่าเฉลี่ยของค่าความแม่นยำของโมเดล All สูงถึง 69% ในขณะที่โมเดลอื่น ๆ มีค่าเฉลี่ยของค่าความแม่นยำ ไม่เกิน 35% ดังรูปที่ 5.15 ซึ่งถือว่าเป็นผลลัพธ์ที่ดี เนื่องจากเพลงไทยมีอารมณ์ Excited เป็นอย่างมาก และโมเดล All ยังเป็น 1 ใน 2 โมเดลร่วมกับ No Pitch ที่สามารถทำนายอารมณ์ Happy ของเพลงไทยได้ถูกต้องทุกเพลง ดังนั้นค่าเฉลี่ย F-measure จึงสูงที่สุดใน 6 โมเดล สรุปประสิทธิภาพโมเดลของเพลงไทยในตารางที่ 5.39



รูปที่ 5.15 ค่าความแม่นยำ ของแต่ละโมเดล จากการทำนายอารมณ์เพลงไทย

ตารางที่ 5.40 ประสิทธิภาพของ 6 โมเดล จากการทำนายอารมณ์เพลงไทย

	Accuracy	Precision	Recall	F-measure
All	**53.6%**	**69%**	**54%**	**41%**
No Pitch	52.8%	33%	53%	39%
No Tempo	52.0%	33%	52%	39%
No Pitch+Tempo	52.8%	35%	53%	40%
K_14	53.6%	34%	54%	40%
K_15	53.6%	34%	54%	40%

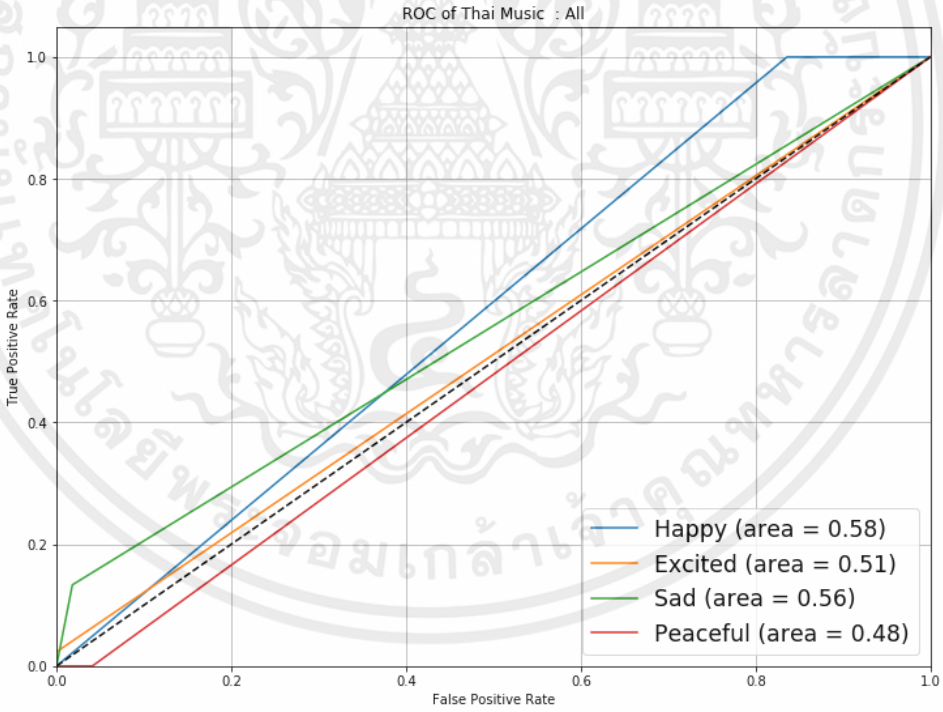
ในขณะที่โมเดล No Tempo ที่ประสิทธิภาพสูงสุดในเพลงตะวันตก เมื่อนำมาทำนายเพลงไทยให้ผลลัพธ์ที่ต่ำที่สุด คู่กับโมเดล No Pitch เมื่อเปรียบเทียบโมเดลที่มีประสิทธิภาพสูงที่สุด และต่ำที่สุดของเพลงตะวันตกและเพลงไทย พบว่ามีผลลัพธ์แตกต่างกันโดยสิ้นเชิง

ตรวจสอบกราฟ ROC เพลงไทย 6 โมเดล ดังรูปที่ 5.16 ถึง 5.21 และสรุปประสิทธิภาพในตารางที่ 5.41 ผลลัพธ์แสดงให้เห็นว่าโมเดล All ที่สามารถทำนายอารมณ์ Excited ได้ถูกต้อง มีค่า ROC อารมณ์ Excited 51% ซึ่งมากกว่าโมเดลอื่น ๆ เล็กน้อย โดยโมเดล K_14 และ K_15 ให้ประสิทธิภาพอารมณ์ Sad 65% และ 64% ตามลำดับ ซึ่งเป็นค่าที่สูงโดดเด่นกว่าอีก 4 โมเดลจาก MLR

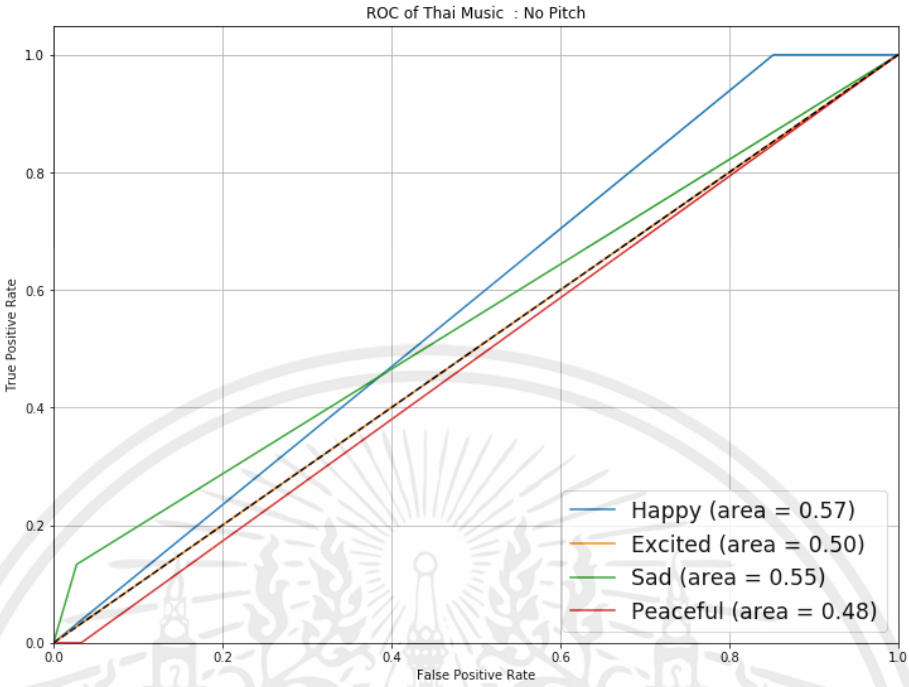
นอกจากนี้ 4 โมเดลจาก MLR ยังให้ประสิทธิภาพอารมณ์ Peaceful ต่ำกว่า 50% ซึ่งถือว่าประสิทธิภาพต่ำกว่าเกณฑ์ และยังไม่เคยเกิดขึ้นในเพลงตะวันตก

ตารางที่ 5.41 ประสิทธิภาพของกราฟ ROC เพลงไทยจาก 6 โมเดล

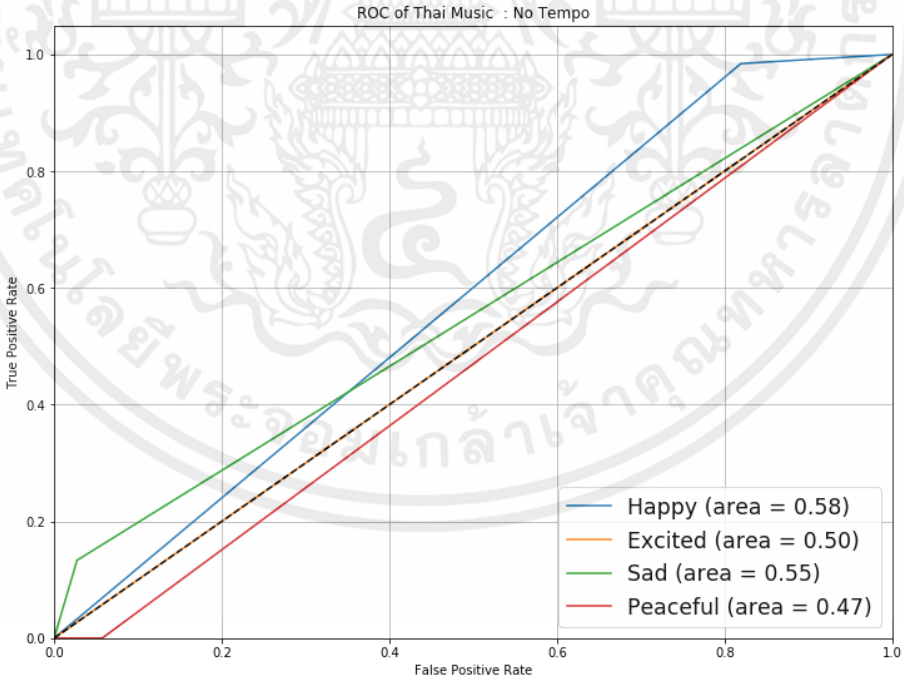
	Happy	Excited	Sad	Peaceful
All	58%	51%	56%	48%
No Pitch	57%	50%	55%	48%
No Tempo	58%	50%	55%	47%
No Pitch+Tempo	58%	50%	59%	48%
K_14	55%	50%	65%	50%
K_15	55%	50%	64%	50%



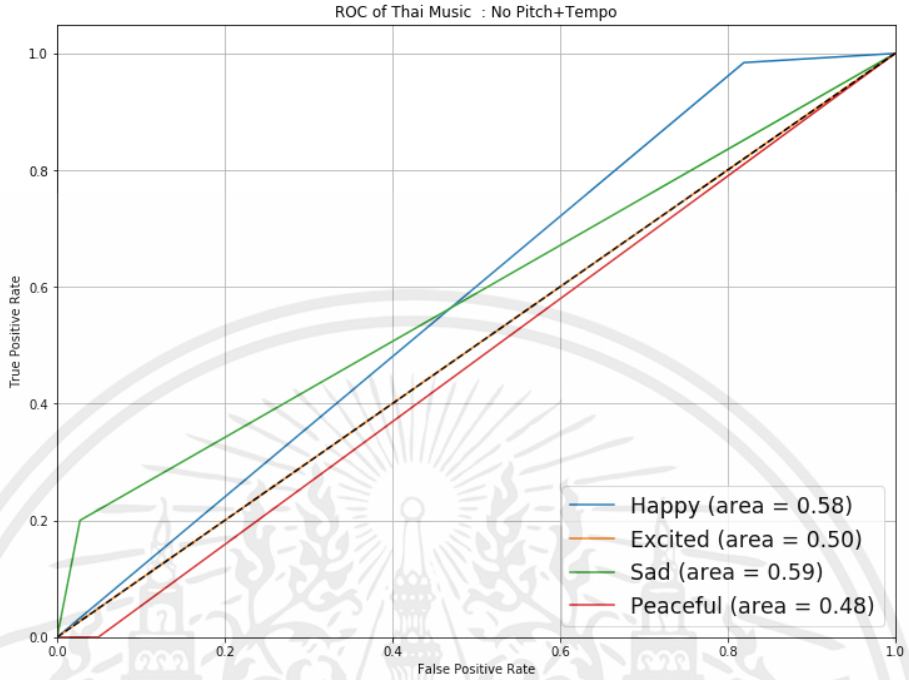
รูปที่ 5.16 กราฟ ROC โมเดล All เพลงไทย



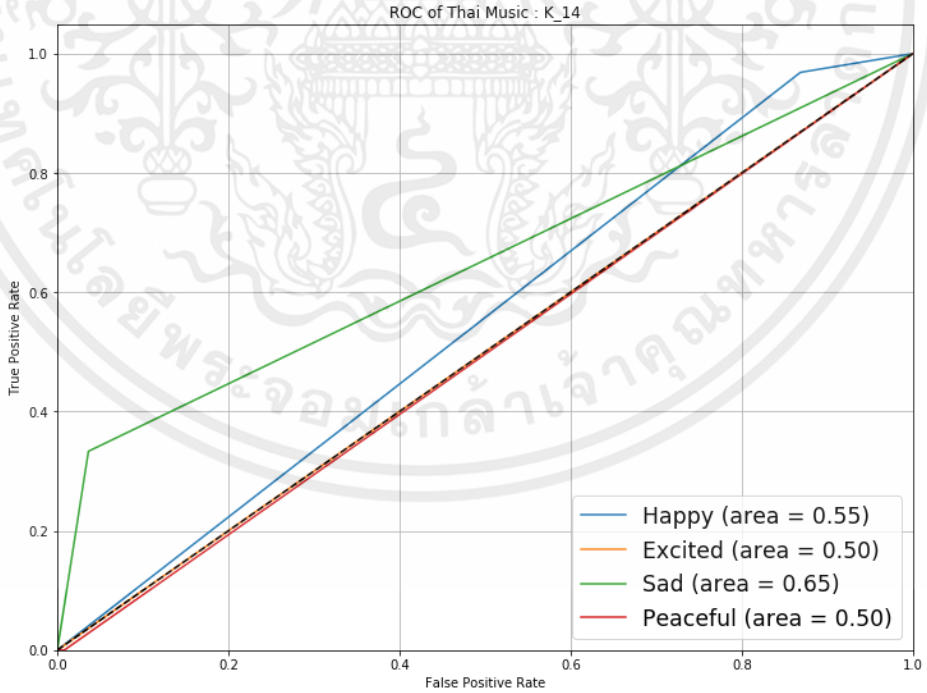
รูปที่ 5.17 กราฟ ROC โมเดล No Pitch เพลงไทย



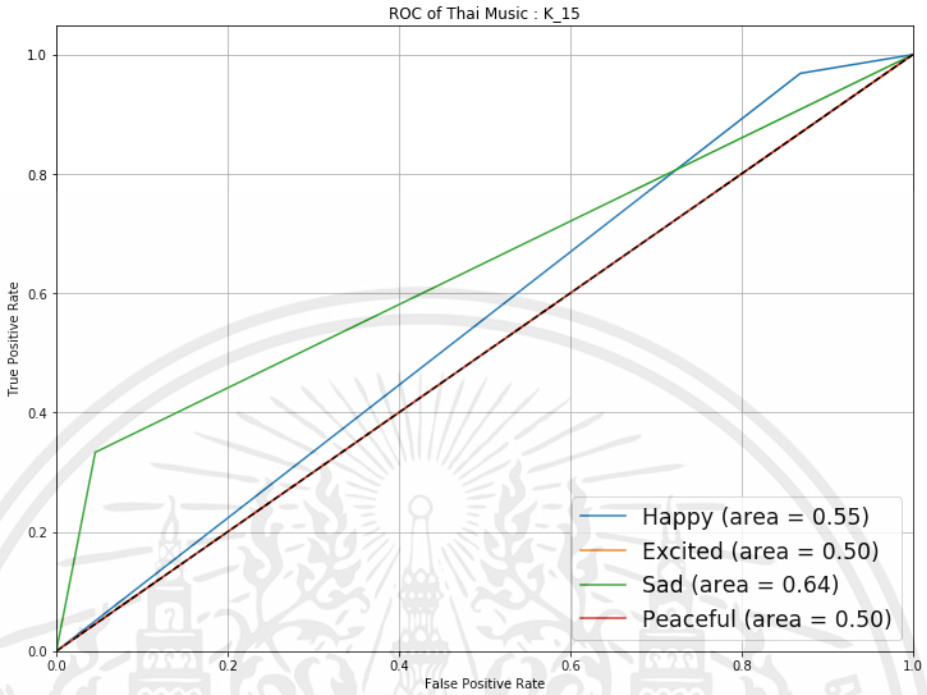
รูปที่ 5.18 กราฟ ROC โมเดล No Tempo เพลงไทย



รูปที่ 5.19 กราฟ ROC โมเดล No Pitch + Tempo เพลงไทย



รูปที่ 5.20 กราฟ ROC โมเดล K_14 เพลงไทย



รูปที่ 5.21 กราฟ ROC โมเดล K_15 เพลงไทย

5.2.4 เปรียบเทียบประสิทธิภาพโมเดล Linear regression เพลงไทย

ผลลัพธ์ในหัวข้อ 5.2.1–5.2.3 แสดงให้เห็นว่าโมเดลที่มีประสิทธิภาพของเพลงตะวันตก และเพลงไทยมีความแตกต่างกันโดยสิ้นเชิง ดังนั้นผู้วิจัยจึงตั้งสมมติฐานอีกหนึ่งอย่าง คือ Multiple linear regression (MLR) และ Support vector regression (SVR) แบบ Linear วิธีใดให้ประสิทธิภาพการทำนายอารมณ์เพลงไทยมากกว่า เพราะในหลาย ๆ งานวิจัยด้านนี้ แสดงให้เห็นว่า Support vector ซึ่งเป็นวิธีระดับสูง คือ วิธีการเรียนรู้ด้วยเครื่องที่ให้ประสิทธิภาพดีที่สุดในเพลงตะวันตก ดังนั้นผู้วิจัยจึงนำ MLR ซึ่งเป็นวิธีพื้นฐานมาเปรียบเทียบประสิทธิภาพการทำนายอารมณ์เพลงไทยโดยเฉพาะ

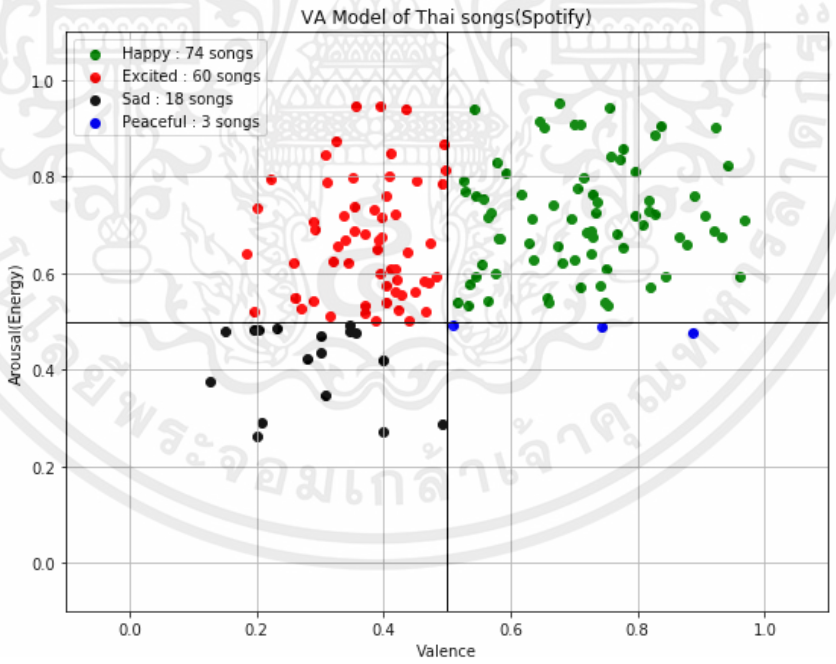
Support vector สามารถจำแนกข้อมูลได้ทั้ง Classification และ Regression ในส่วนนี้เลือกใช้การจำแนกข้อมูลแบบ Regression และเลือกประเภทของโมเดลเป็นแบบ Linear

คุณสมบัติเพลงที่ใช้ในหัวข้อนี้ ผู้วิจัยได้ทำการตัด Tempo ออก เพราะผลลัพธ์ในหัวข้อก่อนหน้านี้แสดงให้เห็นว่าการตัด Tempo ทำให้ได้ผลลัพธ์ที่ดีกว่า และได้ทำการเพิ่มคุณสมบัติเพลง Timbre ทั้งหมด 13 ฟีเจอร์ คุณสมบัติเพลงที่ใช้ในหัวข้อนี้แสดงให้เห็นในตารางที่ 5.42

ตารางที่ 5.42 คุณสมบัติเพลงที่นำไปสร้างโมเดล MLR และ SVR

Features	No. of features	Methods
Pitch	1	Discrete Fourier transform
Dynamic	1	Root mean square energy
Tonality	12	Chromagram
Timbre	13	Mel-frequency cepstral coefficients (MFCCs)

ชุดข้อมูลในส่วนนี้ คือ เพลงไทย 155 เพลง ไม่มีเพลงตะวันตก คุณสมบัติของไฟล์เพลงมีดังต่อไปนี้ ความยาว 45 วินาที 16 บิต อัตราการสุ่มตัวอย่าง 44.1 กิโลเฮิรตซ์ และระบบเสียง Mono แบ่งข้อมูลเป็นอัตรา 80:20 คือ ชุดทดลอง 124 เพลง และ ชุดทดสอบ 31 เพลง โดยค่าอารมณ์ Valence-Arousal(Energy) เท่ากับ 0 ถึง 1 อ้างอิงข้อมูลจาก Spotify API ซึ่งแสดงจำนวนอารมณ์ ดังนี้ Happy : 74, Excited 60, Sad 18 และ Peaceful 3 รูปที่ 5.22 แสดงให้เห็นอารมณ์ของเพลงไทย 155 เพลง บนแผนภาพ Valence-Arousal



รูปที่ 5.22 อารมณ์เพลงไทย 155 เพลง บนแผนภาพ Valence-Arousal

Confusion matrix ของ MLR ในตารางที่ 5.43 และ SVR ตารางที่ 5.45 แสดงข้อมูลที่ถูกต้อง (True positive) ที่แตกต่างกันที่อารมณ์ Excited 1 เพลง เท่านั้น และอารมณ์ Excited มีความถูกต้องที่เพิ่มขึ้นจากหัวข้อก่อนที่ใช้โมเดลเพลงตะวันตกทำนายอารมณ์เพลงไทยอย่างเห็นได้ชัด แม้จะมีรูปแบบการทดลองที่แตกต่างกัน แต่ผลลัพธ์ก็แสดงแนวโน้มว่า การทำนายอารมณ์เพลง ควรใช้เพลงประจำชาตินั้น ๆ ในการสร้างโมเดลและทดลอง ในขณะที่ Peaceful ไม่มีข้อมูลที่ถูกต้อง เป็นเพราะว่ามีข้อมูลให้เรียนรู้เพียงแค่ 3 เพลง และเป็นชุดทดสอบแค่ 1 เพลง ทำให้มีโอกาสต่ำที่จะทำนายถูก ค่าความแม่นยำ ค่าความระลึก และ F-measure ของ MLR แสดงในตารางที่ 5.44 และ SVR ตารางที่ 5.46

ตารางที่ 5.43 Confusion matrix โมเดล MLR เพลงไทย

	Happy	Excited	Sad	Peaceful
Happy	**12**	4	0	0
Excited	4	**6**	0	0
Sad	0	2	**1**	1
Peaceful	1	0	0	**0**

ตารางที่ 5.44 ค่าความแม่นยำ ค่าความระลึก และ F-measure โมเดล MLR เพลงไทย

	Precision	Recall	F-measure
Happy	71%	75%	73%
Excited	50%	60%	55%
Sad	100%	25%	40%
Peaceful	0%	0%	0%
Average	**65%**	**61%**	**60%**

ตารางที่ 5.45 Confusion matrix โมเดล MLR เพลงไทย

	Happy	Excited	Sad	Peaceful
Happy	**12**	3	1	0
Excited	5	**5**	0	0
Sad	0	2	**1**	1
Peaceful	1	0	0	**0**

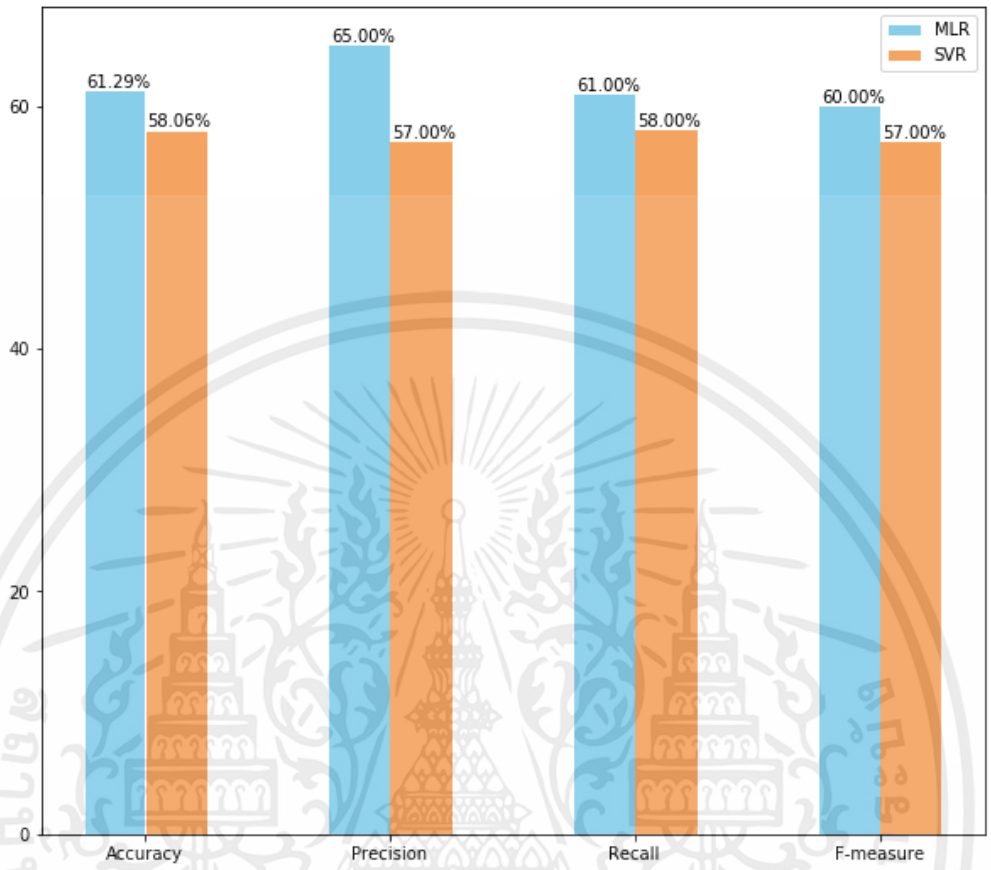
ตารางที่ 5.46 ค่าความแม่นยำ ค่าความระลึก และ F-measure โมเดล SVR เพลงไทย

	Precision	Recall	F-measure
Happy	67%	75%	71%
Excited	50%	50%	50%
Sad	50%	25%	33%
Peaceful	0%	0%	0%
Average	**57%**	**58%**	**57%**

ค่าความแม่นยำ อารมณ์ Sad ของ MLR มีค่าสูงถึง 100% แม้ SVR จะทำนาย Sad ได้ถูกต้อง 1 เพลงเท่ากัน แต่ที่ค่าความแม่นยำอารมณ์ Sad ของ SVR มีค่าแค่ 50% เพราะมีข้อมูล 1 เพลงจากการทำนายอารมณ์ Sad ตรงกับอารมณ์ Happy ทำให้สูญเสียประสิทธิภาพในด้าน Precision และนอกจากนี้ค่าความแม่นยำ ของอารมณ์อื่น ๆ นอกจาก Sad มีแค่เพียง Happy เท่านั้นที่แตกต่างที่ MLR 71% และ SVR 67% ซึ่งแตกต่างกันไม่มาก

ค่าความระลึก อารมณ์ Happy และ Sad ของ MLR และ SVR มีค่าเท่ากัน ในขณะที่อารมณ์ Excited ให้ผลลัพธ์แตกต่างกันที่ MLR 60% และ SVR 50% เป็นเพราะว่า SVR มีการสูญเสียประสิทธิภาพ 1 เพลง ในขณะที่ MLR นำความแตกต่างนั้นมาเป็นค่าข้อมูลที่ถูกต้อง

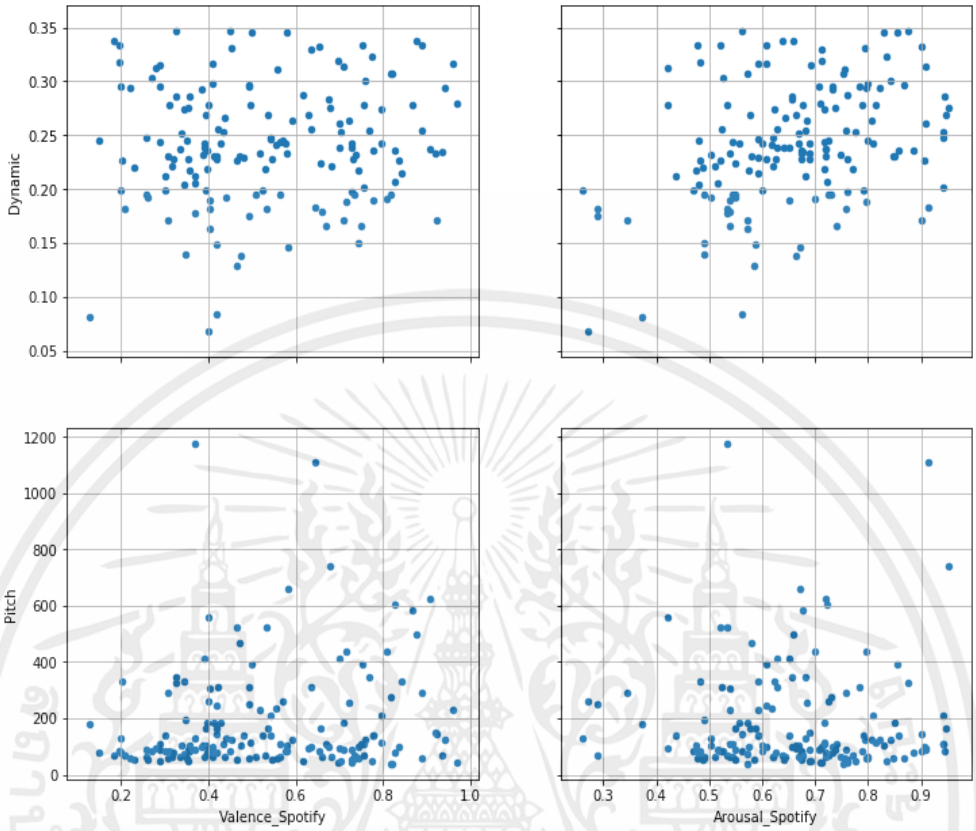
ค่า F-measure แต่ละอารมณ์ของ MLR มีประสิทธิภาพกว่า SVR ทุกอารมณ์ เมื่อนำมาคำนวณค่าเฉลี่ยประสิทธิภาพโมเดลในแต่ละด้าน MVR จึงแสดงประสิทธิภาพได้ดีกว่าในทุกด้าน สรุปให้เห็นในรูปที่ 5.23



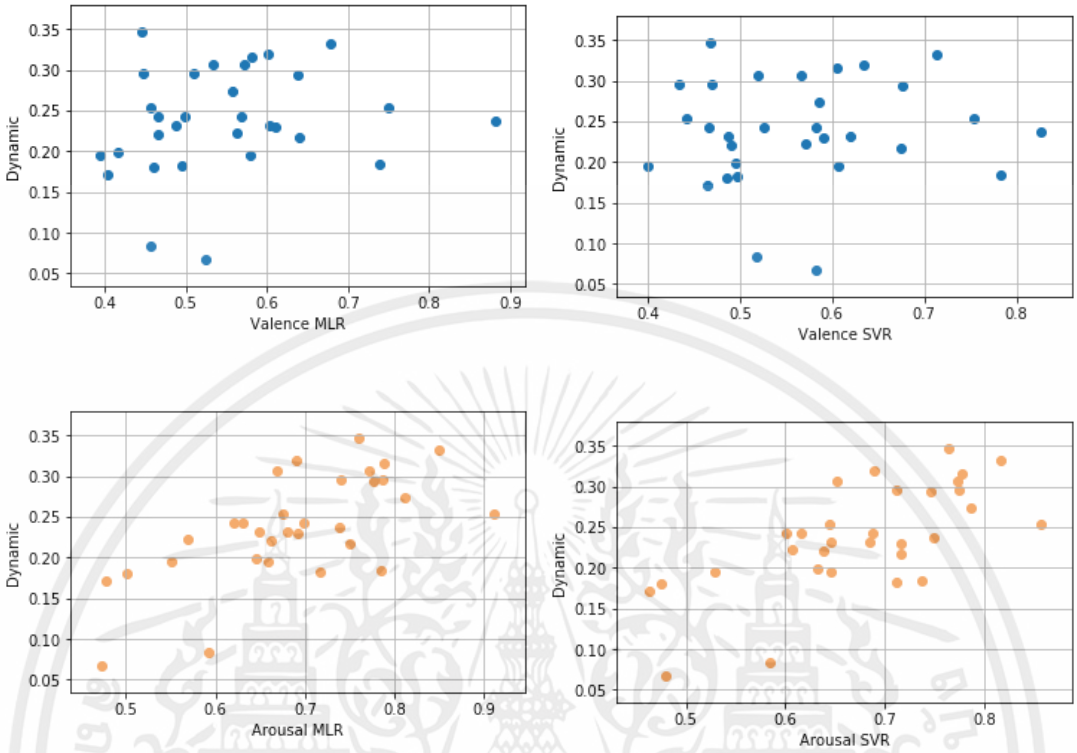
รูปที่ 5.23 เปรียบเทียบค่าความถูกต้อง และค่าเฉลี่ย ค่าความแม่นยำ ค่าความระลึก และ F-measure ของเพลงไทย โมเดล MLR และ SVR

รูปที่ 5.24 แสดงความสัมพันธ์ของคุณสมบัติเพลงและ Valence-Arousal ในชุดทดสอบเพลงไทย 155 เพลง ผลลัพธ์แสดงให้เห็นว่ารูปแบบข้อมูลที่ทำนายออกมามีรูปแบบความสัมพันธ์กับคุณสมบัติเพลงคล้ายคลึงชุดทดสอบเพลงเพลงไทยกับรูปที่ 5.25

Dynamic ยังคงมีรูปแบบข้อมูลสัมพันธ์กับ Valence-Arousal ในเชิง Linear ซึ่งผลลัพธ์เหมือนกับโมเดลเพลงตะวันตก แม้ว่าผลลัพธ์ในการเปรียบเทียบโมเดลเพลงตะวันออกและเพลงไทยจะให้ผลลัพธ์ที่แตกต่างกันโดยสิ้นเชิง รวมไปถึงการเปรียบเทียบโมเดล Regression ของเพลงไทยโดยเฉพาะจะแตกต่างกับการที่มีเพลงตะวันตกร่วมด้วย แต่จุดที่มีร่วมกัน คือ Dynamic เป็นคุณสมบัติเพลงที่ส่งผลต่อประสิทธิภาพโมเดลและอารมณ์เพลงในงานวิจัยนี้ที่สุด



รูปที่ 5.24 ความสัมพันธ์ระหว่าง Valence-Arousal และ Pitch, Dynamic, Tempo ของเพลงไทย 155 เพลง



รูปที่ 5.25 ความสัมพันธ์ระหว่าง Valence-Arousal และ Pitch, Dynamic, Tempo ของเพลงไทยชุดทดสอบ 31 เพลง

บทที่ 6
สรุปผลการวิจัยและข้อเสนอแนะ

6.1 สรุปผลการวิจัย

วัตถุประสงค์ของวิทยานิพนธ์ฉบับนี้ คือ เลือกโมเดลที่มีประสิทธิภาพที่สุดในโมเดลเพลงตะวันตกมาทำนายอารมณ์เพลงไทย และเปรียบเทียบโมเดลที่มีประสิทธิภาพที่สุดระหว่างเพลงตะวันตกและเพลงไทย มีความคล้ายคลึงหรือแตกต่างกัน เริ่มต้นจากการสร้างโมเดล MLR ของเพลงตะวันตก ผลลัพธ์ที่ได้ คือ Valence-Arousal มี R-square 0.11 และ 0.52 ตามลำดับ และ MSE 0.058 และ 0.042 ตามลำดับ ซึ่งในด้านอารมณ์ ไม่ได้เป็นผลลัพธ์ที่ต่ำเกินไป เพราะอารมณ์นั้นยากที่จะนิยามได้ชัดเจน

เมื่อนำค่า Valence-Arousal มาหาคำนิยามอารมณ์บนแผนภาพ Valence-Arousal และตรวจสอบค่าความถูกต้อง ผลลัพธ์ที่ได้ คือ 54.4% และผู้วิจัยต้องการเพิ่มประสิทธิภาพของโมเดล จึงเลือกวิธี Backward elimination ซึ่งจะตัดคุณสมบัติเพลงที่ P-value > 0.05 ออก เมื่อตรวจสอบ P-value ของ Pitch, Dynamic และ Tempo ผลลัพธ์แสดงว่า Dynamic เป็นคุณสมบัติเพลงเพียงอย่างเดียวที่ไม่ควรตัดออก

หลังจากทำการตัด Pitch และ Tempo ทำให้มีโมเดลทั้งหมด 4 แบบ คือ All, No Pitch, No Tempo และ No Pitch+Tempo ซึ่งโมเดล No Tempo มีค่าความถูกต้อง สูงสุดที่ 56% แต่เมื่อตรวจสอบ Confusion matrix ของทุกโมเดลพบว่าข้อมูลที่ถูกต้องไม่มีการกระจายตัวที่ดี ทำให้ค่าความถูกต้อง ไม่น่าเชื่อถือ ควรเลือก F-measure ในการวัดประสิทธิภาพ โดยจะมีการคำนวณ ค่าความแม่นยำ ค่าความระลึก ร่วมด้วย

โมเดล No Tempo มีประสิทธิภาพ F-measure สูงสุดที่ 51% จากผลลัพธ์นี้ทำให้ No Tempo เป็นโมเดลที่มีประสิทธิภาพที่สุดทั้งในด้านค่าความถูกต้อง และ F-measure ในขณะที่อีก 3 โมเดลมีค่า F-measure 50% ซึ่งไม่แตกต่างกับ No Tempo ผู้วิจัยจึงนำทั้ง 4 โมเดล เข้าสู่กระบวนการ KNN ที่ K เท่ากับ 1-30

ค่าความถูกต้อง สูงสุดของ KNN เท่ากับ 55.2% โดยโมเดล No Pitch+Tempo ที่ K เท่ากับ 14 และ 15 ในขณะที่ค่า F-measure 49% ทั้งสองโมเดล แม้จะมีค่าความถูกต้อง ที่สูง แต่ค่า F-measure กลับต่ำกว่าทุกโมเดลจาก MLR ผู้วิจัยเลือก 2 โมเดลนี้ ร่วมกับ 4 โมเดลจาก MLR เพื่อตรวจสอบอารมณ์เพลงไทย รวมเป็น 6 โมเดล ดังนี้ All, No Pitch, No Tempo, no Pitch+Tempo, K_14 และ K_15 ซึ่ง No Tempo คือ โมเดลที่มีประสิทธิภาพที่สุดของเพลงตะวันตก และ All คือ โมเดลที่มีประสิทธิภาพต่ำที่สุด

ผลลัพธ์ของเพลงไทย แสดงว่า All คือ โมเดลที่มีประสิทธิภาพที่สุด โดยค่าความถูกต้อง 53.6% และ F-measure 41% อีกทั้งยังสามารถทำนาย Excited ได้ถูกต้อง ซึ่งไม่มีโมเดลใดสามารถทำได้ ในขณะที่ No Tempo เป็นโมเดลที่มีประสิทธิภาพต่ำที่สุด โดยค่าความถูกต้อง 52% และ F-measure 39% ผลลัพธ์เช่นนี้ทำให้โมเดลที่มีประสิทธิภาพที่สุดของเพลงตะวันตกและเพลงไทยมีความแตกต่างกันโดยสิ้นเชิง

ในอีกส่วนหนึ่ง ผู้วิจัยเปรียบเทียบประสิทธิภาพเพลงไทยโดยเฉพาะ ระหว่างโมเดล MLR และ SVR แบบ Linear ซึ่ง SVR ปรากฏผลลัพธ์ที่ดีกว่าในหลาย ๆ งานวิจัยด้านอารมณ์เพลงตะวันตก โดยเลือกที่จะตัด Tempo ออก และเพิ่ม Timbre ผลลัพธ์ที่ได้ คือ MLR มีประสิทธิภาพดีกว่า SVR ทุกด้าน โดย ค่าความถูกต้อง 61.29%, ค่าความแม่นยำ 65%, ค่าความระลึก 61% และ F-measure 60%

ผลลัพธ์ทั้งหมดจากงานวิจัยนี้แสดงให้เห็นว่า เพลงตะวันตกและเพลงไทยมีความแตกต่างกันอย่างมาก ทั้งที่ไม่มีด้านภาษาเข้ามาเกี่ยวข้อง อาจเป็นไปได้ว่ามีคุณลักษณะทางเสียงที่แตกต่างกัน แม้โดยทั่วไปจะมีแนวดนตรีบนพื้นฐานเดียวกันก็ตาม ดังนั้น หากต้องการที่จะสร้างโมเดลทำนายอารมณ์เพลงประเทศใด ควรที่จะสร้างโมเดลเฉพาะเพลงประเทศนั้น ๆ เพื่อให้เกิดผลลัพธ์ที่มีประสิทธิภาพสูงสุด

6.2 ข้อเสนอแนะ

วิทยานิพนธ์ฉบับนี้ วิจัยจากชุดข้อมูลเพลงตะวันตก MediaEval2013 และ ชุดข้อมูลเพลงไทยส่วนตัว + Spotify API ทำให้ผลลัพธ์มาจากการจำลองไม่ได้มาจากสถานการณ์จริง ในอนาคตจึงคิดจะสร้างโมเดลที่ทดลองกับคนไทยจริง ๆ เพื่อให้การทำนายอารมณ์เพลงไทยเกิดประสิทธิภาพสูงสุด และนำข้อมูลมาสร้างชุดข้อมูลเพลงไทยโดยเฉพาะให้นักวิจัยได้ศึกษา รวมไปถึงหาวิธีปรับปรุงประสิทธิภาพการวิเคราะห์ Tempo ซึ่งควรจะเป็นคุณสมบัติเพลงที่ส่งผลต่ออารมณ์เพลงได้ชัดเจนที่สุด

บรรณานุกรม

[1] J. A. Russell, “A circumplex model of affect,” **Journal of Personality and Social Psychology**, vol. 39, no. 6, 1980, pp. 1161-1178.

[2] R. E. Thayer, **The biopsychology of mood and arousal,** Oxford University, New York, lnc. 1989.

[3] Y. E. Kim, Erik M. Schmidt and R. Migneco, “Music emotion recognition: A state of the art review,” **Proc. of Int. Society for Music Information Retrieval Conference,** 2010, pp. 255-266.

[4] M. G. Viraj and K. L. Jayaratne, “Melody analysis for prediction of the emotions conveyed by Sinhala songs,” **Proc. of Int. Conference on Information and Automation for Sustainability,** 2016, pp.1-6.

[5] S. H. Deshmukh and S. G. Bhirud, “North Indian classical music’s singer identification by timbre recognition using MIR toolbox,” **Int. Journal of Computer Applications,** Vol. 91, No. 4, 2014, pp. 1- 5.

[6] https://developer.spotify.com/documentation/web-api

[7] B. Mcfee, C. Raffel and D. Liang, “Librosa : Audio and music signal and analysis in Python,” **Proc. of the Python in Science Conferences,** 2015, pp. 18-25.

[8] K. M. Han, T. Zin and H. Tun, “Extraction of audio features for emotion recognition system based on music,” **Int. Journal of Scientific & Technology Research**, vol.5, no. 6, 2016, pp. 53–56.

[9] F. Zhang, H. Meng and M. Li, “Emotion extraction and recognition from music,” **Proc. of International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery,** 2016, pp. 1728- 1733.

[10] Y. An, S. Sun and S. Wang, “Naive Bayes classifiers for music emotion classification based on lyrics,” **Proc. Of Int. Conference on Computer and Information Science,** 2017, pp. 635-638.

[11] D. Yang and W.S. Lee. “Music emotion identification from lyrics,” **Proc. of Int. Symposium on Multimedia,** San Diego, CA, USA, 2009, pp. 624-629.

[12] Y. E. Kim, Erik M. Schmidt and R. Migneco, “Music emotion recognition: A state of the art review,” **Proc. of Int. Society for Music Information Retrieval Conference,** pp. 255-266, 2010.

[13] D. Ayata, Y. Yaslan and M. E. Kamasak, “Emotion Based Music Recommendation System Using Wearable Physiological Sensors,” **IEEE Trans. on Consumer Electronics,** vol. 64, no. 2, May. 2018. pp. 196-203.

[14] R. Panda and R. Pedro Paiva, “Music emotion classification: Dataset acquisition and comparative,” **Proc. of Int. Conference on Digital Audio Effects,** York, UK, 2012.

[15] C. Lin, M. Liu, W. Hsiung and J. Jhang, “Music emotion recognition based on two-level support vector classification,” **Proc. of Int. Conference on Machine Learning and Cybernetics,** 2016, pp. 375-379.

[16] J. Bai, J. Peng and J. Shi, “Dimensional music emotion recognition by valence arousal regression,” **Proc. of Int. Conference on Cognitive Informatics & Cognitive Computing,** 2016, pp. 42-49.

[17] Y. H. Yang, Y. C. Lin, Y. F. Su, and H. H. Chen, “A regression approach to music emotion recognition,” **IEEE Trans. Audio, Speech & Language Processing,** Vol. 16, no. 2, 2008, pp. 448–457.

ภาคผนวก

ภาคผนวก ก
บทความวิจัยที่ได้รับการตีพิมพ์

[1] S. sangnark, M. Lertchawakul and C. Benjangkaprasert, “Thai Music Emotion Recognition Based on Western Music,” **Int. Conf. on Computer and Electrical Engineering**, Oct. 2018.

[2] S. sangnark, M. Lertchawakul and C. Benjangkaprasert, “Thai Music Emotion Recognition by Linear Regression,” **Int. Conf. on Automation, Control and Robots**, Oct. 2018.

ประวัติผู้เขียน

ชื่อ-นามสกุล	นายสรวิศ แสงนาค
วัน เดือน ปีเกิด	5 พฤษภาคม 2537 ณ จังหวัด นนทบุรี
เบอร์โทร	086-8951605
E-mail	soravitt.sang@gmail.com
ที่อยู่	105/300 หมู่ 6 ถนนกาญจนาภิเษก ตำบลบางรักพัฒนา อำเภอบางบัวทอง จังหวัด นนทบุรี 11110
ประวัติการศึกษา	2559 วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง